



Study Protocol on Intentional Distortion in Personality Assessment: Relationship with Test Format, Culture, and Cognitive Ability

Eline Van Geert¹, Altan Orhon², Iulia A. Cioca³, Rui Mamede⁴, Slobodan Golušin⁵, Barbora Hubená⁶ and Daniel Morillo^{7*}

¹ Faculty of Psychology and Educational Sciences, KU Leuven (University of Leuven), Leuven, Belgium, ² Department of Psychology, Istanbul Bilgi University, Istanbul, Turkey, ³ ScienceForWork, Milan, Italy, ⁴ Formerly affiliated with Faculty of Psychology and Education Sciences, University of Coimbra, Coimbra, Portugal, ⁵ Faculty of Philosophy, University of Novi Sad, Novi Sad, Serbia, ⁶ Department of Psychology, Faculty of Arts, Masaryk University, Brno, Czech Republic, ⁷ Chair of Psychometric Models and Applications, Department of Social Psychology and Methodology, Faculty of Psychology, Autonomous University of Madrid, Madrid, Spain

OPEN ACCESS

Edited by:

Gabriela Diana Roman,
University of Cambridge, UK

Reviewed by:

Daniel Saverio John Costa,
University of Sydney, Australia
Keith M. Harris,
University of Queensland, Australia

*Correspondence:

Daniel Morillo
daniel.morillo@uam.es

Specialty section:

This article was submitted to
Quantitative Psychology
and Measurement,
a section of the journal
Frontiers in Psychology

Received: 10 March 2016

Accepted: 06 June 2016

Published: 28 June 2016

Citation:

Van Geert E, Orhon A, Cioca IA, Mamede R, Golušin S, Hubená B and Morillo D (2016) Study Protocol on Intentional Distortion in Personality Assessment: Relationship with Test Format, Culture, and Cognitive Ability. *Front. Psychol.* 7:933. doi: 10.3389/fpsyg.2016.00933

Self-report personality questionnaires, traditionally offered in a graded-scale format, are widely used in high-stakes contexts such as job selection. However, job applicants may intentionally distort their answers when filling in these questionnaires, undermining the validity of the test results. Forced-choice questionnaires are allegedly more resistant to intentional distortion compared to graded-scale questionnaires, but they generate ipsative data. Ipsativity violates the assumptions of classical test theory, distorting the reliability and construct validity of the scales, and producing interdependencies among the scores. This limitation is overcome in the current study by using the recently developed Thurstonian item response theory model. As online testing in job selection contexts is increasing, the focus will be on the impact of intentional distortion on personality questionnaire data collected online. The present study intends to examine the effect of three different variables on intentional distortion: (a) test format (graded-scale versus forced-choice); (b) culture, as data will be collected in three countries differing in their attitudes toward intentional distortion (the United Kingdom, Serbia, and Turkey); and (c) cognitive ability, as a possible predictor of the ability to choose the more desirable responses. Furthermore, we aim to integrate the findings using a comprehensive model of intentional distortion. In the Anticipated Results section, three main aspects are considered: (a) the limitations of the manipulation, theoretical approach, and analyses employed; (b) practical implications for job selection and for personality assessment in a broader sense; and (c) suggestions for further research.

Keywords: personality assessment, personnel selection, forced-choice, Thurstonian IRT, faking, ipsativity, cross-cultural comparison

Abbreviations: BFI, big five inventory; ICAR, international cognitive ability resource; IRT, item-response theory; MUPP, multi-unidimensional pairwise preference [model].

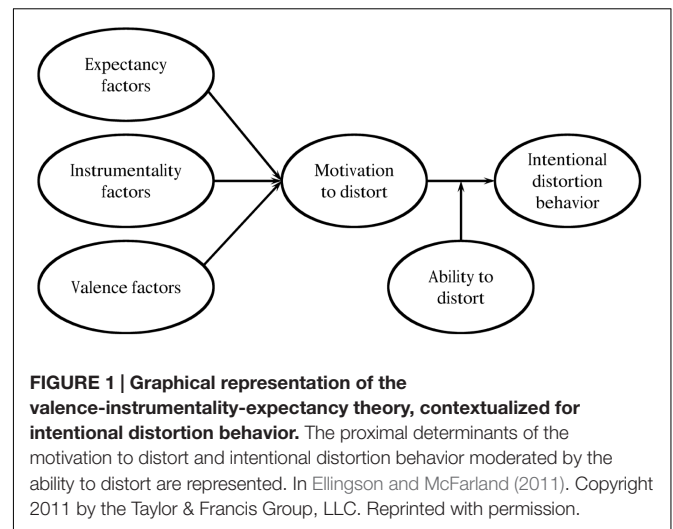
INTRODUCTION

Self-report personality questionnaires are increasingly popular in high-stakes contexts such as personnel selection (Rothstein and Goffin, 2006), college admissions (Sjöberg, 2015), and determining eligibility to stand trial (Archer et al., 2006). However, in these situations, instead of answering honestly, test takers often intentionally distort their answers to increase their chances of being hired (Birkeland et al., 2006). It has been estimated that roughly 30 percent of job applicants intentionally distort their responses (Griffith and Converse, 2011). Intentional distortion is detrimental to the psychometric properties of the assessment instrument, hiring decisions, and the utility of whole-job selection systems (Donovan et al., 2014), although human resources practitioners are largely unaware of the implications (Rothstein and Goffin, 2006). Furthermore, research on intentional distortion suffers from weak theoretical support and over-reliance on empirical and statistical methods (Griffith and Peterson, 2011).

In the literature there is considerable debate on the question whether intentional distortion also decreases the predictive validity of self-report questionnaires. Donovan et al. (2014) conducted a study utilizing a within-subjects design in an actual organizational setting and found not only a negative impact of intentional distortion on the psychometric properties of the non-cognitive self-report measure, but also a negative impact on the quality of the hiring decisions made by the organization. Additionally, people intentionally distorting their answers were found to exhibit lower levels of performance than people answering honestly. This implies that intentional distortion has negative consequences for the predictive validity of the personality test. The opposite argument, however, is based on seeing intentional distortion as a type of intelligence, mostly related to social or emotional intelligence, which consists of the ability to correctly read and interpret cues in social situations. This ability allows test takers to identify correctly the meaning of the test items and the desirable characteristics for the job in question, and later on will also help them perform better at their job, especially if it involves social interactions (Kleinmann et al., 2011). Thus, in this view, the influence of intentional distortion on the personality test leads to an equal or increased predictive validity of the test.

Even when intentional distortion would lead to a better predictive or criterion-related validity of personality tests, it is also important to consider the construct validity of the test. If the test does not measure what it is expected to measure, in this case personality factors, then the construct validity is low. Understanding and reducing the influence of intentional distortion on these measures of personality should lead toward an ideal situation in which a personality test assesses personality (and not intentional distortion), and another test assesses intentional distortion or a related ability, if this variable would have predictive validity for job performance (Kleinmann et al., 2011).

The most comprehensive theoretical model of intentional distortion (see **Figure 1**; Ellingson and McFarland, 2011) is based on the valence-instrumentality-expectancy theory of motivation



(Vroom, 1964). This model states that the predictors of intentional distortion work through three proximal motivational factors: (a) valence, the affective reaction an individual has to a particular outcome of an action; (b) instrumentality, the belief that the action will lead to a particular outcome; and (c) expectancy, the belief that one can perform the action. These three factors determine a person's motivation to engage in intentional distortion; however, the individual's actual ability to enact intentional distortion moderates the effect of the motivation to do so (Ellingson and McFarland, 2011).

Situational characteristics such as test format may offset individuals' capacities for intentional distortion. Forced-choice response formats have been proposed as a way of controlling for intentional distortion in personality assessments (Christiansen et al., 2005). In forced-choice questionnaires, instead of rating items on a graded scale, respondents rank groups of personality statements that seem equivalent in terms of social desirability. Forced-choice questionnaires hinder the identification of advantageous response patterns (Stark et al., 2014), rule out uniform biases such as acquiescence and extreme responding, and are recommended for use in cross-cultural comparisons and high-stakes situations (He et al., 2014). On the other hand, another type of scale format, that of dichotomous answers (yes/no) is rarely used (e.g., Eysenck Personality Questionnaire, Eysenck and Eysenck, 1975), being advantageous in terms of time, it takes to complete the test. However, problems arise with extremely unbalanced response distributions (e.g., everyone answers "yes" to a certain item; Clark and Watson, 1995) which indicates intentional distortion, and the measurement of continuous personality variables through completely polarized items, which minimizes the information obtained for those with real scores in extremities of the distribution (Furr, 2011).

Despite their advantages, forced-choice questionnaires have traditionally been discounted due to problems arising from conventional approaches to scoring them, which produce ipsative scores. These are able to show the relative levels of a trait within an individual (e.g., an individual appears more agreeable than

conscientious), but they cannot be used to compare absolute levels between individuals (Christiansen et al., 2005). An increase on one dimension in an ipsative measurement necessitates a corresponding decrease on other dimensions. This property makes ipsative measures incompatible with methods such as correlation or factor analysis (Cornwell and Dunlap, 1994) and creates issues relating to construct validity, criterion-related validity, and reliability estimates (Brown and Maydeu-Olivares, 2013). Hicks (1970, p. 181) cautioned researchers against using purely ipsative instruments, writing, “[researchers] cannot legitimately manipulate the variables assessed by the test utilizing statistical procedures which assume that independent dimensions are involved.” Methods proposed to address this issue have included increasing the number of dimensions being measured (Hicks, 1970) and compositional data analysis (Aitchison and Egozcue, 2005), yet the relative nature of the inferences remained unresolved (van Eijnatten et al., 2015). However, recent models based on IRT allow the extraction of normative scores from forced-choice responses (Stark et al., 2014; Joubert et al., 2015). Among these, the two state-of-the-art models are the Thurstonian IRT model (Brown and Maydeu-Olivares, 2011) and multi-unidimensional pairwise preference model (MUPP; Stark et al., 2005). These models overcome the problems posed by scoring ipsative measures via classical test theory by explicitly proposing a measurement model, that describes the relationship between items and traits, and a decision model, that describes how the individual selects one item over another (Brown, 2016). This additional level of information is what allows the recovery of normative scores from a forced-choice instrument – on Thurstonian IRT, a structure of correlated latent factors derived from multiple blocks of forced-choice items is fitted to binary outcomes of pairwise comparisons (Brown and Maydeu-Olivares, 2013), whereas MUPP combines multidimensional items with unidimensional pairings and a Bayes modal procedure as means of estimating trait scores (Stark et al., 2005).

The aim of our study is twofold: (a) to present an integrated view of intentional distortion formulated on sound theoretical underpinnings and (b) to reduce the effects of intentional distortion on personality assessment in high-stakes contexts by testing a viable method of scoring forced-choice questionnaires that can overcome earlier difficulties in their use (i.e., the ipsativity problem). Along with this, we will investigate three variables previously found to affect intentional distortion and present the theoretical reasoning behind these hypothesized effects.

First, responses for forced-choice questionnaires should exhibit lower levels of intentional distortion than those for graded-scale questionnaires. Besides the effects of forced-choice format on the ability to distort discussed above (i.e., more difficult identification of advantageous response patterns), having to choose between statements with similar levels of social desirability could induce higher levels of test-taking anxiety in applicants (Converse et al., 2008), lowering cognitive performance and ability to distort. According to Converse et al. (2008), this may come from a perception that in forced-choice format they do not have free choice of answers as well as less opportunity to express their personality qualities related to the

job. Additionally, the forced-choice format could diminish their expectancy beliefs about intentional distortion of their answers (Ellingson and McFarland, 2011).

Second, attitudes toward the appropriateness of a candidate’s use of intentional distortion are associated with several cultural dimensions suggested by the GLOBE study (House et al., 2004), such as uncertainty avoidance (which decreases the appropriateness due to lack of control about the result), or power distance (enhancing the appropriateness due to perceived lack of fairness in societies high in power distance; Fell et al., 2015). These attitudes may act on intentional distortion through (a) valence beliefs, by informing personal standards of behavior, or (b) instrumentality beliefs, by leading to the belief in a more positive outcome of intentional distortion (Ellingson and McFarland, 2011).

Third, because forced-choice questionnaires are more cognitively demanding compared to graded-scale questionnaires (Converse et al., 2008), intentional distortion is expected to relate more strongly to cognitive ability in forced-choice questionnaires than in graded-scale questionnaires. Cognitive ability is on one hand expected to relate positively to the ability of applicants to distort their answers (Christiansen et al., 2005), as it is assumed that more cognitively able applicants will be better able to identify advantageous response patterns in relation to the job requirements. On the other hand, there has also been evidence showing that people with higher cognitive ability distort their answers less often (Austin et al., 2002; Levashina et al., 2009) and do not respond in a more socially desirable manner than other participants (Ones et al., 1996). Reasons for avoiding intentional distortion of their answers include high self-efficacy and believing in one’s own abilities to succeed in assessment without distorting (De Fruyt et al., 2006), short-term outcomes (such as being excluded from the applicants pool for failing social desirability items), or long-term outcomes (such as not being suitable for the role or not fitting into the working team). However, if this would be the case, this relationship would also be evident in the graded-scale questionnaires.

Consequently, our research questions are as follows:

1. Is intentional distortion lower in forced-choice questionnaires than in graded-scale questionnaires?
2. Are there differences in levels of intentional distortion across cultural groups?
3. Do people with higher cognitive ability show more intentional distortion in forced-choice questionnaires than people with lower cognitive ability?

MATERIALS AND EQUIPMENTS

Measures

Big Five Inventory

The BFI is a popular instrument for international studies and it is recommended for use in cross-cultural settings (Schmitt et al., 2007). This 44-item graded-scale-format operationalization (Pervin and John, 1999) of the Big Five theory (John et al., 2008) will be used to assess personality traits. Adaptations of the BFI to

the languages of the target countries, as well as country-specific psychometric properties, are available (Schmitt et al., 2007; Neşe Alkan, “Reliability and Validity of the Turkish Version of the Big Five Inventory,” unpublished manuscript, 2006).

Tailored Forced-Choice Questionnaires

Equivalent forced-choice questionnaires for each country will be constructed by pairing positively keyed items measuring personality traits from the International Personality Item Pool (Goldberg, 1999). Each Big Five trait is represented by 30 items that were selected to reflect the diversity of their respective domains.

In order to ensure that the items being paired to form the blocks in the forced-choice questionnaire are as closely matched in social desirability as possible, we are conducting a study to gauge social desirability levels for each item. Approximately 250 respondents (as in Stark et al., 2005) in each country will be asked to rate the items for their attractiveness. Given that social desirability may be a context-dependent property (Rothstein and Goffin, 2000), the participants will be presented with the job description of the high-stakes condition and prompted to rate the social desirability “as if” applying for that job. Next, the items will be paired using a procedure that (a) generates a list of possible pairs of items on different dimensions of the Big Five, numbering 10 pairs initially; (b) sorts the items from most desirable to least desirable, according to mean ratings; (c) identifies the two items whose means are closest; (d) removes the pair constituted by the two items from the search space; and (e) repeats the process of pairing the closest items until no more pairs remain in the search space, after which the procedure enters the next round of matching. Pairing 150 items in this manner requires eight rounds.

International Cognitive Ability Resource

We will use the 16-item ICAR Sample Test (The International Cognitive Ability Resource Team, 2014) to measure cognitive ability. This instrument, designed for online administration (Condon and Revelle, 2014), is a public-domain measure with four subscales: Letter-Number Series, Matrix Reasoning, 3D Rotation, and Verbal Reasoning. The test has been adapted for use in the native languages of the countries in this study. (Scores will be used for within-culture comparisons only.) Condon and Revelle (2014) report adequate internal consistency for the ICAR Sample Test (Cronbach’s $\alpha = 0.81$, total $\omega = 0.83$) and provide evidence of adequate convergent validity with several widely accepted measures of cognitive ability.

STEPWISE PROCEDURES

Participants

Data will be collected from university students or recent graduates in their early adulthood (aged 18–30) in three countries: the United Kingdom, Serbia, and Turkey. Approximately 250 participants from each country will take part in the study to construct the tailored forced-choice questionnaires and 500 participants from each country will take part in the experimental study. They will be recruited online

(mostly resorting to social media, e.g., Facebook, Twitter), by using university resources (i.e., using online subject pool programs or by administering them to students during or after classes), and by advertising the study in university facilities. In order to maximize participation, the advertisements will be timed to avoid periods that would be associated with decreased participation. The participants of the experimental study will be motivated by the opportunity to enter a raffle for financial prizes and the opportunity to get individual feedback on their personality.

The targeted group of participants are students and graduates in their early adulthood according to Erikson’s (1993) stage of human development. This stage is, besides completing formation adult identity and establishing intimate relationships, typical of finishing one’s education and entering the job market. University students and fresh graduates are likely to be familiar with the situation of applying for jobs, going through job interviews and assessment, including personality assessment. Moreover, the role of assistant manager which is used to introduce the high-stakes condition might be quite realistic and relatively attractive job for wide range of university students and fresh graduates of different specializations with limited work experience (Kleinmann and Klehe, 2011).

Participating countries were chosen based on differences in attitudes toward intentional distortion in job interviews (Fell et al., 2015), which were related to the cultural dimensions (e.g., power distance, in-group collectivism) assessed by the international GLOBE study (House et al., 2004). Our choices are representative of presumed minimum, intermediate, and maximum levels on this attitude index (the United Kingdom, Serbia, and Turkey, respectively), on which a higher score represents a more positive attitude toward intentional distortion. Although Serbia was not included in the GLOBE study, later research provided information on the cultural dimensions in question (Vukonjanski et al., 2012).

Ethics Statement

The study has been given full clearance by the Ethics Committee of Universidad Autónoma de Madrid, which abides law 14/2007 of July 3, 2007 regarding biomedical research, and is fully compliant with the Declaration of Helsinki.

Design and Procedure

Participants will be invited to take a set of online tests in a single session. The tests will be administered via the Qualtrics platform (Qualtrics, Provo, UT, USA). The set includes two self-report questionnaires (graded-scale and forced-choice format), each administered in two conditions (high-stakes and low-stakes), and a test of cognitive ability. In the low-stakes condition, participants will be instructed to respond as sincerely as possible. In the high-stakes condition, they will be instructed to answer as if they were applying for a job—in this case, a management trainee position. Every participant will go through both the high-stakes and the low-stakes condition, with order determined by random assignment (see **Figure 2**). The within-subject design is recommended for studying intentional distortion because it accounts for individual tendencies in the behavior (Viswesvaran

and Ones, 1999). Between the two conditions, respondents will answer a cognitive ability measure, which should have the additional benefit of reducing practice or memory effects for the questionnaires (Grieve and de Groot, 2011). Finally, respondents will answer a single item asking to what extent the described job is attractive for individual participants on a four-point scale (from very unattractive to very attractive). This will allow us to operationalize job attractiveness and possibly control for it. In return for participation, respondents who complete the whole questionnaire will have the possibility to participate in a raffle containing several monetary reward. The participants will also be offered personalized feedback based on the BFI scores in the low-stakes condition which should increase the respondents' motivation to answer the questionnaire in an accurate and honest manner under this instruction.

Proposed Analysis

The Thurstonian IRT model (Brown and Maydeu-Olivares, 2011) has proved to be a flexible, robust model for obtaining normative trait scores from individual responses to forced-choice questionnaires. In contrast to the MUPP (Stark et al., 2005), it does not require precalibration of the item parameters. It can be estimated using the widespread software Mplus (Muthén and Muthén, 2015) and thus does not require any specialized software. Finally, the International Personality Item Pool items that will be used in the forced-choice questionnaires are written in a way that assumes a dominance response model, in which an individual is more likely to answer positively to items assessing traits they are high on; as such, these items are better fit by the Thurstonian IRT than an unfolding model such as the MUPP (Brown and Maydeu-Olivares, 2010). Therefore, Thurstonian IRT is the model of choice to analyze the ipsative data.

This model is based on Thurstone's (1927) Law of Comparative Judgement. It links the utility of each response option to the latent trait it intends to measure, by means of a linear function (Brown and Maydeu-Olivares, 2011). As a result, the probability that a respondent chooses item *i* in a binary comparison between items *i* and *k* in block *l* is expressed by (p. 473),

$$P(Y_l = 1 | \eta_a, \eta_b) = \Phi \left(\frac{-\gamma_l + \lambda_i \eta_a + \lambda_k \eta_b}{\sqrt{\psi_i^2 + \psi_k^2}} \right),$$

where $\Phi_{(x)}$ is the cumulative standard normal distribution function at *x*, γ_l is the threshold for the binary comparison of the two items block *l*, λ_i and λ_k are the factor loadings of items *i* and *k* on their respective factors *a* and *b*, ψ_i^2 and ψ_k^2 the unique variances of items *i* and *k*, and η_a and η_b a respondent's scores in factors *a* and *b*, respectively. By fitting a confirmatory factor-analytic model to the data, item and population parameters can be estimated for the model. Then, normative person parameters can be obtained through a maximum *a posteriori* estimator. Brown and Maydeu-Olivares (2012) provide and document an Excel macro that can be used to generate the necessary input files to Mplus for a given forced-choice questionnaire – the

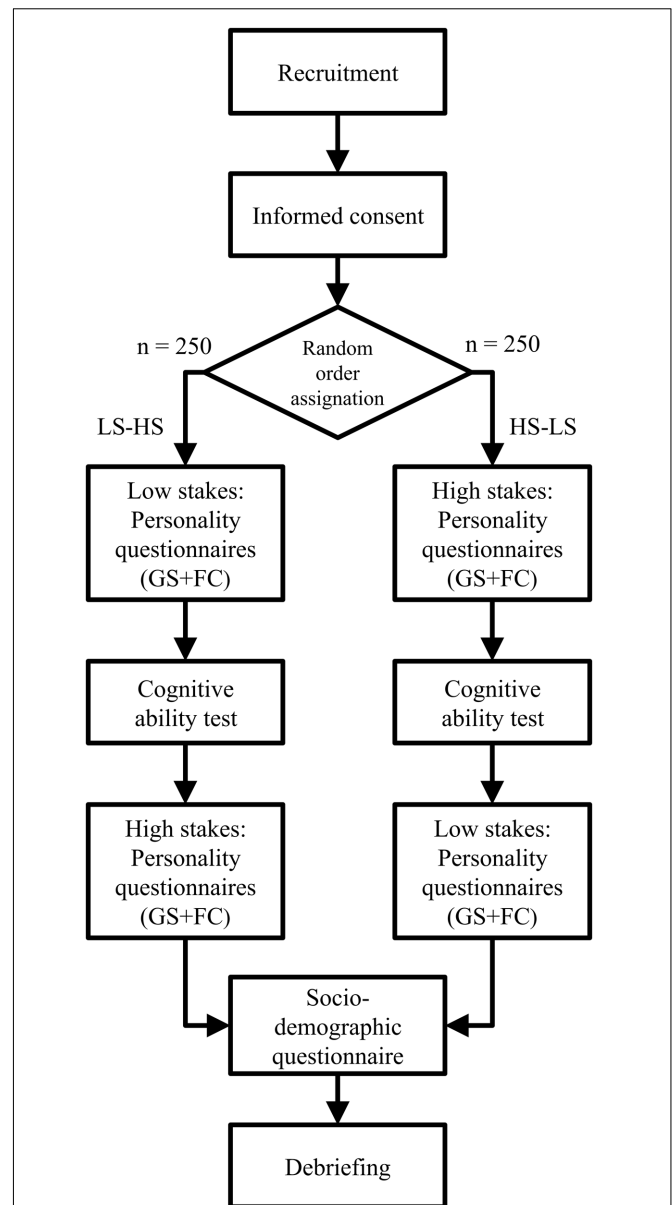


FIGURE 2 | Research design flowchart. LS, low stakes; HS, high stakes; GS, graded scale; FC, forced choice.

output allows estimation and scoring according to a Thurstonian IRT model that fits the data, computing item loadings, item thresholds, and factor scores.

The Thurstonian IRT model will be integrated into a wider structural equation model, where the responses to the forced-choice questionnaire and the graded-scale questionnaire will be jointly modeled. The bivariate information from the low-stakes condition will then be used to fit the structural equation model, and an invariance analysis will be conducted to check for invariance of the two order conditions. Then, a multitrait-multimethod matrix will be assessed for construct, convergent, and discriminant validity. After that, *maximum a posteriori*

scores for the respondents' latent traits in both conditions will be obtained using Mplus (Brown and Maydeu-Olivares, 2012).

Two intentional distortion scores will be obtained for each respondent by subtracting the IRT-based trait scores in the low-stakes (reference score) from those in the high-stakes condition: one concerning each test format (graded-scale versus forced-choice). To answer the first research question, those intentional distortion scores will be tested for significant differences using Rao's *F*-test (Christiansen et al., 2005). To test the second research question, intentional distortion scores of the three cultural samples will be tested for differences across country groups using an analysis of variance test (Converse et al., 2010). Finally, a linear regression analysis will be conducted between intentional distortion scores and cognitive ability scores to answer the third research question.

ANTICIPATED RESULTS

The present study intends to clarify the influence of test format, culture, and cognitive ability on intentional distortion in self-report personality measures. Hypotheses made concerning the influence of test format, culture, and cognitive ability are based on and integrated in the theoretical model of intentional distortion by Ellingson and McFarland (2011). However, the proposed underlying processes are still to be tested in further research.

Firstly, tests that use a forced-choice item format have been proposed to reduce the effects of respondents' intentional distortion on the test results. However, they have proved to be impractical up to now because forced-choice questionnaire items generate ipsative data. By using an IRT-based data analysis, the present study aims to increase the applicability of the forced-choice tests, and provide a valuable alternative for practitioners to reduce the effects of intentional distortion in personality assessment. As forced-choice format makes it more difficult to identify advantageous response patterns (Stark et al., 2014) and might also decrease expectancy beliefs (i.e., belief in ability to successfully distort), it is expected that intentional distortion will be lower in forced-choice questionnaires than in graded-scale questionnaires.

The results of our study regarding the test format will be of practical relevance for the assessment in high-stakes situations, such as personnel selection, where important decisions are made based on candidates' scores on personality tests. Future research could explore the utility of the assessment method for other high-stakes contexts, such as establishing eligibility for trial. In the long term, this will enable a more accurate and fairer assessment of participants in high-stakes contexts.

Secondly, it is expected that cultures differ in the extent of intentional distortion they display. More specifically, it is expected that participants from cultures scoring low, medium, or high on the index of positive attitude toward intentional distortion (the United Kingdom, Serbia, and Turkey, respectively), will show, respectively, low, medium, and high levels of intentional distortion. This influence of culture on intentional distortion may act through valence beliefs (i.e., informing personal attitude toward intentional distortion) and

instrumentality beliefs (i.e., affecting belief that intentional distortion will lead to positive outcomes; Ellingson and McFarland, 2011).

Cross-national work-related mobility is increasing nowadays, and likewise with the reach of multinational enterprises. Practitioners conducting personality assessment in such cross-national contexts need to understand the differences in their respondents' tendencies to complete personality tests in certain ways. By investigating the phenomenon of intentional distortion in three countries that differ in their attitude toward this practice, the present study will have further implications for international assessment.

Thirdly, we will also explore the relationship between a person's general cognitive ability and intentional distortion, both on graded-scale and forced-choice items. In graded-scale questionnaires, no influence of cognitive ability on intentional distortion is expected. In forced-choice questionnaires, a positive relation of cognitive ability and the ability to distort is hypothesized, as it is expected that participants should be more able to identify the advantageous response patterns. Moreover, cognitive ability might also reinforce a person's motivation to distort by raising their expectancy beliefs of how successful they will be at distorting their answers.

Nevertheless, a potential rejection of this hypothesis could indicate support for an alternative explanation. Participant's cognitive ability can be negatively related to their motivation to distort as more cognitively able applicants would be more aware of possible short-term consequences (such as being excluded from the applicants pool for failing social desirability items), or long-term consequences (such as not being suitable for the role or not fitting into working team) of distorting answers in high-stakes contexts. Yet another reason for choosing not to distort in participants with high cognitive skills is associated with higher self-efficacy and belief that they can score high without distorting (Levashina et al., 2009), so their expectancy belief may be that distorting is not worth the effort and risk-taking. However, because of the simulated nature of the high-stakes manipulation, the motivational processes to distort may differ from those in an actual high-stakes situation, for example because the long-term consequences are less taken into consideration, which threatens the ecological validity of the results. Simulating the high-stakes situations is a common practice in this field of research (see, e.g., Christiansen et al., 2005), but future studies with real job applicants would be recommended to validate our findings and their applicability in real-life situations. Additionally, although the nature of the specific instruction set given in the high-stakes context (i.e., "respond as if applying for a job") was chosen to be as ecologically valid as possible in a simulated context, this instruction set does not distinguish between the short- and long-term consequences possibly influencing the motivation to distort, thereby compromising internal and external validity. To disentangle both motivations, further studies could include an additional high-stakes condition focusing on short-term consequences specifically (e.g., "respond so as to maximize your chances of getting hired").

Understanding how cognitive ability and intentional distortion relate in the context of assessment is important to clarify aspects of predictive and construct validity of personality tests. Although a high predictive validity is useful in practice, it is essential to understand what the test actually measures. We have tried to achieve this by anchoring the study design in a solid theoretical framework that not only contributes to explaining the interrelations between concepts but also can guide future research to build a deeper and more comprehensive understanding of intentional distortion.

Limitations of our experimental design include the use of student groups as representative populations, lack of control over the physical testing environment and sample equivalence, and the possibility of a high rate of attrition leading to less diversity in sample. However, we try to mitigate the effect of the first aspect by advertising the study to recent graduates and students in final years, who are confronting (or will soon be confronting) the challenge of obtaining their first job. Regarding control over the physical environment, online assessment is an increasingly common practice, with 81% of the companies that use assessment administering it online (Kantrowitz, 2014), despite its potential disadvantages. Furthermore, it appears that online tests and pen-and-paper versions are roughly equal in their susceptibility to intentional distortion (Grieve and de Groot, 2011); therefore, research on intentional distortion in online assessment is still needed. Weigold et al. (2013) describe two studies comparing results for surveys administered via traditional means (e.g., on paper and in lab settings) and surveys administered either online or in a take-home format. The instruments used in these studies purportedly measured personality and social desirability. The authors reported that paper-and-pencil and

online survey administration were generally equivalent except for some auxiliary aspects such as response rates and completion time. However, Joinson (1999) described an effect whereby participants reported lower social anxiety and social desirability influence in an online survey compared to a paper-based survey, and when they were anonymous compared to being identified. In the case of the present study, it is expected that most participants will provide some personally identifying information in the course of enrolling for the raffle. The present study attempts to reproduce the conditions of high-stakes assessment in a job selection context. Having participants identify themselves matches more closely the conditions of real-life job selection, and a hypothetical increase in susceptibility to social desirability likewise matches what we intend to study. Because of this, our choice of methodology might be more appropriate for drawing conclusions for this type of assessment.

AUTHOR CONTRIBUTIONS

All authors listed, have made substantial, direct and intellectual contribution to the work, and approved it for publication.

ACKNOWLEDGMENTS

This research was made possible by the Junior Researcher Programme (<http://jrp.psycholars.org/>). We would like to thank everyone involved in the organization of the Programme for their assistance.

REFERENCES

- Aitchison, J., and Egozcue, J. J. (2005). Compositional data analysis: where are we and where should we be heading? *Math. Geol.* 37, 829–850. doi: 10.1007/s11004-005-7383-7
- Archer, R. P., Buffington-Vollum, J. K., Stredny, R. V., and Handel, R. W. (2006). A survey of psychological test use patterns among forensic psychologists. *J. Pers. Assess.* 87, 84–94. doi: 10.1207/s15327752jpa8701_07
- Austin, E. J., Deary, I. J., Whiteman, M. C., Fowkes, F. G. R., Pedersen, N. L., Rabbitt, P., et al. (2002). Relationships between ability and personality: does intelligence contribute positively to personal and social adjustment? *Pers. Individ. Dif.* 32, 1391–1411. doi: 10.1016/S0191-8869(01)00129-5
- Birkeland, S. A., Manson, T. M., Kisamore, J. L., Brannick, M. T., and Smith, M. A. (2006). A meta-analytic investigation of job applicant faking on personality measures. *Int. J. Sel. Assess.* 14, 317–335. doi: 10.1111/j.1468-2389.2006.00354.x
- Brown, A. (2016). Item response models for forced-choice questionnaires: a common framework. *Psychometrika* 81, 135–160. doi: 10.1007/s11336-014-9434-9
- Brown, A., and Maydeu-Olivares, A. (2010). Issues that should not be overlooked in the dominance versus ideal point controversy. *Ind. Organ. Psychol.* 3, 489–493. doi: 10.1111/j.1754-9434.2010.01277.x
- Brown, A., and Maydeu-Olivares, A. (2011). Item response modeling of forced-choice questionnaires. *Educ. Psychol. Meas.* 71, 460–502. doi: 10.1177/0013164410375112
- Brown, A., and Maydeu-Olivares, A. (2012). Fitting a Thurstonian IRT model to forced-choice data using Mplus. *Behav. Res. Methods* 44, 1135–1147. doi: 10.3758/s13428-012-0217-x
- Brown, A., and Maydeu-Olivares, A. (2013). How IRT can solve problems of ipsative data in forced-choice questionnaires. *Psychol. Methods* 18, 36–52. doi: 10.1037/a0030641
- Christiansen, N. D., Burns, G. N., and Montgomery, G. E. (2005). Reconsidering forced-choice item formats for applicant personality assessment. *Hum. Perform.* 18, 267–307. doi: 10.1207/s15327043hup1803_4
- Clark, L. A., and Watson, D. B. (1995). Constructing validity: basic issues in scale development. *Psychol. Assess.* 7, 309–319. doi: 10.1037/1040-3590.7.3.309
- Condon, D. M., and Revelle, W. (2014). The international cognitive ability resource: development and initial validation of a public-domain measure. *Intelligence* 43, 52–64. doi: 10.1016/j.intell.2014.01.004
- Converse, P. D., Oswald, F. L., Imus, A., Hedricks, C., Roy, R., and Butera, H. (2008). Comparing personality test formats and warnings: effects on criterion-related validity and test-taker reactions. *Int. J. Sel. Assess.* 16:155. doi: 10.1111/j.1468-2389.2008.00420.x
- Converse, P. D., Pathak, J., Quist, J., Merbedone, M., Gotlib, T., and Kostic, E. (2010). Statement desirability ratings in forced-choice personality measure development: implications for reducing score inflation and providing trait-level information. *Hum. Perform.* 23, 323–342. doi: 10.1080/08959285.2010.501047
- Cornwell, J. M., and Dunlap, W. P. (1994). On the questionable soundness of factoring ipsative data: a response to Saville and Willson (1991). *J. Occup. Organ. Psychol.* 67, 89–100. doi: 10.1111/j.2044-8325.1994.tb00553.x
- De Fruyt, F., Aluja, A., Garcia, L. F., Rolland, J. P., and Jung, S. C. (2006). Positive presentation management and intelligence and the personality differentiation by intelligence hypothesis in job applicants. *Int. J. Sel. Assess.* 14, 101–112. doi: 10.1111/j.1468-2389.2006.00337.x

- Donovan, J. J., Dwight, S. A., and Schneider, D. (2014). The impact of applicant faking on selection measures, hiring decisions, and employee performance. *J. Bus. Psychol.* 29, 479–493. doi: 10.1007/s10869-013-9318-5
- Ellingson, J. E., and McFarland, L. A. (2011). Understanding faking behavior through the lens of motivation: an application of VIE theory. *Hum. Perform.* 24, 322–337. doi: 10.1080/08959285.2011.597477
- Erikson, E. H. (1993). *Childhood and Society*. New York, NY: W. W. Norton and Company.
- Eysenck, H. J., and Eysenck, S. B. G. (1975). *Manual of the Eysenck Personality Questionnaire (Junior and Adult)*. Kent: Hodder and Stoughton.
- Fell, C. B., König, C. J., and Kammerhoff, J. (2015). Cross-cultural differences in the attitude toward applicants' faking in job interviews. *J. Bus. Psychol.* 31, 65–85. doi: 10.1007/s10869-015-9407-8
- Furr, R. M. (2011). *Scale Construction and Psychometrics for Social and Personality Psychology*. London: Sage Publications.
- Goldberg, L. R. (1999). "A broad-bandwidth, public domain, personality inventory measuring the lower-level facets of several five-factor models," in *Personality Psychology in Europe*, eds I. Mervielde, I. Deary, F. de Fruyt, and F. Ostendorf (Tilburg: Tilburg University Press), 7–28.
- Grieve, R., and de Groot, H. T. (2011). Does online psychological test administration facilitate faking? *Comput. Hum. Behav.* 27, 2386–2391. doi: 10.1016/j.chb.2011.08.001
- Griffith, R. L., and Converse, P. D. (2011). "The rules of evidence and the prevalence of applicant faking," in *New perspectives on faking in personality assessments*, eds M. Ziegler, C. McCann, and R. Roberts (Oxford: Oxford University Press).
- Griffith, R. L., and Peterson, M. H. (2011). One piece at a time: the puzzle of applicant faking and a call for theory. *Hum. Perform.* 24, 291–301. doi: 10.1080/08959285.2011.597474
- He, J., Bartram, D., Inceoglu, I., and van de Vijver, F. J. R. (2014). Response styles and personality traits: a multilevel analysis. *J. Cross. Cult. Psychol.* 45, 1028–1045. doi: 10.1177/0022022114534773
- Hicks, L. E. (1970). Some properties of ipsative, normative, and forced-choice normative measures. *Psychol. Bull.* 74, 167–184. doi: 10.1037/h0029780
- House, R. J., Hanges, P. J., Javidan, M., Dorfman, P. W., and Gupta, V. (eds). (2004). *Culture, Leadership, and Organizations: The GLOBE Study of 62 Societies*. Thousand Oaks, CA: Sage Publications.
- John, O. P., Naumann, L. P., and Soto, C. J. (2008). "Paradigm shift to the integrative big-five trait taxonomy: history, measurement, and conceptual issues," in *Handbook of Personality: Theory and Research*, eds O. P. John, R. W. Robins, and L. A. Pervin (New York, NY: Guilford Press), 114–158.
- Joinson, A. (1999). Social desirability, anonymity, and internet-based questionnaires. *Behav. Res. Methods Instrum. Comput.* 31, 433–438. doi: 10.3758/BF03200723
- Joubert, T., Inceoglu, I., Bartram, D., Dowdeswell, K., and Lin, Y. (2015). A comparison of the psychometric properties of the forced choice and Likert scale versions of a personality instrument. *Int. J. Sel. Assess.* 23, 92–97. doi: 10.1111/ijsa.12098
- Kantrowitz, T. M. (2014). *2014 Global Assessment Trends Report*. Arlington, VA: CEB.
- Kleinmann, M., Ingold, P. V., Lievens, F., Jansen, A., Melchers, K. G., and König, C. J. (2011). A different look at why selection procedures work: the role of candidates' ability to identify criteria. *Organ. Psychol. Rev.* 1, 128–146. doi: 10.1177/2041386610387000
- Kleinmann, M., and Klehe, U. C. (2011). Selling oneself: construct and criterion-related validity of impression management in structured interviews. *Hum. Perform.* 24, 29–46. doi: 10.1080/08959285.2010.530634
- Levashina, J., Morgeson, F. P., and Campion, M. A. (2009). They don't do it often, but they do it well: exploring the relationship between applicant mental abilities. *Int. J. Select. Assess.* 17, 271–281. doi: 10.1111/j.1468-2389.2009.00469.x
- Muthén, L. K., and Muthén, B. O. (2015). *Mplus User's Guide*, 7th Edn. Los Angeles, CA: Muthén and Muthén.
- Ones, D. S., Viswesvaran, C., and Reiss, A. D. (1996). Role of social desirability in personality testing for personnel selection: the red herring. *J. Appl. Psychol.* 81, 660–679. doi: 10.1037/0021-9010.81.6.660
- Pervin, L. A., and John, O. P. (eds) (1999). *Handbook of Personality: Theory and Research*, 2nd Edn. New York, NY: Guilford Press.
- Rothstein, M. G., and Goffin, R. D. (2000). "The assessment of personality constructs in industrial-organizational psychology," in *Problems and Solutions in Human Assessment*, eds R. D. Goffin and E. Helmes (New York, NY: Springer), 215–248.
- Rothstein, M. G., and Goffin, R. D. (2006). The use of personality measures in personnel selection: what does current research support? *Hum. Resour. Manag. Rev.* 16, 155–180. doi: 10.1016/j.hrmmr.2006.03.004
- Schmitt, D. P., Allik, J., McCrae, R. R., and Benet-Martinez, V. (2007). The geographic distribution of Big Five personality traits. *J. Cross. Cult. Psychol.* 38, 173–212. doi: 10.1177/0022022106297299
- Sjöberg, L. (2015). Correction for faking in self-report personality tests. *Scand. J. Psychol.* 56, 582–591. doi: 10.1111/sjop.12231
- Stark, S., Chernyshenko, O. S., and Drasgow, F. (2005). An IRT approach to constructing and scoring pairwise preference items involving stimuli on different dimensions: the multi-unidimensional pairwise-preference model. *Appl. Psychol. Meas.* 29, 184–203. doi: 10.1177/0146621604273988
- Stark, S., Chernyshenko, O. S., Drasgow, F., Nye, C. D., White, L. A., Heffner, T., et al. (2014). From ABLE to TAPAS: a new generation of personality tests to support military selection and classification decisions. *Mil. Psychol.* 26, 153–164. doi: 10.1037/mil0000044
- The International Cognitive Ability Resource Team (2014). "International Cognitive Ability Resource" (Measurement Instrument). Available at: <http://icar-project.com/>
- Thurstone, L. L. (1927). A law of comparative judgment. *Psychol. Rev.* 34, 273–286. doi: 10.1037/h0070288
- van Eijnatten, F. M., van der Ark, L. A., and Holloway, S. S. (2015). Ipsative measurement and the analysis of organizational values: an alternative approach for data analysis. *Qual. Quant.* 49, 559–579. doi: 10.1007/s11135-014-0009-8
- Viswesvaran, C., and Ones, D. S. (1999). Meta-analyses of fakability estimates: implications for personality measurement. *Educ. Psychol. Meas.* 59, 197–210. doi: 10.1177/00131649921969802
- Vroom, V. H. (1964). *Work and Motivation*. New York, NY: Wiley.
- Vukonjanski, J., Nikolić, M., Hadžić, O., Terek, E., and Nedeljković, M. (2012). Relationship between GLOBE organizational culture dimensions, job satisfaction, and leader-member exchange in Serbian organizations. *J. East Eur. Manag. Stud.* 17, 333–368. doi: 10.2307/23281716
- Weigold, A., Weigold, I. K., and Russell, E. J. (2013). Examination of the equivalence of self-report survey-based paper-and-pencil and internet data collection methods. *Psychol. Methods* 18, 53–70. doi: 10.1037/a0031607

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2016 Van Geert, Orhon, Cioca, Mamede, Golušin, Hubená and Morillo. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.