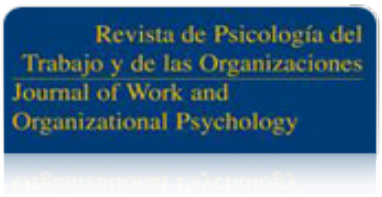


Revista de Psicología del Trabajo y de las Organizaciones
ISSN: 1576-5962
ISSN: 2174-0534
Colegio Oficial de la Psicología de Madrid

Martínez, Alexandra; Salgado, Jesús F.; Lado, Mario

Quasi-ipsative Forced-Choice Personality Inventories and the Control of Faking: The Biasing Effects of Transient Error

Revista de Psicología del Trabajo y de las Organizaciones,
vol. 38, no. 3, 2022, December, pp. 241-248
Colegio Oficial de la Psicología de Madrid

DOI: https://doi.org/10.5093/jwop2022a16

Available in: https://www.redalyc.org/articulo.oa?id=231374563010

Journal of Work and Organizational Psychology (2022) 38(3) 241-248

Journal of Work and Organizational Psychology

https://journals.copmadrid.org/jwop

# Quasi-ipsative Forced-Choice Personality Inventories and the Control of Faking: The Biasing Effects of Transient Error

Alexandra Martínez, Jesús F. Salgado, and Mario Lado

*University of Santiago de Compostela, Spain*

ARTICLE INFO

*Article history:*
Received 16 March 2022
Accepted 7 September 2022
Available online 19 October 2022

*Keywords:*
Faking
Transient error
Quasi-ipsative forced-choice inventories
Latin-square design
Personnel selection

A B S T R A C T

To date, experimental research on the effect of faking on personality measures has used two types of designs: within-subject designs and between-subjects designs. None of these designs permit us to control for the effects of transient error on faking. Using a Latin-square design (LSD), the current study examines the effects of faking on the Big Five as assessed by a quasi-ipsative forced-choice (FC) personality inventory. LSD is a type of experimental design that simultaneously permits us to control for between-subject differences, within-subject variability, and transient error. The sample consisted of 246 participants (four experimental groups, assessed twice, 2-3 week interval). The results showed that (1) faking effect size can be largely attributed to transient error and (2) the quasi-ipsative FC format shows great resistance to faking behavior. The average effect size (Cohen's *d*) for the Big Five was 0.21, 0.12, and 0.09 for observed faking, transient error, and true faking, respectively. On average, 62% of observed faking effect size can be attributed to transient error. To conclude, we discuss the implications of these findings for the research and practice of personnel selection.

## Los inventarios de personalidad de elección forzosa cuasi-ipsativos y el control del *faking*: los efectos de sesgo del error temporal

*Palabras clave:*
Falseamiento
Error temporal
Inventarios de elección forzosa cuasiipsativos
Diseño de cuadrado latino
Selección de personal

R E S U M E N

La investigación experimental sobre los efectos del *faking* o falseamiento en las medidas de personalidad ha utilizado dos tipos de diseños: diseños intrasujeto y diseños entre sujetos. Pero ninguno de ellos nos permite controlar los efectos del error temporal en el *faking*. Usando un diseño de cuadrado latino (DCL), este estudio examina los efectos del *faking* en los Cinco Grandes evaluados con un inventario de personalidad de elección forzosa (EF) cuasi-ipsativo. El DCL es un diseño experimental que simultáneamente nos permite controlar las diferencias entre sujetos, la variabilidad intrasujeto y el error temporal. La muestra estuvo compuesta por 246 participantes (cuatro grupos experimentales, evaluados dos veces en un intervalo de 2-3 semanas). Los resultados mostraron que (1) el tamaño del efecto del *faking* se puede atribuir en gran medida a un error temporal y (2) el formato de EF causi-ipsativo muestra una gran resistencia al *faking*. El tamaño del efecto promedio (*d* de Cohen) para los Cinco Grandes fue 0.21, 0.12 y 0.09 para el *faking* observado, el error temporal y el *faking* verdadero, respectivamente. En promedio, el 62 % del tamaño del efecto del *faking* observado se puede atribuir a un error temporal. Para concluir, se discuten las implicaciones de estos resultados.

Faking is one of the most pervasive phenomena in work and organizational psychology (W/O) and, particularly, in the personnel selection area. For instance, empirical findings have shown that applicants can fake their responses to personality measures (e.g., the Big Five) when they are involved in an assessment process (such as personnel selection, work promotions processes, or academic decisions) increasing or decreasing their actual scores on these measures (Martínez, 2019; Martínez et al., 2021a, 2021b; Martínez & Salgado, 2021; Morgeson et al., 2007; Murphy, 2005; Salgado, 2016; Zickar & Gibby, 2006). Consequently, this behavior may have serious consequences for hiring and organizational decision-making because the quality of the decisions may be questionable (see, for instance, Donovan et al., 2014; Griffith et al., 2007; Martínez, 2019; Morgeson et al., 2007; Salgado, 2016). The effects of faking do not affect personality inventories only, but also other assessment tools such as personnel interviews, biodata, and integrity tests, to mention but a few examples. Furthermore, this is a widespread phenomenon as anyone could commit faking, and any organization that uses

Cite this article as: Martínez, A., Salgado, J. F., & Lado, M. (2022). Quasi-ipsative forced-choice personality inventories and the control of faking: The biasing effects of transient error. *Journal of Work and Organizational Psychology, 38*(3), 241-248. https://doi.org/10.5093/jwop2022a16

Funding: This research was partially supported by grant PID2020-116409GB-I00 from the Spanish Ministry of Science and Innovation.
Correspondence: alexandra.martinez@usc.es (A. Martínez).

personality inventories could have their selection processes affected by faking (Griffith & Converse, 2012). For these reasons, a major concern in applied personality measurement contexts (e.g., personnel selection) is to, firstly, know the extent to which personality scores are affected by faking and, secondly, to develop instruments, methods and assessment strategies that permit us to control and reduce the effects of faking. This paper aims to shed further light on these issues.

With regard to the first concern, the extent of faking in personality scores over the years, many studies have been carried out to estimate the effects of faking on personality inventories. Those studies have shown that faking has significant negative consequences on the psychometric properties of personality inventories (Salgado, 2016). For instance, faking increases the mean and decreases the standard deviation (*SD*) of the score distribution of personality variables. In addition, faking also produces a decrease in reliability and criterion validity and can modify the factor structure of personality inventories (e.g., Birkeland et al., 2006; Douglas et al., 1996; Hooper, 2007; Hough et al., 1990; Martínez et al., 2021a, 2021b; Salgado, 2016; Salgado & Lado, 2018; Viswesvaran & Ones, 1999).

Typically, research has examined the effects of faking by comparing the scores under two conditions: (a) responding under honest instructions and (b) responding following faking instructions (see, for instance, Cao & Drasgow, 2019; Martínez & Salgado, 2021; Mesmer-Magnus & Viswesvaran, 2006; Salgado & Lado, 2018; Viswesvaran & Ones, 1999). Typically, these two sets of scores are obtained within a single group of participants (within-subject design) or with two independent groups of participants (between-subject design). To the best of our knowledge, all the studies, both correlational and experimental, that have examined the effects of faking on personality inventories adopted one of these two designs. Nevertheless, they have some methodological limitations that might contribute to the overestimation or underestimation of the extent of faking, as they do not control some sources of variance that may be affecting the measurements (e.g., idiosyncratic variations or transient variations). Despite these design limitations, to date, no studies have been conducted with other experimental designs that might control for these sources of error.

Regarding the second above-mentioned concern, i.e., how to control and reduce faking, several assessment methods have emerged over the years with different degrees of effectiveness in their capacity to control and reduce faking. Among them, recent empirical evidence has shown that forced-choice (FC) personality inventories stand out as methods capable of reducing the effects of faking on personality measures (see, for instance, Converse et al., 2006; Dilchert & Ones, 2012; MacCann et al., 2012; Morillo et al., 2019; Salgado, 2017; Salgado & Lado, 2018). Typically, FC personality inventories require that individuals choose between response options with the same or very similar degrees of social desirability. This characteristic makes it more difficult for candidates to voluntarily distort their responses.

Hence, the aim of this research is to contribute to the study of the effects of faking on FC personality measures using an experimental design never applied to date in the study of faking behavior, the Latin-square experimental design. The Latin-square design (LSD) is a type of experimental design that consists of randomly assigning subjects to different instructional sets created with the aim of controlling several sources of variance in order to know, in a more specific way, the effects of a concrete phenomenon on the relationship between two variables (Cochran & Cox, 1978; Kirk, 2013). In the next sections, we discuss the concept and effects of faking, we point out the characteristics of the various FC personality methods and their faking resistance, we review the literature on the experimental designs used to estimate the effects of faking, and we introduce the LSD as a better method for estimating the true effect of faking on personality scores.

## Concept and Effects of Faking

Faking can be described as a type of response bias wherein individuals voluntarily distort their responses to non-cognitive instruments (e.g., personality tests, interviews, biodata, etc.) trying to give a portrait of themselves that provides them some benefit or advantage in the assessment processes (e.g., Birkeland et al., 2006; Donovan et al., 2014; García-Izquierdo et al., 2020; Griffith & Converse, 2012; Salgado, 2016). As noted by Ziegler et al. (2012), faking occurs when there is a perception on the part of the subject that there is an imbalance between the situational demands and his/her individual characteristics.

In this regard, numerous studies that have shown that faking can severely and negatively affect some selection instruments, particularly the single-stimulus (SS) personality inventories (e.g., NEO-PI-R, MMPI-II, California Personality Inventory, Hogan Personality Inventory, 16PF-V), perhaps the most commonly personality instruments used to make hiring decisions (Rothstein & Goffin, 2006). A main characteristic of SS personality inventories is that every statement must be rated in to describe individual personality (for instance, using Likert scale, yes/no, or true/false answer format). Many researchers and practitioners were (and are) concerned with the potential sensitivity of SS to be faked when hiring decisions are to be taken (see Christiansen et al., 2005; Murphy, 2005).

Recently, Salgado (2016) proposed a theoretical account to explain the psychometric effects of faking. According to this theory, faking is a source of error variance that produces two simultaneous artifactual effects: increases the mean and reduces the standard deviation of the distributions. In other words, faking produces an artificial homogenization of the samples reducing the range of scores obtained by the individuals. Also, this causes that the individuals seem more similar to each other than they really are (due to the restriction of the range of scores). Likewise, these artifactual effects produce a reduction in the reliability and predictive validity of personality instruments and affect the factorial structure (construct validity) of the SS measures. Accordingly, faking would have a direct consequence on the selection processes: applicants that commit faking would be undeservedly in higher ranking positions than those applicants who have not faked. Consequently, the evaluators could make wrong hiring decisions during assessment processes (Griffith et al., 2007; Komar et al., 2008; Salgado, 2016).

Empirical findings have shown that the main effect of faking on SS inventories was an artificial increase of the mean, a reduction in the standard deviations of the scores, and the lowering of the reliability coefficients (Salgado, 2016). These results have also been supported by the meta-analysis of Viswesvaran and Ones (1999) in experimental settings, and by the meta-analyses of Birkeland et al., (2006), Hooper (2007), and Salgado (2016) in occupational settings, where they compared real job applicants' samples and non-applicants' samples (i.e., incumbents or respondents in a non-selection setting). Empirical evidence concluded, therefore, that individuals can and do distort their scores on SS instruments if they are motivated to fake.

Accordingly, it was suggested that forced-choice (FC) personality inventories might be a robust alternative procedure to SS personality inventories in order to cope with the effects of faking in practical assessment contexts (e.g., Cao & Drasgow, 2019; Fisher et al., 2019; Martínez, 2019; Martínez & Salgado, 2021; Morgeson et al., 2007; Salgado & Táuriz, 2014).

## Forced-Choice Personality Inventories

In the last few years, FC personality inventories have been recommended as measures that might better control the effects of faking than SS tests do (Adair, 2014; Cao & Drasgow, 2019; Martínez et al., 2021a, 2021b; Martínez & Salgado, 2021). FC personality

inventories are characterized by presenting sets of items (usually, the options are grouped in pairs, triads, or tetrads) that have the same degree of social desirability (see, for instance, Baron, 1996; Bartram, 1996; Cao & Drasgow, 2019; Salgado & Táuriz, 2014). Respondents must choose the alternative that best describes them and, in some cases, the alternative that worst describes them. Given that the answer options are similar in their level of social desirability, it will be more difficult for the participants to distort their responses. Consequently, the use of FC personality inventories might reduce the effects of faking (Brown & Maydeu-Olivares, 2013; Christiansen et al., 2005; Converse et al., 2006; Dilchert & Ones, 2012; Jackson et al., 2000; Martínez, 2019; Martínez & Salgado, 2021; Morillo et al., 2019; Salgado, 2017; Salgado & Lado, 2018).

Typically, three types of FC scores (i.e., normative, ipsative, and quasi-ipsative) are distinguished depending on how the answer-choice is made, each of them with specific psychometric characteristics (see Abad et al., 2022; Clemans, 1966; Hicks, 1970; Meade, 2004; Salgado et al., 2015; Salgado & Táuriz, 2014). Normative FC scores present unidimensional items, each item being a scale that just evaluates one personality factor. So, the normative scores allow for the analysis of the results on an inter-individual level. In the case of ipsative FC scores, all the alternatives presented in each item are scored by the respondents. Therefore, there is dependence in the scores between factors, that is, the score for each dimension depends on an individual's scores on the other graded dimensions. Consequently, the sum of the scores obtained for each individual is a constant, and ipsative scores only allow intra-individual comparisons. The third FC scores, the quasi-ipsative ones, do not meet all the criteria of pure ipsative scores but present some characteristics associated with them (Clemans, 1966). Specifically, a score is quasi-ipsative when it presents the following characteristics (Hicks, 1970; Horn, 1971; Salgado & Táuriz, 2014): (1) the results for each factor vary between the individuals over a certain range of scores, (2) even though these inventories have some properties in common with the ipsative FC, the scores do not add up to the same constant for all individuals, (3) the increase in the score on one personality factor does not necessarily produce a decrease in the other factors. In fact, we can distinguish two types of quasi-ipsative FC scores (Horn, 1971): (1) quasi-ipsative, algebraically-dependent scores, which show some degree of ipsativization of scores or, in other words, a metric dependence exists between the scores, and (2) quasi-ipsative, non-algebraically-dependent scores, where the score for each personality factor is not influenced by the score in other personality dimensions.

Empirical evidence has shown that FC inventories show an important degree of faking resistance. For instance, the meta-analyses of Adair (2014), Cao and Drasgow (2019), and Martínez and Salgado (2021) have shown that, compared with the SS personality inventories, the three types of FC personality inventories reduce the effects of faking on personality scores, both in experimental studies and in real personnel selection contexts. In particular, the quasi-ipsative FC inventories stand out from the other two formats in faking-control effectiveness, as the effects sizes found were smaller than those found for normative and ipsative scores (Martínez & Salgado, 2021).

## Experimental Designs for Estimating the Effects of Faking

Despite extensive research on faking, a matter of debate is how faking should be estimated and/or measured. The difficulties encountered when evaluating this construct are largely due to its intrinsic nature and the diverse set of behaviors that it comprises, but also the different methods used to estimate it (Ellingson et al., 2007; Martínez, 2019; Ones et al., 1996; Viswesvaran & Ones, 1999). Different methods have been applied to quantify the degree of faking, for instance, social desirability scales or the application of confidence intervals.

The most widely used method to examine the effects of faking are experimental studies that allow comparisons among the scores obtained both under honest and under faking instructions conditions (Mesmer-Magnus & Viswesvaran, 2006; Smith & McDaniel, 2012; Viswesvaran & Ones, 1999). This strategy can be carried out by combining two types of characteristics: (a) the research context and (b) the type of design used.

In relation to the research context, two types of studies can be distinguished: (1) correlational studies (i.e., studies in real contexts) and (2) experimental studies (i.e., lab studies). In personnel selection, the first type of studies refers to the comparison between actual job applicant samples and incumbent samples. This perspective assumes that applicants are motivated to fake but that incumbents are not motivated to voluntarily modify their answers. Therefore, the differences between applicants and incumbents would indicate whether or not faking has been committed (see Mesmer-Magnus & Viswesvaran, 2006). This method can be considered an indicator of typical performance on the test, that is, differences that occur naturally in a real personnel selection context. However, given the difficulties involved in obtaining equivalent samples of applicants and incumbents, most studies are carried out in experimental contexts in which selection processes are simulated.

In experimental settings, two different strategies have been used until now. In the first strategy, participants are randomly assigned to a fake good response-condition (equivalent to the role of applicants) and to an honest response-condition (equivalent to the role of incumbents) before responding to the tests. In the fake good response-condition, the participants are instructed to answer the test by trying to show an improved ("the best") image of themselves. In the honest response-condition, they are instructed to be totally honest and candid when answering the tests. In those cases, the experimental design is a between-subject one. The second strategy requires the participants to take the personality inventory twice, first in one condition (e.g., under honest-response instructions) and next in the second condition (e.g., under faking-response instructions). The comparison between the scores in both conditions would produce an estimate of faking. In those cases, the experimental design is a within-subject one.

In experimental settings, it is considered that the results reflect the effects of faking in contexts of maximum performance in which it is clearly intended to find the greatest difference between both experimental conditions.

Salgado (2016) carried out a meta-analytic review of the effects of faking in which the two scenarios were considered (i.e., real contexts and lab contexts). The results pointed out that faking estimates obtained in a real context (i.e., applicants vs. incumbents) and faking estimates obtained via an experimental design produce very similar effect sizes. This finding has an important implication for empirical research since it can be assumed that the results found in experimental studies have the same accuracy as those found in real contexts.

In examining the consequences on faking estimates of using one design type or the other, on the one hand, within-subject designs allow us to compare the individual differences that can occur within the same group under different conditions (see Viswesvaran & Ones, 1999). However, this methodology is not without limitations. The most prominent one is the suggestion that the results may be affected by the subject's pretest experience (remember that they complete the same test twice), as well as by the personal characteristics of each individual evaluated (Cook & Campbell, 1979; Mesmer-Magnus & Viswesvaran, 2006; Viswesvaran & Ones, 1999). On the other hand, between-subject designs are carried out with two independent samples; therefore, the main criticism of between-subject designs is that it is assumed that the effect of faking will be the same for all subjects even though only a part of the sample responds under

faking-responding instructions and the other part responds under honest-responding instructions.

An issue that arises about these two types of designs is whether equally accurate results are obtained regardless of which of them is applied. Viswesvaran and Ones (1999) studied the incidence of faking in personality measures using both designs and compared the resulting estimators in each one. They concluded that within-subject designs are much more accurate in explaining the effects of faking. They assumed that the within-subject designs have greater statistical power by allowing the two groups to be fully equated. However, both are typical "before-after" designs (designs where we only obtain one answer under each condition) and some researchers have pointed out that this type of design does not control some sources of variance, particularly transient measurement error, whose effects could be misassigned to faking (Heggestad et al., 2006; Martínez, 2019). Consequently, other experimental strategies that control these sources of error, like for instance Latin-square designs (LSD), should be applied to get a more accurate estimate of faking effects (Kirk, 2013).

## Latin Square Design as an Alternative Design for Studying Faking

LSD is a type of experimental design with repeated measures that is appropriate to use when it is necessary to control two sources of variability, i.e., error due to the treatment (faking) and transient errors. In LSD, the number of levels of the main factor (faking/honest behavior in our case) must coincide with the number of levels of the secondary factor (for instance, the order of the instructions). Also, it must be assumed that there is no interaction between any pair of variables (Cochran & Cox, 1978; Kirk, 2013).

Let $K$ be the number of levels of each of the factors, then LSD uses $K^2$ blocks and each of these blocks corresponds to one of the possible level combinations of the two control variables. In each block, a single experimental condition is applied so that each experimental condition must appear with each of the levels of the two control variables (see Gao, 2005; Grant, 1948; Kirk, 2013). That is, if we consider a table where the rows and columns represent each of the two-block variables and the cells the levels of the main variable, this means that each variable must appear once in each row and in each column (see Figure 1).

Therefore, the sample in an LSD is divided in several independent experimental groups, each of which is single "evidence" of the combinations of variables. Furthermore, as the combination of variables is assigned randomly in each experimental group, systematic biases are avoided. In other words, LSD allows us to isolate the differences between the samples and experimental and environmental conditions (Cochran & Cox, 1978; Grant, 1948; Kirk, 2013).

In summary, LSD can be more informative on the effect of the main variable studied (i.e., faking behavior in our case), but it will be also more informative about the differences observed in the samples due to the individual and contextual characteristics (e.g., transient error). Therefore, from a methodological viewpoint, using an LSD will permit us to better examine the effects of faking on FC personality inventories and to obtain a more accurate estimate of faking effect size.

## Aims of the Study and Research Hypothesis

The current study aims to contribute to the knowledge on faking in FC personality measures by using an experimental design never applied to date in the study of the faking effects, the LSD. Specifically, the main objective is to know the amount of faking in a Big Five FC quasi-ipsative personality inventory, after controlling for the two main sources of variability that may be affecting the results, i.e., honest vs. faking behavior and transient error. Therefore, with this design, we aim to isolate the differences between the samples that are because of faking from those that are due to individual or environmental differences (i.e., transient measurement error).

Based on the previous above cited theoretical and empirical evidence, we state the following hypotheses.

*Hypothesis 1:* Faking produces an increase in the mean and a reduction in the standard deviation of the scores of the personality factors.

*Hypothesis 2:* Quasi-ipsative FC personality inventories show smaller observed faking effect as compared with the one found for SS personality inventories revealed in previous meta-analyses (e.g., Birkeland et al., 2004; Viswesvaran & Ones, 1999; Salgado, 2016).

*Hypothesis 3:* The transient error in personality scores accounts for a significant percentage of the observed faking effect size.

*Hypothesis 4:* When the effects of transient error on personality scores are discounted, the magnitude of the faking effect size in a Big Five FC personality inventory is very small on average ($d < 0.10$).

# Method

## Sample

The sample was composed of 246 undergraduates from a Spanish university; 65.85% of the sample were women ($n$ = 162) and 34.15% were men ($n$ = 84). All of them were first-year students. The average age was 19.70 years old ($SD$ = 3.73). Participation in this experimental study was voluntary. Small face-to-face groups were organized to answer the tests. The subjects provided informed consent to participate in the study.

## Measures

### ***Quasi-ipsative Personality Inventory*** (QI5F_tri; Salgado, 2014)

The QI5F_tri is a quasi-ipsative forced-choice personality inventory designed to assess the Big Five personality factors. The test consists of 140 triads that evaluate the Big Five (each factor is assessed by 28 triads). In each triad, individuals are required to indicate which option best describes them and which option worst describes them. An example of a triad is: "I am a person who is (a) altruistic (agreeableness), (b) perfectionist (conscientiousness), (c) very imaginative (openness to experience)". The QI5F_tri personality

| 2 x 2 | |
|---|---|
| A | B |
| B | A |

| 3 x 3 | | |
|---|---|---|
| A | B | C |
| B | C | A |
| C | A | B |

| 4 x 4 | | | |
|---|---|---|---|
| A | B | C | D |
| B | C | D | A |
| C | D | A | B |
| D | A | B | C |

**Figure 1.** Examples of Latin Squares Designs.

inventory provides a score for each factor that is algebraically non-dependent from the score on the other factors, that is, it implements Horn's strategy: the items used to evaluate a dimension are not used to evaluate other dimensions. Consequently, the factors are algebraically non-dependent even though the nature of the score is quasi-ipsative (Horn, 1971; Salgado, 2014; Salgado & Lado, 2018). The internal consistency coefficients (McDonald's ω) were .74, .77, .86, .72, and .81, and .87, .83, .93, .85, and .89 (great lower bound; GLB) for emotional stability (ES), extraversion (EX), openness to experience (OE), agreeableness (A), and conscientiousness (C), respectively (Salgado, 2014). Test-retest reliabilities ranged from .72 for agreeableness to .86 for openness to experience (Salgado, 2014). In the present sample, test-retest reliabilities were ,82, .83, ,83, .78, and .84, for ES, EX, OE, A and C, respectively. Evidence of the convergent validity of the QI5F_tri using an SS personality inventory was reported by Otero et al. (2020) and Martínez et al. (2021a; 2021b).

## Experimental Design and Procedure

To carry out this study, we used a 2 x 2 Latin-square experimental design (LSD) with repeated measures. The participants answered the personality questionnaire twice at two different times. Therefore, data collection was performed in two sessions with each group, leaving a time interval between each of the sessions of between 2 and 3 weeks. In each session, participants were randomly assigned to one of two experimental conditions (honest or faking), so that from the combination of the experimental conditions and the number of sessions the participants were included in one of the following four experimental groups (instructional sets): honest-honest (H-H), honest-faking (H-F), faking-honest (F-H), and faking-faking (F-F) (see Figure 2).

| | | Response-Instructions | |
|---|---|---|---|
| | | H | F |
| Sequence | H | HH ($n$ = 65) | HF ($n$ = 63) |
| | F | FH ($n$ = 55) | FF ($n$ = 63) |

**Figure 2.** Latin Square Design Used in the Study.

A random number table was used to randomize participants for each response condition in each session. Specifically, a number was assigned to each participant, which had previously been randomly associated with an experimental condition.

Regarding the instructions to answer the QI5F-tri personality inventory, for the honest condition, the participants received the following instructions provided by the instrument:

*In this questionnaire you will be presented with sets of phrases grouped into triads. Try to rank them by first identifying the one that best describes you, the one that second best describes you, and finally the one that describes you least. In each item, mark a plus sign (+) next to the phrase that best describes you and a minus sign (-) next to the phrase that least describes you. You should leave blank the one you considered second.*

For the faking condition, the instructions were modified in such a way that participants were encouraged to fake. The following paragraph was added to the above instructions:

*When answering, assume that you are in the last step of a selection process for a very attractive job. Since it offers you a great opportunity to advance your professional career, you want to get that job. To do this, you must answer the test trying to give a better image of yourself.*

In both response conditions the inventory was administrated using the Google Forms application and the participants only had access to the test during the time they attended the study. Due to confidentiality reasons, the data set is only available from the first author upon a reasonable request.

## Results

Table 1 reports the descriptive statistics for the Big Five in each experimental group. Taken together, the means and the standard deviations obtained for all groups and experimental conditions are very similar. There are no big differences in the values for all experimental conditions, though we can appreciate slight variations in the results across groups. For instance, comparing the results for the H-H condition ($M$ = 23.45~29.63) with the results for the F-F condition ($M$ = 22.46~30.10) it can be observed that there is some degree of variability in the magnitude of the mean and standard deviation values, even though the participants answered both times under the same response-instructions. These differences were more noticeable between the honest and faking response-conditions in the groups in which the participants answered the inventory under both instructional sets. Specifically, the mean ranged from 22.57 to 31.75 in the H-F response-conditions and from 21.51 to 30.39 in F-H instructions. In particular, the results showed that the means are slightly higher under faking response-conditions than under honest response-conditions and that the values of the standard deviations are smaller under faking instructions compared to the values under honest response-conditions. These findings supported Hypothesis 1.

The results also showed that there is some degree of variability in the means when the participants respond under the same experimental instructional sets (i.e., H-H and F-F). Clearly, this fact indicates that transient measurement error is present in estimated observed faking as predicted by Hypothesis 2.

In order to estimate the effect size for the transient error, we

**Table 1.** Descriptive Statistics of the Big Five in the Four Experimental Groups of the Latin Square

| | H-H | | | | F-F | | | | H-F | | | | F-H | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | T1 | | T2 | | T1 | | T2 | | T1 | | T2 | | T1 | | T2 | |
| | *M* | *SD* | *M* | *SD* | *M* | *SD* | *M* | *SD* | *M* | *SD* | *M* | *SD* | *M* | *SD* | *M* | *SD* |
| ES | 24.60 | 7.52 | 23.45 | 6.95 | 22.46 | 5.84 | 23.79 | 6.40 | 22.57 | 7.21 | 24.87 | 6.68 | 23.11 | 6.39 | 21.51 | 6.74 |
| EX | 26.66 | 7.61 | 26.60 | 7.25 | 27.10 | 7.86 | 26.19 | 7.36 | 26.59 | 7.68 | 25.62 | 7.96 | 26.94 | 7.66 | 27.45 | 8.54 |
| OE | 25.66 | 7.88 | 24.77 | 8.24 | 25.92 | 8.50 | 25.33 | 8.94 | 24.40 | 8.60 | 26.81 | 8.46 | 27.85 | 6.98 | 27.49 | 8.27 |
| A | 28.82 | 6.09 | 29.63 | 6.87 | 30.10 | 6.03 | 28.87 | 7.04 | 31.75 | 7.12 | 30.89 | 6.51 | 28.70 | 6.13 | 30.39 | 6.83 |
| C | 27.94 | 7.53 | 28.49 | 7.42 | 28.29 | 7.95 | 29.29 | 8.48 | 27.05 | 8.89 | 30.94 | 9.07 | 29.66 | 7.93 | 28.06 | 8.51 |

*Note.* $N_{H\text{-}H}$ = 65; $N_{F\text{-}F}$ = 63; $N_{H\text{-}F}$ = 63; $N_{F\text{-}H}$ = 55; ES = emotional stability; EX = extroversion; OE = openness to experience; A = agreeableness; C = conscientiousness; H = honest condition; F = faking condition; T1 = first time; T2 = second time.

used the formulas developed by Dunlap et al. (1996) to estimate *d* coefficient. Dunlap et al.'s formulas for *d* allow for a more accurate estimate of effect size than the conventional Cohen's *d* formula, as they take into account the correlation between matched groups. This means that these formulas control to a greater extent the inside design variability. The results obtained are reported in Table 2.

As can be seen, the *d* values ranged from -0.12 to 0.16 when comparing the values among H-H conditions. The *d* values ranged from -0.22 to 0.19 for comparisons among F-F conditions. The absolute (i.e., sign ignored) average *d* value was 0.09 for the H-H condition and it was 0.14 for the F-F condition. Therefore, these results showed that there was variability in the response of the participants even when the participants answered the personality inventory under the same response-instructions both times. Thus, transient error showed a small but relevant effect size in the stability of participants' personality scores under both honest and faking instructional sets.

In the H-F experimental condition, the values of *d* ranged from -0.41 to 0.13, with an absolute average d of 0.25. Lastly, the values of *d* ranged from -0.26 to 0.24 for the F-H experimental condition, with an absolute *d* average of 0.16. Pooling the two sets of estimates, the absolute average *d* is 0.21 for the comparison between honest and faking instructional sets.

**Table 2.** Estimates of *d* for the Comparisons among the Experimental Groups of the Latin Square

| | H-H | F-F | H-F | F-H |
|---|---|---|---|---|
| ES | 0.16 | -0.22 | -0.33 | 0.24 |
| EX | 0.01 | 0.12 | 0.12 | -0.06 |
| OE | 0.11 | 0.07 | -0.28 | 0.05 |
| A | -0.12 | 0.19 | 0.13 | -0.26 |
| C | -0.07 | -0.12 | -0.41 | 0.20 |
| *Average* | *0.02* | *0.01* | *-0.15* | *0.03* |
| *Absolute average* | *0.09* | *0.14* | *0.25* | *0.16* |

*Note.* $N_{H\text{-}H}$ = 65; $N_{F\text{-}F}$ = 63; $N_{H\text{-}F}$ = 63; $N_{F\text{-}H}$ = 55; ES = emotional stability; EX = extroversion; OE = openness to experience; A = agreeableness; C = conscientiousness; H = honest condition; F = faking condition.

Regarding the effect of faking on personality scores, the first column of Table 3 reports the degree of observed faking for the Big Five (average of d estimates of H-F and F-H conditions, sign reversed). The results show *d* coefficients that can be considered moderately small for conscientiousness and emotional stability (0.31 and 0.29, respectively) to small or very small for agreeableness, openness to experience, and extraversion (0.20, 0.17, and 0.09, respectively). These findings indicate that the quasi-ipsative FC format of the current personality inventory is very robust against the faking effects, if we compared these estimates with those reported in previous meta-analyses for the SS personality inventories. For instance, Viswesvaran and Ones (1999) reported *d* values ranging from 0.47 to 0.93 (average *d* = 0.66). Therefore, the current *d* estimates are remarkably smaller than those previous estimates which supports Hypothesis 2. Also, the observed estimates of faking effects found in this study concur with the findings of Cao and Drasgow (2019) and Martínez and Salgado (2021) for the quasi-ipsative FC.

The second column of Table 3 reports the effect size for the transient error in the personality scores under the same instructional set (average of *d* estimates of H-H and F-F conditions, sign reversed). The results show *d* coefficients ranging from 0.07 to 0.19, with an average *d* of 0.12.

**Table 3.** Estimates of d for Observed Faking, Transient Error, and True Faking

| | (A) Observed Faking *d* | (B) Transient Error *d* | (A-B) True Faking *d* | % Observed Faking Due to Transient Error |
|---|---|---|---|---|
| ES | 0.29 | 0.19 | 0.10 | 66 |
| EX | 0.09 | 0.07 | 0.02 | 78 |
| OE | 0.17 | 0.09 | 0.08 | 53 |
| A | 0.20 | 0.16 | 0.04 | 80 |
| C | 0.31 | 0.10 | 0.20 | 32 |
| Average | 0.21 | 0.12 | 0.09 | 62 |

*Note.* ES = emotional stability; EX = extroversion; OE = openness to experience; A = agreeableness; C = conscientiousness; H = honest condition; F = faking condition.

The third column of Table 3 reports the effect sizes of true faking for the Big Five. We label "true faking" the faking estimate obtained when the effect size of transient error is discounted from the observed faking effect size. As can be seen, true faking effect sizes are small or very small, ranging from 0.02 to 0.20, with an average size of 0.09. This finding means that the effect of faking on the personality scores obtained with a quasi-ipsative FC inventory are of almost no practical importance when the effect of transient error is discounted. The fourth column shows that, on average, transient error accounts for 62% of observed faking effect size in personality scores. As a whole, these findings fully support Hypotheses 3 and 4.

## Discussion

The main purpose of this study was to expand the literature on the effects of faking on personality measures by using an experimental design that permits us to control for the effects of transient error on the estimates of faking. As we previously pointed out, LSD is a type of experimental design that, due to its characteristics, allows us to control the effects of some sources of variability, such as transient measurement error, that other experimental designs (e.g., within-subject or between-subject designs) widely used in faking research cannot control. Because of this, LSDs are a more accurate way to know the consequences of faking because they reflect with more precision the true (real) effects of faking and the variance produced by other characteristics of the sample or the design.

Therefore, the current study has contributed to the study of faking and its effects in several ways. First, this study provides empirical evidence of the robustness of quasi-ipsative FC inventories resistance to scoring inflation in personality dimensions (i.e., the Big Five). In the current study this fact was shown by using a quasi-ipsative FC inventory that provides non-algebraically dependent scores.

Secondly, in accordance with the psychometric theory of faking effects (Salgado, 2016), the results showed that faking behavior increases the mean and decreases the standard deviation of personality scores which confirms the predictions of the theory.

The third contribution, and a unique one, has been to show that previous estimates of faking obtained with within-subject and between-subject experimental designs were affected by transient error to some extent. The transient error is a critical source of measurement error that can be substantial in some measures, e.g., faking (Becker, 2000; Schmidt et al., 2003; Vautier & Jmel, 2003). This study demonstrated that the transient error has a dramatic effect on faking estimates, as 62% of the faking effect size was due, on average, to transient errors. Consequently, it is of critical importance to take into account the potential effects of transient errors while keeping in mind the effect of transient error permits us to obtain two estimates of faking effect size, observed faking and true (real) faking. To the best

of our knowledge, previous research has not distinguished between these two estimates of faking so far, and researchers have attributed the longitudinal variations in responses to personality measures due to random variation in respondents' psychological states across time (i.e., transient error; Schmidt et al., 2003) to faking. Making this attribution, researchers implicitly assumed that the total amount of observed faking was true faking. Clearly this assumption was not supported in the current study.

From the methodological point of view, a third relevant contribution of the study has been to use an LSD to establish the real effects of faking. To the best of our knowledge, no previous research has used this design to examine faking effects. The LSD enabled us to estimate the amount of variability in honest and faking response conditions that it is misassigned to faking. The LSD showed that the experimental groups that responded both times under the same response instructions showed random variations that are not relevant to the construct under study, i.e., faking. In other words, the effect sizes were different from zero, contrary to the assumption of previous research.

The fourth contribution has been to show that, in comparison with SS personality inventories, quasi-ipsative FC inventories without algebraic dependence are a method capable of controlling the effects of faking on the scores. Globally, the *d* values provided by the current study are substantially smaller than those obtained for SS measures (e.g., Viswesvaran & Ones, 1999).

The fifth contribution has been to show that faking does not affect the Big Five personality factors equally in the personnel selection domain. Conscientiousness ($d = 0.20$) is the only personality dimension shown to be affected by faking but with no important practical consequences.

## Implications for Research and Practice and Limitations of the Study

The findings reported in the current study have some implications for researchers and practitioners in applied personality assessment. From a research point of view, this is the first study that provides empirical evidence of the transient error effects of faking on personality measures. Therefore, empirical research should distinguish between observed faking and true faking when this construct is operationalized. The frequently used within-subject and between-subject designs do not permit us to distinguish between and to estimate the precise amount of these two estimates of faking, and this is critical as true faking is what it is truly relevant from an applied point of view.

From a practical perspective, the results supported the idea that quasi-ipsative FC personality inventories are robust instruments that control faking effects. Therefore, we suggest practitioners use quasi-ipsative FC personality inventories in applied settings (i.e., personnel selection) because, in addition to being relevant predictors of organizational outcomes (e.g., Martínez et al., 2021a; Salgado et al., 2015; Salgado & Táuriz, 2014), they are robust against faking.

This study, like any empirical study, has some limitations that should be considered in future research. First, this study used a single quasi-ipsative FC inventory, with its peculiarities (i.e., non-algebraic dependence). In this sense, future research should expand by using other FC formats (e.g., ipsative and normative FC inventories, and quasi-ipsative FC inventories with algebraic dependence). Second, although the findings of this study regarding transient error probably generalize to SS personality inventories, future studies should examine the extent to which transient error affects the faking estimates obtained for SS personality inventories. Hence, this study should be replicated with other types of personality inventories to confirm the generalization and extension of the findings.

## Conflict of Interest

The authors of this article declare no conflict of interest.

## References


Abad, F. J., Kreitchmann, R. S., Sorrel, M. A., Nájera, P., García-Garzón, E., Garrido, L. E., & Jiménez, M. (2022). Building adaptive forced choice tests "On the Fly" for personality measurement. *Psychologist Papers, 43*(1), 29-35. https://doi.org/10.23923/pap.psicol.2982

Adair, C. (2014). *Interventions for addressing faking on personality assessments for employee selection: A meta-analysis* (Unpublished doctoral dissertation). DePaul University.

Baron, H. (1996). Strengths and limitation of ipsative measurement. *Journal of Occupational and Organizational Psychology, 69*(1), 49-56. https://doi.org/10.1111/j.2044-8325.1996.tb00599.x

Bartram, D. (1996). The relationship between ipsatized and normative measures of personality. *Journal of Occupational and Organizational Psychology, 69*(1), 25-39. https://doi.org/10.1111/j.2044-8325.1996.tb00597.x

Becker, G. (2000). How important is transient error in estimating reliability? Going beyond simulation studies. *Psychological Methods, 5*(3), 370-379. https://doi.org/10.1037/1082-989X.5.3.370

Birkeland, S. A., Manson, T. M., Kisamore, 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*(4), 317-335. https://doi.org/10.1111/j.1468-2389.2006.00354.x

Brown, A., & Maydeu-Olivares, A. (2013). How IRT can solve problems of ipsative data in forced-choice questionnaires. *Psychological Methods, 18*(1), 36-52. https://doi.org/10.1037/a0030641

Cao, M., & Drasgow, F. (2019). Does forcing reduce faking? A meta-analytic review of forced-choice personality measures in high-stakes situations. *Journal of Applied Psychology, 104*(11), 1347-1368. https://doi.org/10.1037/apl0000414

Christiansen, N. D., Burns, G. N., & Montgomery, G. E. (2005). Reconsidering forced-choice item formats for applicant personality assessment. *Human Performance, 18*(3), 267-307. https://doi.org/10.1207/s15327043hup1803_4

Clemans, W. V. (1966). An analytical and empirical examination of some properties of ipsative measures. *Psychometric Monographs, 14,* 1-56.

Cochran, W. G., & Cox, G. M. (1978). *Experimental designs.* John Wiley & Sons, Inc.

Converse, P. D., Oswald, F. L., Imus, A., Hedricks, C., Roy, R., & Butera, H. (2006). Forcing choices in personality measurement. In R. L. Griffith & M. H. Peterson (Eds.), *A closer examination of applicant faking behavior* (pp. 263-282). Information Age.

Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation: Design and analysis issues for field settings.* Houghton Mifflin.

Dilchert, S., & Ones, D. S. (2012). Measuring and improving environmental sustainability. In S. E. Jackson, D. S. Ones, & S. Dilchert (Eds.), *Managing HR for environmental sustainability* (pp. 187-221). Jossey-Bass/Wiley.

Donovan, J. J., Dwight, S. A., & Schneider, D. (2014). The impact of applicant faking on selection measures, hiring decisions, and employee performance. *Journal of Business and Psychology, 29,* 479-493. https://doi.org/10.1007/s10869-013-9318-5

Douglas, E. F., MacDaniel, M. A., & Snell, A. F. (1996). The validity of non-cognitive measures decays when applicants fake. *Academy of Management Proceedings, 1996*(1), 127-131. https://doi.org/10.5465/ambpp.1996.4979062

Dunlap, W, P., Cortina, J. M., Vaslow, J. B., & Burke, M. J. (1996). Meta-analysis of experiments with matched groups or repeated measures designs. *Psychological Methods, 1*(2), 170-177. https://doi.org/10.1037/1082-989X.1.2.170

Ellingson, J. E., Sackett, P. R., & Connelly, B. S. (2007). Personality assessment across selection and development contexts: Insights into response distortion. *Journal of Applied Psychology, 92*(2), 386-395. https://doi.org/10.1037/0021-9010.92.2.386

Fisher, P. A., Robie, C., Christiansen, N. D., Speer, A. B., & Schneider, L. (2019). Criterion-related validity of forced-choice personality measures: A cautionary note regarding Thurstonian IRT versus classical test theory scoring. *Personnel Assessment and Decisions, 5*(1), 49-61. https://doi.org/10.25035/pad.2019.01.003

Gao, L. (2005). *Latin squares in experimental design.* Unpublished manuscript. Michigan State University.

García-Izquierdo, A. L., Ramos-Villagrasa, P. J., & Lubiana, M. A. (2020). Developing biadata for public selection purposes: A comparison between fuzzy logic and traditional methods. *Journal of Work and Organizational Psychology, 36*(3) 231-242. https://doi.org/10.5093/jwop2020a22

Grant, D. A. (1948). The latin square principle in the design and analysis of psychological experiments. *Psychological Bulletin, 45*(5), 427-442. https://doi.org/10.1037/h0053912

Griffith, R. L., Chmielowski, T., & Yoshita, Y. (2007). Do applicants fake? An examination of the frequency of applicant faking behavior. *Personnel Review, 36*(3), 341-355. https://doi.org/10.1108/00483480710731310

Griffith, R. L., & Converse, P. D. (2012). The rules of evidence and the prevalence of applicant faking. In M. Ziegler, C. MacCann, & R. D. Roberts (Eds.), *New perspectives on faking in personality assessment* (pp. 34-52). Oxford University Press.

Heggestad, E. D., George, E., & Reeve, C. L. (2006). Transient error in personality scores: Considering honest and faked responses. *Personality and Individual Differences, 40*(6), 1201-1211. https://doi.org/10.1016/j.paid.2005.10.014

Hicks, L. E. (1970). Some properties of ipsative, normative, and forced-choice normative measures. *Psychological Bulletin, 74*(3), 167-184. https://doi.org/10.1037/h0029780

Hooper, A. C. (2007). *Self-presentation on personality measures in lab and field settings: A meta-analysis* (Unpublished doctoral dissertation). University of Minnesota.

Horn, J. L. (1971). Motivation and dynamic calculus concepts from multivariate experiment. In R. B. Cattell (Ed.), *Handbook of multivariate experimental psychology* (2nd printing, pp. 611-641). Tand McNally.

Hough, L. M., Eaton, N. K., Dunnette, M. D., Kamp, J. D., & McCloy, R. A. (1990). Criterion-related validities of personality constructs and the effect of response distortion on those validities. *Journal of Applied Psychology, 75*(5), 581-595. https://doi.org/10.1037/0021-9010.75.5.581

Jackson, D. N., Wroblewski, V. R., & Ashton, M. C. (2000). The impact of faking on employment tests: Does forced choice offer a solution? *Human Performance, 13*(4), 371-388. https://doi.org/10.1207/S15327043HUP1304_3

Kirk, R. E. (2013). *Experimental design: Procedures for the behavioral sciences* (4th Edition). Sage.

Komar, S., Brown, D. J., Komar, J. A., & Robie, C. (2008). Faking and the validity of conscientiousness: A Monte Carlo investigation. *Journal of Applied Psychology, 93*(1), 140-154. https://doi.org/10.1037/0021-9010.93.1.140

MacCann, C., Ziegler, M., & Roberts, R. D. (2012). Faking in personality assessment: Reflections and recommendations. In M. Ziegler, C. MacCann, & R. D. Roberts (Eds.), *New perspectives on faking in personality assessment* (pp. 309-329). Oxford University Press.

Martínez, A. (2019). *Evaluación empírica de un modelo teórico de los efectos del faking sobre las medidas de personalidad ocupacional* [Empirical assessment of a theoretical model of the effects of faking on the scores of occupational personality] (Unpublished doctoral dissertation). University of Santiago de Compostela.

Martínez, A., Moscoso, S., & Lado, M. (2021a). Effects of faking on the predictive validity of a quasi-ipsative forced-choice personality inventory: Implications for sustainable personnel selection. *Sustainability, 13*(8), 4398. https://doi.org/10.3390/su13084398

Martínez, A., Moscoso, S., & Lado, M. (2021b). Faking effects on the factor structure of a quasi-ipsative forced-choice personality inventory. *Journal of Work and Organizational Psychology, 37*(1), 1-10. https://doi.org/jwop2021a7

Martínez, A., & Salgado, J. F. (2021). A meta-analysis of the faking resistance of forced-choice personality inventories. *Frontiers in Psychology, 12,* Article 732241. https://doi.org/10.3389/fpsyg.2021.732241

Meade, A. W. (2004). Psychometric problems and issues involved with creating and using ipsative measures for selection. *Journal of Occupational and Organizational Psychology, 77*(4), 531-552. https://doi.org/10.1348/0963179042596504

Mesmer-Magnus, J., & Viswesvaran, C. (2006). Assessing response distortion in personality tests: A review pf research designs and analytic strategies. In R. L. Griffith & M. H. Peterson (Eds.), *A closer examination of applicant faking behavior* (pp. 85-114). Information Age.

Morgeson, F. P., Campion, M. A., Dipboye, R. L., Hollenbeck, J. R., Murphy, K., & Schmitt, N. (2007). Reconsidering the use of personality tests in personnel selection contexts. *Personnel Psychology, 60*(3), 683-729. https://doi.org/10.1111/j.1744-6570.2007.00089.x

Morillo, D., Abad, F. J., Kreitchmann, R. S., Leenen, I., Hontangas, P., & Ponsoda, V. (2019). The journey from Likert to forced-choice questionnaires: Evidence of the invariance of item parameters. *Journal of Work and Organizational Psychology, 35*(2), 75-83. https://doi.org/10.5093/jwop2019a11

Murphy, K. R. (2005). Why don't measures of broad dimensions of personality perform better as predictors of job performance? *Human Performance, 18*(4), 343-357. https://doi.org/10.1207/s15327043hup1804_2

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*(6), 660-679. https://doi.org/10.1037/0021-9010.81.6.660

Otero, I., Cuadrado, D., & Martínez, A. (2020). Convergent and predictive validity of the Big Five Factors assessed with single stimulus and quasi-ipsative questionnaires. *Journal of Work and Organizational Psychology, 36*(3), 215-222. https://doi.org/10.5093/jwop2020a17

Rothstein, M. G., & Goffin, R. D. (2006). The use of personality measures in personnel selection: What does current research support? *Human Resource Management Review, 16*(2), 155-180. https://doi.org/10.1016/j.hrmr.2006.03.004

Salgado, J. F. (2014). *Reliability, construct, and criterion validity of the Quasi-Ipsative Personality Inventory (QI5F/Tri).* Unpublished manuscript. University of Santiago de Compostela.

Salgado, J. F. (2016). A theoretical model of psychometric effects of faking on assessment procedures: Empirical findings and implications for personality at work. *International Journal of Selection and Assessment, 24*(3), 209-228. https://doi.org/10.1111/ijsa.12142

Salgado, J. F. (2017). Moderator effects of job complexity on the validity of forced-choice personality inventories for predicting job performance. *Journal of Work and Organizational psychology, 33*(3), 229-238. https://doi.org/10.1016/j.rpto.2017.07.001

Salgado, J. F., Anderson, N., & Táuriz, 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 Psychology, 88*(4), 797-834. https://doi.org/10.1111/joop.12098

Salgado, J. F., & Lado, M. (2018). Faking resistance of a quasi-ipsative forced-choice personality inventory without algebraic dependence. *Journal of Work and Organizational Psychology, 34*(3), 213-216. https://doi.org/10.5093/jwop2018a23

Salgado, J. F., & Táuriz, 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*(1), 3-30. https://doi.org/10.1080/1359432X.2012.716198

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*(2), 206-224. https://doi.org/10.1037/1082-989X.8.2.206

Smith, D. B., & McDaniel, M. (2012). Questioning old assumptions: Faking and the personality-performance relationship. In M. Ziegler, C. MacCann, & R. D. Roberts (Eds.), *New perspectives on faking in personality assessment* (pp. 53-70). Oxford University Press.

Vautier, S., & Jmel, S. (2003), Transient error or specificity? An alternative to the staggered equivalent split-half procedure. *Psychological Methods, 8*(2), 225-238. https://doi.org/10.1037/1082-989X.8.2.225

Viswesvaran, C., & Ones, D. S. (1999). Meta-analyses of fakability estimates: Implications for personality measurement. *Educational and Psychological Measurement, 59*(2), 197-210. https://doi.org/10.1177/00131649921969802

Zickar, M. J., & Gibby, R. E. (2006). A history of faking and socially desirable responding on personality tests. In R. L. Griffith & M. H. Peterson (Eds.), *A closer examination of applicant faking behavior* (pp. 21-42). Information Age.

Ziegler, M., MacCann, C., & Roberts, R. D. (2012). Faking: knowns, unknowns, and points of contention. In M. Ziegler, C. MacCann, & R. D. Roberts (Eds.), *New perspectives on faking in personality assessment* (pp. 3-16). Oxford University Press.