Cross-cultural differences in social desirability scales: Influence of cognitive ability

Author:
Aletta Odendaal[1]

Affiliation:
[1]Department of Industrial Psychology and People Management, University of Johannesburg, South Africa

Correspondence to:
Aletta Odendaal

Email:
aodendaal@uj.ac.za

Postal address:
PO Box 781615, Sandton 2146, South Africa

Orientation: The use of personality tests for selection and screening has been consistently criticised resulting from the risk of socially desirable responding amongst job applicants.

Research purpose: This study examined the magnitude of culture and language group mean-score differences amongst job applicants and the moderating effect of race on the relationship between social desirability and cognitive ability.

Motivation for the study: The influence of cognitive ability and potential race and ethnic group differences in social desirability scale scores, which can lead to disproportional selection ratios, has not been extensively researched in South Africa.

Research design, approach and method: A quantitative, cross-sectional research design, based on secondary datasets obtained from the test publisher, was employed. The dataset consisted of 1640 job applicants across industry sectors.

Main findings: Moderated multiple regression analyses revealed that the relationship between social desirability and general reasoning was moderated by culture and language, with group differences in social desirability being more pronounced at the low general reasoning level. This suggests that social desirability scales may be an ambiguous indicator of faking as the scales may indicate tendency to fake, but not the ability to fake, that is likely to be connected to the level of cognitive ability of the respondent.

Practical/managerial implications: Individual differences in social desirability are not fully explained by cognitive ability as cultural differences also played a role. Responding in a certain manner, reflects a level of psychological sophistication that is informed by the level of education and socio-economic status. In relation to selection practice, this study provided evidence of the potentially adverse consequences of using social desirability scales to detect response distortion.

Contribution/value-add: The exploration of cross-cultural differences in the application of social desirability scales and the influence of cognitive ability is seen as a major contribution, supported by possible explanations for the differences observed and recommendations regarding the practice of universal corrections and adjustments.

Introduction

The inferences made from social desirability scales included in personality instruments in cross-cultural settings remain questionable. This is so despite the fact that the use of personality instruments in personnel selection has increased in the last decade because these instruments have been shown to predict job performance and other related behaviours across employment settings (Hough & Oswald, 2008; Ones, Dilchert, Viswesvaran & Judge, 2007; Sackett, 2011; Viswesvaran, Deller & Ones, 2007). Furthermore, well-constructed personality instruments have sound psychometric properties, are relatively inexpensive to administer and score, and are likely to cause less adverse impact on minority groups than cognitive ability tests (Ones, Viswesvaran & Reiss, 1996; Schmidt & Hunter, 2004). Adverse impact in personnel selection typically occurs when a specific selection strategy gives members of a specific group a lower likelihood of being selected than members of another group (Theron, 2007).

Notwithstanding the advantages, the use of personality tests for selection and screening has been consistently criticised consequent to the risk of socially desirable responding amongst job applicants (Birkeland, Manson, Kisamore, Brannick & Smith, 2006; Griffith, Chmielowski & Yoshita, 2007; Hogan, Barrett & Hogan, 2007; Morgeson *et al.*, 2007). Evidence suggested that the self-report format of personality instruments is highly susceptible to response distortion by applicants, as individuals can intentionally distort their responses to create a favourable impression of themselves (Holden & Passey, 2010; O'Connell, Kung & Tristan, 2011;

Visweswaran & Ones, 1999). Empirical studies have further demonstrated that applicants actually engage in response distortion behaviour (Barrick & Mount, 1996; Schmit & Ryan, 1993). The possibility that applicants can, and indeed do distort their responses, has serious implications for employers as one might hire an individual based on their ability to assumedly distort their responses on an instrument rather than as a result of the characteristic being measured.

Despite the criticism, personality instruments are widely used in South Africa for selection and development purposes. Although the effect of socially desirable responses on the validity and utility of personality testing in employment settings has been extensively debated and researched in the international literature, the issue remains unresolved (e.g. Birkeland *et al.*, 2006; Dilchert, Ones, Viswesvaran & Deller, 2006; Ellingson, Sackett & Connelly, 2007; Li & Bagger, 2006; O'Connell *et al.* 2011; Ones *et al.*, 1996). In addition to this, there is a growing recognition that the cross-cultural transferability of constructs has not been systematically examined in the multi-cultural and multi-lingual South African context (De Beer, 2004; Meiring, Van de Vijver, Rothmann & Barrick, 2005; Schaap & Vermeulen, 2008). Specifically, the influence of potential race and ethnic group differences in social desirability scale scores, and the relationship between social desirability and cognitive ability amongst job applicants in a cross-cultural context, have also not been extensively researched (Dilchert & Ones, 2005).

Considering the relationship between social desirability and cognitive ability a meta-analytic review by Ones *et al.* (1996) suggested a weak negative relationship. The estimates were, however, not solely based on applicant samples and in a subsequent study, using a job applicant sample, Dilchert and Ones (2005) found that race and ethnic group differences in social desirability scale mean-scores is partially explained by group differences in cognitive ability. Limited research has been conducted in South Africa examining the relationship between social desirability and cognitive ability. A study by Meiring *et al.* (2005) investigating method bias in a selection battery for entry-level police officials in the South African Police Services, found that the extent of cross-cultural differences between the language groups was not influenced by socially desirable responding or cognition. Greene (2000) and also Dilchert and Ones (2005) reviewed the influence of group differences in cognitive ability on social desirability scale mean-scores, and suggested that socially desirable responding appear to be associated with a certain form of social naïveté, likely to be connected to cognitive ability.

The magnitude of the social desirability and cognitive ability relationship amongst job applicants in a cross-cultural context remains an open question that requires further exploration. It is against this background that this article aims to report on the findings of a study that examined the influence of cognitive ability on group differences in social desirability amongst job applicants in South Africa, addressing the following research questions:

- What is the relationship between socially desirable responding and cognitive ability?
- Are any race differences in social desirability scores related to differences in cognitive ability?

More specifically the objectives of the study are to:

1. Examine the magnitude of culture and language group mean-score differences on social desirability scores and cognitive ability amongst job applicants.
2. Examine whether or not race moderates the relationship of social desirability with cognitive ability.

The main contribution of this research is not only the exploration of cross-cultural differences in the application of social desirability scales and the influence of cognitive ability, but also the provision of possible explanations for the differences observed. The author also provides recommendations regarding the practice of universal corrections and adjustments.

In the next section the construct social desirability will be described, followed by a review of evidence regarding cross-cultural differences in socially desirable responding, and the influence of cognitive ability, in order to advance our understanding regarding the pattern of relationships that requires further exploration.

Literature review

Socially desirable responding is viewed as an important component of self-report inventories that has inspired much debate, and has generated mixed and at times contradictory research results, depending on the operational definition of social desirability and research design employed. Reviewing the large body of research on social desirability revealed that various terms have been used to describe the construct. These terms include impression management (Hogan *et al.*, 2007; Paulhus, 1984), faking (Barrick & Mount, 1996; Ones *et al.*, 1996; Rossè, Stecher, Miller & Levin, 1998), self-deception (Paulhus, 1984, 2002) and self-enhancement (Heine, 2005; Heine & Lehman, 1997). These terminologies are regularly used interchangeably and are conceptualised as a unitary construct, notwithstanding clear differences in meaning and application (Griffith & Robie, 2013; Li & Bagger, 2006; Ones *et al.*, 1996). Although many of the terms are conceptually distinct they all relate to the elevation of scores on a self-report inventory. It should, however, be noted that socially desirable responding is not only restricted to personality inventories but is also a concern in any assessment conducted for the purpose of decision-making (selection, promotion or development opportunities) regardless of the assessment instrument or medium (Dilchert *et al.*, 2006).

The terms 'distortion' and 'faking' are furthermore perceived to be misleading concepts because they imply that there is a 'true' response that can be determined independently of the

behaviour of the test-taker (Mueller-Hanson, Heggestad & Thornton, 2006). The term faking implies that test-takers are aware of the 'amount of a psychological construct they actually possess, and they respond to items in a way that is knowingly inconsistent with this (known) possessed amount of the psychological construct' (Davies, Norris, Turner & Wadlington, 2005, p. 4). Faking, therefore, has a negative connotation and is applied only to personality test responses and not to cognitive ability (where changes in cognitive test scores are commonly referred to as practice effects) or any other type of responses that may vary in a similar manner (Davies *et al.*, 2005; Hogan *et al.*, 2007). In this study social desirability was operationalised as response patterns that can result from both self-deception and impression management. Faking was defined as the purposeful misrepresentation and conscious distortion of responses in order to score favourably and was, thus, viewed as a form of impression management. In applied settings psychologists work towards eliminating sources of bias or systematic error, such as self-enhancement, which are not relevant to the measured attributes through the use of different measures of socially desirable responding.

Measures of socially desirable responding

In an attempt to address the effects of socially desirable responding, different strategies are employed. These strategies largely depend on the purpose and level of application, broadly classified as (1) identification versus prevention strategies and (2) item and scale versus person level strategies (Aurthur & Glaze, 2011; Dilchert & Ones, 2011). Identification strategies typically aim to detect response distortion amongst test-takers, whereas prevention strategies attempt to discourage test-takers from engaging in response distortion by making socially desirable responding more difficult (Burns & Christiansen, 2011; Rothstein & Goffin, 2006). Social desirability scales are, therefore, employed as a strategy to identify response distortion on scale or test level. Common approaches to identify socially desirable responding in personality instruments include the use of one or more social desirability, impression management or faking scale(s), referred to as validity scales (Burns & Christiansen, 2011; Ellingson, Heggestad & Makarius, 2012) and should be distinguished from other response style indicators such as acquiescence and extreme response sets. Validity scale items are typically dispersed amongst the personality items in the personality instrument. These validity scales examine the pattern of responses and infer the credibility of the personality profile obtained (Holden & Passey, 2010; O'Connell *et al.*, 2011). A survey conducted by Goffin and Christiansen (2003) reported that 80% of commercial self-report inventories include social desirability scales, such as the Balanced Inventory of Desirable Responding (BIDR, Paulhus, 1984), the Marlowe-Crowne Scale (Crowne & Marlowe, 1960), or the Edwards Social Desirability Scale (Edwards, 1957). It was further evident from the survey that the majority of psychologists using personality questionnaires reported that they interpret validity scales despite the lack of clear directives on how these scales should be interpreted.

Several independent social desirability scales are also available that can be administered separately from the other assessments in a battery. In this regard different short forms of the Marlowe-Crowne Scale have subsequently been developed with reported internal consistencies around .60 (Barger, 2002). A meta-analyses conducted by Ones *et al.* (1996) reported a mean estimate of the social desirability scales' reliability of .74 across 119 reliabilities with an associated standard deviation of .14. It is evident that none of these reported reliabilities suggest that scores on these scales should be used to make decisions about individuals (cf. Nunnally & Bernstein, 1994). Irrespective of reported reliabilities, current personality instruments widely used in South Africa include one or more validity scale, and the practice remains that Industrial and Organisational (IO) Psychologists treat scores on these scales as signs of response bias and evidence of faking.

Given the pervasive influence of socially desirable responding on all types of behaviours, in everyday life, Tett, Freund, Christiansen, Fox & Coaster (2011) argue that it is important to examine process models of how responses are generated and develop an understanding of the antecedents of response bias behaviour. This argument is in accordance with more recent definitions of socially desirable responding that view faking as representing a response set aimed at providing a description of the self in an attempt to achieve personal goals. According to this theory faking occurs 'when this response set is activated by situational demands and person characteristics to produce systematic differences in test scores that are not due to the attribute of interest' (Ziegler, MacCann & Roberts, 2011, p. 8). Despite this definition little is known about the actual process of socially desirable responding in a cross-cultural context.

In an attempt to obtain answers to questions regarding how responses are generated, and to questions concerning what people think when completing a self-report inventory, various cognitive (Krosnick, 1999; Tett & Simonet, 2011; Ziegler, 2012) and psychological process models have been conceptualised (cf. McFarland & Ryan, 2000, 2006; Mueller-Hanson *et al.*, 2006; Snell, Sydell & Lueke, 1999). Cognitively, people go through a four-step process of responding which, according to Krosnick (1999), consists of comprehension, retrieval, judgement and mapping. When people are motivated to respond in a sincere manner an optimising strategy is followed. In contrast, a satisficing strategy is used when factors such as motivation, cognitive ability and fatigue influence optimal responding. The cognitive process model, thus, supports the assumption that a person's ability and motivation for faking influences their strategy of either optimising or satisficing (Ziegler & Bühner, 2009).

In order to explain the psychological processes that underlie faking behaviour Snell *et al.* (1999) proposed an interactional framework for understanding both individual differences (ability and willingness or motivation to fake) and also situational differences in successful faking. In an attempt

to address the conceptual limitations of the interactional model, McFarland and Ryan (2006) proposed a second model based on the theory of planned behaviour (Ajzen, 1991). A central factor in the theory of planned behaviour concerns the individual's intention to engage in the behaviour. Within this model faking is viewed as a result of three conceptually independent determinants, namely:

• attitude towards faking (beliefs about rightness or wrongness of faking)
• subjective norms (beliefs about how others view faking and the perceived social pressure to perform or not perform the behaviour)
• perceived behavioural control (beliefs about the ease or difficulty of faking).

Some support was found in research for this model but it was limited as it did not address the impact of dispositional factors on faking intentions (Mcfarland & Ryan, 2006).

To address limitations from the earlier models, Mueller-Hanson *et al.* (2006) integrated the models of faking proposed by McFarland and Ryan (2000) and Snell *et al.* (1999), to develop an integrative model of faking behaviour that explains the predictors of individual differences in the motivation and ability to distort responses. This integrative model of faking (Mueller-Hanson *et al.*, 2006) includes both dispositional and attitudinal antecedents. In accordance with the theory of planned behaviour (Ajzen, 1991) these antecedents precede intentions, which precede behaviour. Antecedents include:

• a person's perception of the situation (based on belief in the importance of faking, perceived behavioural control and subjective norms)
• ability to fake (operationalised as knowledge)
• willingness to fake
• the two core personality characteristics of conscientiousness and emotional stability.

Whilst a full review of all process models of faking is outside the scope of this article, the preceding discussion clearly indicates that explaining human behaviour and specifically socially desirable responding is a complex and cognitively demanding task. The fact that an individual can fake responses to an item when instructed to do so, and can present themselves as the ideal employee for a position, entails a complex set of cognitive processes and it is expected that respondents, higher in general ability, will be better at enhancing performance on a self-report inventory (O'Connell *et al.*, 2011; Tett & Simonet, 2011). Process models of socially desirable responding, therefore, provide useful conceptual frameworks for understanding how responses are generated, and for working towards the implementation of interventions that may be effective in changing them.

The evidence provided further suggests that people differ in relation to how much they will fake on a personality instrument, with some people faking substantially and others faking little or not at all. The extent to which individuals fake is partially determined by their perception of the situation, their willingness and ability to fake and also their personality characteristics (Mueller-Hanson *et al.*, 2006). The evidence also suggests that people tend to implement a total test strategy of self-presentation and take into account how specific responses may relate to other items in the instrument, in an effort to optimise the impact of their overall performance (Hogan *et al.*, 2007; Tett & Simonet, 2011). These findings bring into question the use of social desirability scales to determine the validity of a personality profile. In addition, the literature leaves unaddressed a central theoretical and practical consideration, regarding potential race and ethnic group differences, in relation to the use of social desirability scales in selection decisions.

Group differences in socially desirable responding

The importance of group differences in a cross-cultural context cannot be underestimated as it can lead to disproportional selection ratios and possible adverse impact. For example, if scores on the social desirability scale are used to make a judgement regarding the validity of the profile, and subsequent corrections are made, then individuals from different groups will receive systematically different scores on these scales. In addition, political and social changes in South Africa over the last 20 years present additional challenges to the use of personality instruments in selection decisions, as the differences between race groups and people from diverse backgrounds are constantly changing (Foxcroft & Roodt, 2009).

The issue of race and language group differences in social desirability has been researched internationally and has yielded mixed results. Hough and Ones (2001) found small to moderate group mean-score differences between some racial and ethnic groups on social desirability scales. They reported *d*-values of -.05, .56, .03, .40 for the black-white, Hispanic-white, American Indian-white and Asian-white group comparisons, respectively. The effect sizes for the Hispanic-white and Asian-white comparisons were large. Dilchert and Ones (2005) criticised this study because it did not focus specifically on job applicants and did not explore potential explanations for the race and ethnic group differences (e.g., cognitive ability). In addition, relatively small sample sizes for the American Indian and Asian groups limited the strength of the estimates.

There is currently strong evidence that job applicants' score distributions are significantly different from those of the general population and, therefore, the use of job applicant samples is imperative when studying group differences in social desirability (Rossé *et al.*, 1998; Tett & Simonet, 2011; Viswesvaran, Ones, Cullen, Drees & Langkamp, 2003). Two independent studies conducted by Dilchert and Ones (2005), which made use of a sample of over 50 000 job applicants, from two occupational groups, showed that black, Hispanic and Asian groups displayed moderately higher mean-scores on social desirability scales than white applicants

(differences ranging from .14 to .16 standard deviations). This study concluded that race and ethnic group differences in socially desirable responding scale mean-scores can partially be explained by group differences in cognitive ability. However, the magnitude of the relationship between social desirability and cognitive ability amongst job applicants in a South African context remains an open question that requires further exploration.

Influence of cognitive ability on socially desirable responding

Ability and aptitude tests have commonly been used as predictors in personnel selection. A typical assessment battery for selection purposes in South Africa (and many other countries) tends to include both personality and ability instruments. The relationship between cognitive ability and work performance has been extensively researched with a substantial body of evidence indicating that general cognitive ability (or general mental ability, 'g') is the strongest predictor of learning and acquisition of job knowledge and also overall job performance for virtually every job (Arneson, Sackett & Beatty, 2011; Kuncel & Hezlett, 2010; Schmidt & Hunter, 2004). In reviewing research evidence, Ones, Dilchert and Viswesvaran (2012) asserted that general mental ability is also relevant for understanding and predicting other important behaviours and outcomes in occupational settings (e.g. leadership effectiveness, innovation, counter productivity, and work attitudes). It is, further, well documented that g-tests show mean sub-group differences in cognitive ability test performance by race and ethnicity, sex, and age both within the United States and internationally (see Ones *et al.*, 2012 for a quantitative review of group differences; Roth, Bevier, Bobko, Switzer & Tyler, 2001).

The cross-cultural comparison of cognitive test scores is also not new in South Africa, with results reflecting those commonly reported in the international literature (for a review see Odendaal, 2013). It should, however, be noted that these studies were conducted prior to the 1990s, and labour force participation and occupational distribution of women and ethnicity in the workplace are totally different to 20–30 years ago. In addition, changes in the nature and demands of jobs (e.g. greater complexity and technological demands) may manifest differently in cognitive ability relations. Important outcomes of early research is the recognition that home environment, schooling, language proficiency, nutrition and other factors may impact cognitive ability measures in a multi-cultural society such as South Africa (Claassen, 1997; Meiring *et al.*, 2005). Reviewing research it is further evident that mean-scores on cognitive ability measures have been documented to be steadily increasing, referred to as the Flynn effect, after the researcher who first documented the narrowing of mean group differences on cognitive ability tests over time (cf. Ang, Rodgers & Wänström, 2010; Rushton & Jensen, 2010).

Typical standardised tests of cognitive ability used in South Africa assess verbal ability, numerical ability, and deductive

reasoning. Following Cattell's theory, these areas can be viewed as indicators of crystallised intelligence and often concern specialised skills or knowledge required by a given culture (Taylor, 1994). South African studies have found that race, level of education, socio-economic status, language and understanding of English are the main factors impacting the construct and item comparability of cognitive and personality tests (Abrahams & Mauer, 1999; Meiring *et al.*, 2005; Stephen, Welman & Jordaan, 2004; Van Zyl & Visser, 1998). A study by Watkins and Elliot (1997) also raised serious questions regarding the functionality of g in the prediction of work performance in South Africa. These authors rejected the notion of a g-factor and argued instead in support of the notion of seven distinct intelligences (logical-mathematical, musical, intra-personal, interpersonal, bodily-kinesthetic and spatial intelligence) based on the work of Gardner (Watkins & Elliot, 1997).

Given the educational differences in South Africa, research has focused on the educability and trainability of South Africans. In this regard Taylor (1994) suggested the identification of learning potential as an alternative to conventional cognitive ability assessment. This suggestion is based on the belief that cognitive ability is not fixed but can change and, following Vygotsky's view, supports the approach that performance on its own is not a true reflection of cognitive ability (Bedell, Van Eeden & Van Staden, 1999). Based on this argument several non-verbal tests have been developed to measure fluid intelligence, which is a relatively culture-reduced form of mental efficiency and is related to a person's inherent capacity to learn and solve problems (Schaap & Vermeulen, 2008). There is, furthermore, recognition that different cultures have different views regarding intelligence, for example, the speed at which a herdsman recognises his own cattle amongst a big herd may be perceived as intelligence in some African tribes., whereas other forms of pattern recognition might be acknowledged as intelligence within Western cultures (De Klerk, 2008).

The comparability of test scores may, therefore, be influenced by ability patterns that are influenced by socio-cultural patterns. In contrast, the main cognitive processes and functions (fluid intelligence) are universally shared properties of intellectual life and may result in highly varied crystallised performances across cultures (Berry, Poortinga, Segall & Dasen, 2002). It is also evident that the greater the amount of information that needs to be manipulated the more important g becomes (Arneson *et al.*, 2011; Kunzel & Hezlett, 2010; Schmidt & Hunter, 2004). In practical terms this means that as the information-processing demands of a position increase, a person with lower general mental ability is less likely to be successful than a person with a higher g. Although this research study did not aim to investigate the taxonomy of cognitive ability it remains important to emphasise the impact of cross-cultural differences on cognitive performance.

In this regard a study by Helms-Lorentz, Van de Vijver, and Poortinga (2003) examined the cultural loadings of

test material. The term cultural loading refers to a specific cultural context that is reflected in the instrument or in the administration thereof. The cultural context is usually that of the test developer and it can create intergroup differences that are not related to the construct measured by the instrument. The results of the study by Helms-Lorentz *et al.* (2003) suggested that cultural complexity (*c*) was as important as *g* in the explanation of performance differences between cultural groups on cognitive tests.

According to the discussion of process models of faking, the ability to fake successfully implies that a respondent must be able and motivated to distort responses (cf. Tett *et al.*, 2011). According to English *et al.* (2005) individuals who are able to fake must have analytical ability to apply problem-solving to understand the construct being measured and also to understand the advantages of faking such behaviour. In this regard they reported that individuals high in *g* recognise and solve problems more successfully than individuals low in *g*. In addition, individuals high in *g* can understand the items in the instrument, can detect desirable answers and can respond accordingly. However, the study found that intelligence did not predict response distortion but that job knowledge moderated (strengthened) the ability to fake. The findings were consistent with previous research, which indicated that general mental ability has a major effect on the acquisition of job knowledge (Arneson *et al.*, 2011; Kunzel & Hezlett, 2010). People higher in general mental ability acquire more job knowledge and at a faster rate than people with lower general mental ability. In addition, research conducted by Wrensen and Biderman (2005) indicated that cognitive ability was positively related to the ability to fake extroversion, conscientiousness and stability (individual differences). t is, therefore, important to investigate whether or not race moderates the relationship between social desirability and cognitive ability.

Research design

Research approach

To meet the main objective of this study a quantitative, cross-sectional research design based on secondary data were employed. The secondary datasets were obtained from Psytech South Africa, the test publisher for the measuring instruments utilised in the study. The use of secondary data is appropriate as the datasets were collected anonymously. The researcher, therefore, need not be concerned with ethical issues concerning the protection of participant identity (Spector, cited in Anderson, Ones, Sinangil & Viswesvaran, 2001). The main limitations of using secondary data involve the researcher's inability to control for data collection errors, the lack of control over the selection of samples and comparison groups and also the quality of the sampling frame which can influence the generalisability of the results (Mouton, 2001). In order to counter the effects of sample selection, both language and race were utilised as independent variables as they are of particular concern in South Africa when evaluating an instrument for the presence of bias (Van

de Vijver & Rothman, 2004). Language in South Africa also varies in relation to culture. South Africa recognises eleven official languages and the custom is, therefore, to assess in the language used in the workplace. The dominant language of business and industry in South Africa is English and all of the measuring instruments utilised in this study were administered in English.

Research method

Measuring instruments

In this study the influence of cognitive ability on socially desirable responding was examined utilising a social desirability measure and a cognitive ability measure. The Social Conformity scale of the Occupational Personality profile (OPP) was used to operationalise social desirability. The Marlow-Crowne Scale (Crowne & Marlow, 1960) forms the basis of the Social Conformity scale in the OPP and consists of 8 items with a 5–point response format ranging from (1) 'strongly agree' to (5) 'strongly disagree'. Budd (1991) reported that the reliability of the Social Conformity scale, as estimated by the Cronbach alpha coefficient, is .59. The reliability of the Social Conformity scale is below the acceptable standard of .70, but was considered acceptable for research purposes as the personality inventory with the social conformity scale is currently in use in South Africa (cf. Aguinis, Henle & Ostroff, 2001). The General Reasoning Test Battery (GRT2), which measures general verbal, numerical and abstract reasoning, was employed to measure cognitive ability. The three sub-tests within this battery have been shown to demonstrate a good standard of reliability, as reported by the following reliability coefficients:

- verbal reasoning (35 items, $\alpha = .83$)
- numerical reasoning (25 items, $\alpha = .84$)
- abstract reasoning (25 items, $\alpha = .83$) (Budd, 1993).

Research participants and procedure

Participants were 1640 adult job applicants[1] (595 female and 1045 male) who completed both the Social Conformity Scale of the OPP and the GRT2. Job applicant data sets were utilised as there is currently strong evidence that the job applicants' score distributions are significantly different from those of the general population and, therefore, the use of job applicant samples is imperative when studying group differences in social desirability (Rossè *et al.*, 1998; Viswesvaran *et al.*, 2003). The average age of the participants was 26 years. Closer inspection revealed that comparing the collective black Nguni languages representing Zulu, Xhosa, Ndebele and Swazi (*n* = 638), and black Sotho language groupings representing Tswana, Pedi and South Sotho (*n* = 517), with that of white Afrikaans (*n* = 272) and white English speakers (*n* = 212) would provide a large enough sample for comparison. As there were insufficient numbers of participants from the coloured[2] and Asian ethnic groups,

1.Limitations for the use of secondary datasets are the inability to control for level of education and gender representation.

2.The term 'coloured' is used to refer to people of mixed racial descent and is used by the South African government as part of its official racial categorisation scheme.

these were excluded from the study. The data were collected from various South African companies who used the OPP and GRT2 for selection purposes. During test administration all participants provided consent to the Test Publisher that information can be utilised for research purposes. In addition, all psychometric data were dealt with confidentially and strict ethical publishing practices were followed (Ethical Rules of Conduct, HPCSA, 2006).

Statistical analysis

To achieve the main research objectives of the study the abstract, numerical and verbal reasoning sub-tests of the GRT2 were subjected to a principal axis factor analysis with iterated communalities, followed by a multi-group confirmatory factor analysis. The objective of these analyses was to establish whether or not an invariant total General Reasoning factor could be extracted from the three sub-tests. Next, mean differences, standard deviations and effect sizes were calculated using the white English-speaking group as the reference group. To examine the relationships between social desirability, cognitive ability and culture and language, a moderated multiple regression (MMR) was conducted, where the English group served as the reference group.

Next, the results are presented followed by a discussion of the results to provide possible explanations for the differences observed.

Results

The main objective of the study was to determine whether or not any potential race group differences in social desirability scores are the result of potential group differences in cognitive ability. As a first step it was important to establish whether or not an invariant total General Reasoning factor could be extracted from the three sub-tests of the General Reasoning Test battery (GRT) employed to measure cognitive ability. The Abstract, Numerical and Verbal Reasoning sub-tests were subjected to a principal axis factor analysis with iterated communalities, followed by a multi-group confirmatory factor analysis. The principal axis factor analysis was chosen as the extraction method because the objective of the analysis was to detect structure and to estimate the proportion of variance that each item has in common with other items. A one-factor solution produced a very good fit with the data, with all the correlation residuals very close to zero (the largest residual was 0.01) showing that no more than one meaningful factor could be extracted. The three eigenvalues of the unreduced correlation matrix were 2.39, 0.32, and 0.29, which also points to the retention of a single factor. The factor loadings of the three sub-tests on this factor were as follows:

- Abstract Reasoning = 0.85
- Numerical Reasoning = 0.82
- Verbal Reasoning = 0.82.

In addition, the multi-group confirmatory factor analysis showed a satisfactory fit for a one-factor model requiring equal unstandardised factor loadings across the four ethnic

TABLE 1: Descriptive statistics for general reasoning and social conformity scale.

Variable	M	SD	Effect size (*d*)
General Reasoning			
English	45.44	14.89	*
Afrikaans	42.71	14.48	-0.18
Nguni	26.10	11.41	-1.30
Sotho	26.30	11.37	-1.29
Social Conformity Scale			
English	22.86	4.05	*
Afrikaans	22.67	4.14	-0.04
Nguni	24.52	4.48	0.38
Sotho	24.66	4.42	0.38

M, mean; SD, standard deviation
*, Note. The white English-speaking group served as the reference group.

groups [$\chi^2(6) = 53.808$, $p < .001$; CFI = 0.973; RMSEA = 0.070]. On the basis of these results it was deemed appropriate to calculate a single General Reasoning score for each person by aggregating scores with unit weighting over the three sub-tests. The General Reasoning score was conceptualised as being representative of general mental ability or *g*. The results of this study are consistent with the test developers' intention of constructing a test that measures general reasoning ability (Budd, 1993). The mean General Reasoning score was subtracted from each respondent's observed score, which resulted in a centred General Reasoning variable with a mean of zero. These centred General Reasoning scores were used in the subsequent moderated multiple regression analyses (cf. Aguinis, 2004).

Next, the means, standard deviations and effect sizes of the culture and language groups for the Social Conformity Scale and General Reasoning were examined. The uncentred means and standard deviations of the four culture and language groups, for OPP Social Conformity Scale and General Reasoning, are provided in Table 1, which also contains the group mean differences.

Using the white English-speaking group as the reference group, the effect sizes of the differences in means for General Reasoning were as follows: black Sotho-speaking, *d* = -1.29; black Nguni-speaking, *d* = -1.30; and white Afrikaans-speaking, *d* = -0.18. On average, the two black groups scored substantially lower than the two white groups on General Reasoning. However, the differences between the two black groups and between the two white groups were small.

Table 1 also shows that the two black groups scored higher than the two white groups on the OPP Social Conformity Scale. However, within the black group the Sotho and Nguni groups obtained very similar mean-scores. Similarly, within the white group the Afrikaans and English groups obtained very similar scores. Using the white English-speaking-group as the reference group, the effect sizes of the mean differences for OPP Social Conformity were as follows: black Sotho-speaking, *d* = 0.38; black Nguni-speaking, *d* = 0.38; and white Afrikaans-speaking, *d* = -0.04.

To examine the relations of the General Reasoning and cultural and language group with OPP Social Conformity Scale a

TABLE 2: Hierarchical moderated multiple regression of the relations between OPP Social Conformity Scale, cognitive ability, and language/culture.

| Model	R	R^2	Adjusted R^2	SE of the estimate	Change statistics				
					ΔR^2	ΔF	$df1$	$df2$	p
1	0.191	0.037	0.036	4.344	0.037	62.292	1	1638	< .001
2	0.215	0.046	0.044	4.326	0.010	5.454	3	1635	.001
3	0.228	0.052	0.048	4.318	0.006	3.247	3	1632	.021

SE, standard error; *df*, degrees of freedom.

moderated multiple regression (MMR) was undertaken. For the purpose of the MMR the language group variable was dummy coded. The English group served as the reference group and three coded vectors were created to represent the Afrikaans, Nguni, and Sotho groups. The MMR proceeded in three steps as indicated in Table 2. General Reasoning (centred) was entered in the first step, the dummy coded language group vectors were added in the second step, and the products of General Reasoning (centred) and the dummy coded language group vectors were added in the third step.

General Reasoning, language group, and their interaction jointly accounted for approximately 5.2% of the variance in OPP Social Conformity Scale, $R^2 = 0.052$, $F (7, 1632) = 12.747$, $p < .001$. Inspection of the ΔR^2 for the third step shows that the interaction of General Reasoning and language group made a statistically significant contribution to the prediction of OPP Social Conformity above and beyond the contributions made by the General Reasoning and language group, $\Delta R^2 = 0.006$, $F (3, 1632) = 3.247$, $p = .021$. The data, therefore, support the proposition that culture and language moderates the relationship between General Reasoning and the OPP Social Conformity Scale. The effect size for the interaction of General Reasoning and language group was $f^2 = 0.006$. Aguinis and Henle (2003) reported a mean effect size of 0.009 with 95% confidence intervals of 0.006 and 0.012 in their review of studies, using MMR in top tier industrial and organisational psychology and management journals. The effect size for the interaction obtained in this study is consistent with those reported in the industrial psychology literature in general.

Inspection of the ΔR^2 for the second step shows that the cultural group made a small but statistically significant contribution to the prediction of the OPP Social Conformity Scale above and beyond the contribution of General Reasoning [$\Delta R^2 = 0.010$, $F(3, 1635) = 5.454$, $p = .001$]. Hence, the data also support the proposition that there are cultural differences in social desirability when cognitive ability is held constant. Finally, the first step of the MMR showed that General Reasoning was a statistically significant predictor of OPP Social Conformity [$R^2 = 0.037$, $F (3, 1638) = 62.292$, $p < .0001$].

Against the background of the significant interaction between General Reasoning and culture, indicating that the slope of the regression lines differs across groups, separate regression equations were calculated for the four language groups. Predicted scores for each of the four language groups were calculated at one standard deviation below the mean and one standard deviation above the mean of General
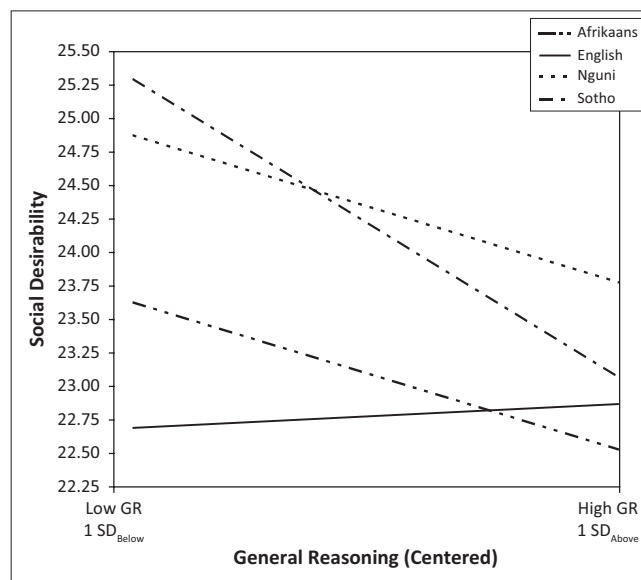
The English group served as the reference group.

FIGURE 1: Relationship of general reasoning with social desirability.

Reasoning. These predicted scores are plotted in Figure 1, which shows that the relations between cognitive ability and social desirability differ across groups.

Figure 1 shows that at low General Reasoning levels, the biggest difference in predicted Social Conformity scores was observed for the Sotho and English groups (with the Sotho group having higher predicted Social Desirability scores). In contrast, at high General Reasoning levels the biggest absolute difference was observed for the Afrikaans and Nguni groups (with the Nguni group having higher predicted Social Conformity scores). Finally, Figure 1 shows that the group differences in Social Conformity are much more pronounced at the lower end of General Reasoning. Although the group differences at the upper end of General Reasoning are much smaller they remain clearly visible. For the English group there is virtually no relationship, whereas for the remaining three groups there is a clear trend towards individuals with high cognitive ability tending to give less socially desirable responses.

To conclude, the exploratory factor analysis showed that a one-factor solution produced a very good fit and the total score of the Verbal, Numerical and Abstract reasoning tests was, therefore, used to operationalise General Reasoning ability. Examining the magnitude of culture and language group mean-score differences on social desirability scores and cognitive ability amongst job applicants, the results showed that the two black groups scored higher than the two white groups on the Social Conformity Scale. However,

within the black group the Sotho and Nguni groups obtained very similar mean-scores. Similarly, within the white group the Afrikaans and English groups obtained very similar scores. In terms of the relationship of social desirability with cognitive ability, the results showed group mean differences in General Reasoning with the Nguni and Sotho groups scoring lower on General Reasoning and higher on Social Desirability than the Afrikaans and English groups. The relationship of Social Desirability and General Reasoning is, therefore, moderated by culture and language with group differences in Social Desirability more pronounced at the low General Reasoning level.

Discussion

Early on in the conceptualisation of social desirability Crowne and Marlowe (1960) suggested that people respond in a manner that is culturally acceptable in order to obtain social approval. Culture was therefore recognised as an important factor when determining whether opinions and behaviours of people are desirable or not (Johnson & Van de Vijver, 2003). However, questions have never been seriously examined in South Africa regarding the relationship between socially desirable responding and cognitive ability amongst job applicants, and whether or not race differences in social desirability scores are related to differences cognitive ability. The results of this study show that on average, Nguni-speaking and Sotho-speaking participants scored lower on the GRT2 and slightly higher on the Social Conformity scale than their Afrikaans and English-speaking counterparts. Of greater significance is the finding that general reasoning is negatively related to social desirability and that this relationship is moderated by ethnicity. Possible explanations for group differences in the South African context can be attributed to:

1. the level of education, socio-economic status, language and understanding of English (Abrahams & Mauer, 1999; Meiring et al., 2005; Stephen et al., 2004; Van Zyl & Visser, 1998).
2. cultural loadings of test material that include implicit and explicit references to a specific cultural context, usually that of the test author, in the instrument or its administration (Helms-Lorent et al., 2003). In this regard cultural complexity (c) was as important as g in the explanation of performance differences between cultural groups on cognitive tests.

In totality, the results support the proposition that culture moderates the relationship between General Reasoning and the OPP Social Conformity Scale. It also supports suggestions that this moderation may be attributed to social naïveté or conformity, and is likely to be connected to the level of cognitive ability of the respondent (Dilchert & Ones, 2005; Greene, 2000; Tett et al., 2011). In addition, Mueller-Hanson et al. (2006) provided evidence that in order to distort responses successfully the respondent must be able and motivated to distort responses. The ability to distort responses is connected to the analytical ability to apply problem-solving

to understand the construct being measured, and also to understand the advantages of faking behaviour (English et al., 2005; Kunzel & Hezlett, 2010; Tett et al., 2011). Thus, individuals high in g recognise and solve problems more successfully than those low in g. In addition, individuals high in g understand the items in the instrument, can detect desirable answers and respond accordingly. Research evidence has also shown that intelligence does not predict response distortion but that job knowledge moderates the ability to fake (Tett & Simonet, 2011). In support of the ability and also the motivation to fake, Wrensen and Biderman (2005) reported that social desirability was negatively related to faking ability, as those high in social desirability obtained the lowest faking ability scores. The results of this study are consistent with the findings of previous research, which suggest that social desirability scales may be an ambiguous indicator of faking as the scales may indicate propensity for faking (tendency to fake) but not the ability to fake.

If one assumes that cognitive ability is a primary selection tool, then it appears that there is a substantial threat of adverse impact (with proportionally more white participants being selected than black participants). If on top of this high scores on social desirability are used to eliminate candidates suspected of faking, the adverse may be exacerbated (with proportionally even more white participants being selected).

Against the background that (1) there are group mean differences in social desirability scores, (2) there are large group mean differences in cognitive ability scores and (3) cognitive ability is differentially related to social desirability across the groups, it also appears unreasonable to apply uniform corrections for social desirability for all groups. Such corrections are likely to penalise individuals with lower cognitive ability scores who tend to give more socially desirable responses. Individual differences in social desirability are also not fully explained by General Reasoning; cultural differences also played a role. This is consistent with findings by Greene (2000) that suggested a link between cognitive ability and social desirability, as responding in a certain manner reflects a level of psychological sophistication informed by the level of education and socio-economic status, thus supporting the literature of acculturation (Johnson & Van de Vijver, 2003; Shuttleworth-Jordan, 1996).

Based on the discussion of the results, the practical implications of the study follow.

Practical implications

As alluded to in the introduction, one of the biggest concerns raised by practitioners in the use of personality inventories is the potential impact of socially desirable responding on selection decisions. The most popular strategy to address response distortion is the inclusion of social desirability scales in personality inventories. In applied settings Industrial Psychologists use these scales to eliminate sources of bias or

systematic error that are not relevant to the measured attribute, to identify applicants who are deliberately presenting themselves in a positive manner, to adjust personality scale scores to compensate for socially desirable responding and to flag potentially invalid personality profiles.

An important implication of this study is the confirmation that the relationship between social desirability and general reasoning is moderated by culture and language, with group differences in social desirability being more pronounced at the low general reasoning level. This suggests that social desirability scales may be an ambiguous indicator of faking as the scales may indicate propensity for faking (tendency to fake) but not the ability to fake.

The results of this study also suggest that it is ethically questionable to deny someone a job opportunity based on the proposed validity of the personality profile, as a result of a score on a social desirability scale. This type of ethical implication is rarely discussed on a practical level. For example, the use of social desirability scales in personality instruments means that the typical instructions on a personality inventory (there are no right or wrong answers on this test) are, therefore, not true (or even ethical) as different strategies are used to identify potential fakers and corrections are then made based on the results (Dilchert & Ones, 2011).

Finally, in terms of selection practice, this study provided evidence of potential adverse consequences of using social desirability scales to detect response distortion and to disqualify applicants from the selection process. This study reported large group mean differences in cognitive ability and also social desirability scores, with the differences more pronounced at the lower cognitive ability level. The practical implication is that the use of a social desirability scale could adversely impact black applicants in ways that are not job related. If multiple predictors, such as cognitive ability, are utilised for selection the adverse impact may be exacerbated (with proportionally even more white applicants being selected).

Limitations and recommendations

Although the study made significant contributions to the body of knowledge concerning social desirability in a multi-cultural context, several limitations should be noted and addressed in future research. Firstly, the study used a cross-sectional design and, therefore, the relationships between variables cannot be interpreted causally. It is recommended that future research make use of longitudinal analytical methods to explore how the impact of socially desirable responding unfolds over time.

Secondly, the study used secondary data and the researcher was, therefore, unable to control for data collection errors. The researcher's lack of control over the selection of samples and comparison groups and also the quality of the sampling frame (e.g. gender representation and level of education) should be noted. This could potentially influence the generalisability of the results across groups. In an attempt to counter the effects

of sample selection, a decision was taken to compare groups across race and language. However, although respondents may share a common language they may be separated by a large cultural distance that requires further research, especially as there are assumed to be large cultural distances between indigenous and western cultures in South Africa. The results of this study provided evidence that suggests that culture, socio-economic status, level of education and language are possible sources of item or test bias.

The literature review furthermore provided evidence that job-relevant predictor composites often contain cognitive ability measures that produce fairly substantial group mean-score differences, contributing to potential adverse impact. These findings were replicated in the current study. The challenge of achieving accurate predictions (criterion-related validity) whilst also achieving similar selection ratios for sub-groups (reduced adverse impact), requires further research. Using linear programming methods, De Corte, Lievens and Sackett (2007) proposed a procedure for forming a weighted composite that reduces adverse impact as much as possible, given a specified level of validity. It is recommended that future research be undertaken in the South African context to examine this procedure in order to understand the sensitivity of predictor weights on adverse impact and validity outcomes.

The use of multiple assessments is further seen as a best practice standard in applied settings and is recommended when using personality measures (Hough & Ones, 2001; Mueller-Hanson *et al.*, 2006). The identification of potential adverse impact resulting from the use of social desirability scales and also cognitive ability measures, highlights the importance of accumulating evidence regarding the impact of multiple predictors on selection decisions. To this end it is recommended that employers using personality inventories in high-stakes selection settings need to accurately assess the requirements of the work context (job analysis) to identify appropriate predictors that may or may not have adverse impact on some groups (see Hough & Oswald, 2008). The legal context in South Africa must also be taken into consideration, as the pressure to ensure job relevance, reliability, validity and lack of bias of instruments, administered as part of a selection battery, remains a priority. In order to make cross-cultural comparisons continuous research must be undertaken to establish the cross-cultural equivalence of assessment outcomes and to address possible causes of cultural bias.

Conclusion

Given the prevalent use of social desirability scales in personality assessment in South Africa, the study provided evidence that there are culture and language group mean differences in social desirability scores. Within the black group the Sotho and Nguni groups, and within the white group the Afrikaans and English groups obtained very similar scores. The data support the hypothesis that culture and language moderates the relationship between General Reasoning (cognitive ability) and OPP Social Conformity (social desirability).

Results further show that the relations between cognitive ability and social desirability differ across groups. For the English group there is virtually no relation, whereas for the remaining groups there is a trend where individuals with high cognitive ability tend to give less socially desirable responses. The results also show that the differences in group means for social desirability are not fully explained by differences in cognitive ability. Cultural differences appear to play a role above and beyond the role of differences in cognitive ability.

Cognitive ability is, therefore, differentially related to social desirability across culture and language groups and it appears unreasonable to apply uniform corrections for social desirability across culture and language groups. The differences in cognitive ability and social desirability mean-scores across the different culture and language groups can lead to differential selection ratios between groups and, thus, potentially to adverse impact. It is further evident from this study that the validity and fairness of social desirability scales to detect applicant faking in the operational setting should be seriously questioned.

In the South African context the following should be taken into account:

> [*It does not seem*] unreasonable to attribute at least some part of the systematic group-related differences, especially on the measure of cognitive ability, to a socio-political system that systematically denied the members of specific groups the opportunity to develop and acquire those crystallised abilities required to succeed on the criterion. (Theron, 2007, p. 114)

The solution lies in affirmative development interventions aimed at developing those attainments and dispositions needed to succeed. This will present numerous exiting and stimulating challenges to the IO psychologist in South Africa.

Acknowledgements

Competing interests

The author declares that she has no financial or personal relationship(s) that may have inappropriately influenced her in writing this article.

References

Abrahams, F., & Mauer, K.F. (1999). Qualitative and statistical impact of home language on responses to the items of the sixteen personality factor questionnaire (16PF) in South African context. *South African Journal of Psychology, 29*(2)*,* 76–86. http://dx.doi.org/10.1177/008124639902900204

Aguinis, H. (2004). *Regression analysis for categorical moderators*. New York, NY: The Guilford Press.

Aguinis, H., & Henle, C.A. (2003). The search for universals in cross-cultural organizational behaviour. In J. Greenberg (Ed.), *Organizational behaviour: The state of the science* (pp. 373–411). London, UK: Sage.

Aguinis, H., Henle, C.A., & Ostroff, C. (2001). Measurement in work and organizational psychology. In N. Anderson, D.S. Ones, H.K. Sinangil, & C. Viswesvaran (Eds.), *Handbook of industrial work and organizational psychology* (Vol. 1, pp. 27–50). London, UK: Sage. http://dx.doi.org/10.4135/9781848608320.n3

Ajzen, I. (1991). Theory of planned behaviour. *Organizational Behaviour and Human Decision Processes*, *50*, 179–211. http://dx.doi.org/10.1016/0749-5978(91)90020-T

Anderson, N., Ones, D.S., Sinangil, H.K., & Viswesvaran, C. (2001). *Handbook of industrial work and organizational psychology*. London, UK: Sage.

Ang, S., Rodgers, J.L., & Wänström, L. (2010). The Flynn Effect within subgroups in the U.S.: Gender, race, income, education, and urbanization differences in the NLSY-children data. *Intelligence*, *38*, 367–384. http://dx.doi.org/10.1016/j.intell.2010.05.004

Arneson, J.J., Sackett, P.R., & Beatty, A.S. (2011). Ability performance relationships in education and employment settings: Critical tests of more-is-better and the good-enough hypothesis. *Psychological Science, 22,* 1336–1342. http://dx.doi.org/10.1177/0956797611417004

Aurthur, W., & Glaze, R.M. (2011). Cheating and response distortion on remotely delivered assessments. In N.T. Tippins & S. Adler (Eds.), *Technology-enhanced assessment of talent* (pp. 99–152). San Francisco: Jossey-Bass Publication. http://dx.doi.org/10.1002/9781118256022.ch4

Barger, S.D. (2002). The Marlowe-Crowne affair: Short forms, psychometric structure and social desirability. *Journal of Personality Assessment, 79,* 286–305. http://dx.doi.org/10.1207/S15327752JPA7902_11

Barrick, M.R., & Mount, M.K. (1996). Effects of impression management and self-deception on the predictive validity of personality constructs. *Journal of Applied Psychology, 81,* 261–272. http://dx.doi.org/10.1037/0021-9010.81.3.261

Bedell, B., Van Eeden, R., & Van Staden, F. (1999). Culture as moderator variable in psychological test performance: Issues and trends in South Africa. *SA Journal of Industrial Psychology, 25,* 1–7. http://dx.doi.org/10.4102/sajip.v25i3.681

Berry, J.W., Poortinga, Y.H., Segall, M.H., & Dasen, P.R. (2002). *Cross-cultural psychology: Research and applications.* (2nd edn.). Cambridge, MA: Cambridge University Press. http://dx.doi.org/10.1177/10697102036003005

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*, 317–335. http://dx.doi.org/10.1111/j.1468-2389.2006.00354.x

Budd, R.J. (1991). *The OPP technical manual*. Bedfordshire, UK: Psytech International Limited.

Budd, R.J. (1993). *Manual general, critical & graduate test battery*. Letchworth Herts, UK: Psytech International Limited.

Burns, G.N., & Christiansen, N.D. (2011). Methods of measuring faking behavior. *Human Performance, 24,* 358–372. http://dx.doi.org/10.1080/08959285.2011.597473

Claassen, N.C.W. (1997). Culture differences, politics and test bias in South Africa. *European Review of Applied Psychology, 47,* 297–307.

Crowne, D.P., & Marlowe, D. (1960). A new scale of social desirability independent of psychopathology. *Journal of Consulting Psychology, 24*(4)*,* 349–354. http://dx.doi.org/10.1037/h0047358

Davies, S., Norris, D., Turner, J., & Wadlington, P. (2005, April). *Cheating, guessing, faking, and self-presentation in assessment responses.* Paper presented at the 20th Annual Conference of the Society for Industrial and Organizational Psychology, Los Angeles, CA.

De Beer, M. (2004). Use of differential item functioning (DIF) analysis for bias analysis in test construction. *SA Journal of Industrial Psychology, 30*(4)*,* 52–58. http://dx.doi.org/10.4102/sajip.v30i4.175

De Corte, W., Lievens, F., & Sackett, P.R. (2007). Combining predictors to achieve optimal trade-offs between selection quality and adverse impact. *Journal of Applied Psychology, 92,* 1380–1393.

De Klerk, G. (2008). Cross-cultural testing. In M. Born, C.D. Foxcroft, & R. Butter (Eds.), *Online readings in testing and assessment.* International Test Commission: http://www.intestcom.org/Publications/ORTA.php

Dilchert, S., & Ones, D.S. (2005, April). *Race differences in social desirability scale scores: Cognitive ability influences among job applicants*. Poster presented at the 20th Annual Conference of the Society for Industrial and Organisational psychology, Los Angeles, CA.

Dilchert, S., & Ones, D.S. (2011). Application of preventive strategies. In M. Ziegler, C. MacCann, & R.D. Roberts (Eds.), *New perspectives on faking in personality assessment* (pp. 177–200). New York, NY: Oxford University Press. http://dx.doi.org/10.1093/acprof:oso/9780195387476.003.0054

Dilchert, S., Ones, D.S., Viswesvaran, C., & Deller, J. (2006). Response distortion in personality measurement: Born to deceive, yet capable of providing valid self-assessments? *Psychology Science, 25*(3), 209–225.

Edwards, A.L. (1957). *The social desirability variable in personality assessment and research*. New York, NY: Dryden.

Ellingson, J.E., Heggestad, E.D., & Makarius, E.E. (2012). Personality retesting for managing intentional distortion. *Journal of Personality and Social Psychology, 102*, 1063–1076. http://dx.doi.org/10.1037/a0027327

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

English, A., Griffith, R., Yoshita, Y., Gujar, A., Malm, T., & Socin, R. (2005, April). *It's all how you frame it: The impact of instructional sets on applicant faking behaviour.* Paper presented at the 20th Annual Society for Industrial and Organisational Psychology Conference, Los Angeles, CA.

Ethical rules of conduct for practitioners registered under the Health Professions Act, 1974. (2006, 4 August). *Government Gazette*, p. 29079, Cape Town.

Foxcroft, C.D., & Roodt, G. (2009). *Introduction to psychological assessment in the South African context.* (3rd edn.). Cape Town, South Africa: Oxford University Press Southern Africa.

Goffin, R.D., & Christiansen, N.D. (2003). Correcting personality tests for faking: A review of popular personality tests and an initial survey of researchers. *International Journal of Selection and Assessment, 11,* 340–344. http://dx.doi.org/10.1111/j.0965-075X.2003.00256.x

Greene, R.L. (2000). *The MMPI-2: An interpretative manual.* (2nd edn.). Needhan Heights, MA: Allyn & Bacon.

Griffith, R.L., Chmielowski, T., & Yoshita, Y. (2007). Do applicants fake? An examination of the frequency of applicant faking behavior. *Personnel Review, 36,* 341–357. http://dx.doi.org/10.1108/00483480710731310

Griffith, R.L., & Robie, C. (2013). Personality testing and the 'F–Word' revisiting seven questions about faking. In N.D. Christiansen & R.P. Tett (Eds.), *Handbook of personality at work* (pp. 253–280). New York: Routledge.

Heine, S.J. (2005). Where is the evidence for pancultural self-enhancement? A reply to Sedikides, Gaertner, and Toguchi (2003). *Journal of Personality and Social Psychology, 89,* 531–538. http://dx.doi.org/10.1037/0022-3514.89.4.531

Heine, S.J., & Lehman, D.R. (1997). The cultural construction of self-enhancement: An examination of group-serving biases. *Journal of Personality and Social Psychology, 72,* 1268–1283. http://dx.doi.org/10.1037/0022-3514.72.6.1268

Helms-Lorentz, A., Van de Vijver, F., & Poortinga, Y.H. (2003). Cross-cultural differences in cognitive performance and Spearman's hypothesis: *G* or *c*? *Intelligence, 31,* 9–29. http://dx.doi.org/10.1016/S0160-2896(02)00111-3

Hogan, J., Barrett, P., & Hogan, R. (2007). Personality measurement, faking and employment selection. *Journal of Applied Psychology, 92*(5), 1270–1285. http://dx.doi.org/10.1037/0021-9010.92.5.1270

Holden, R.R., & Passey, J. (2010). Socially desirable responding in personality assessment: Not necessarily faking and not necessarily substance. *Personality and Individual Differences, 49,* 446–450. http://dx.doi.org/10.1016/j.paid.2010.04.015

Hough, L.M., & Ones, D.S. (2001). The structure, measurement, validity, and use of personality variables in industrial, work, and organizational psychology. In N. Anderson, D.S. Ones, H.K. Sinangil, & C. Viswesvaran (Eds.), *Handbook of industrial work and organizational psychology* (Vol. 1, pp. 233–377). London, UK: Sage. http://dx.doi.org/10.4135/9781848608320.n13

Hough, L.M., & Oswald, F.L. (2008). Personality testing and industrial-organisational psychology: Reflections, progress and prospects. *Industrial and Organisational Psychology, 1,* 272–290. http://dx.doi.org/10.1111/j.1754-9434.2008.00048.x

Johnson, T.P., & Van de Vijver, F.J.R. (2003). Social desirability in cross-cultural research. In J.A. Harkness, F.J.R. Van de Vijver, & P.P. Mohler (Eds.), *Cross-cultural survey methods* (pp. 195–204). New York, NY: Wiley.

Krosnick, J.A. (1999). Survey research. *Annual Review of Psychology, 50,* 537–567. http://dx.doi.org/10.1146/annurev.psych.50.1.537

Kuncel, N. R., & Hezlett, S. A. (2010). Fact and fiction in cognitive ability testing for admissions and hiring decisions. *Current Directions in psychological Science, 19,* 339–345. http://dx.doi.org/10.1177/0963721410389459

Li, A., & Bagger, J. (2006). Using the BIDR to distinguish the effects of impression management and self-deception on the criterion validity of personality measures: A meta-analysis. *International Journal of Selection and Assessment, 14,* 131–141. http://dx.doi.org/10.1111/j.1468-2389.2006.00339.x

McFarland, L.A., & Ryan, A.M. (2000). Variance in faking across noncognitive measures. *Journal of Applied Psychology, 85,* 812–821. http://dx.doi.org/10.1037/0021-9010.85.5.812

McFarland, L.A., & Ryan, A.M. (2006). Towards an integrated model of applicant faking behaviour. *Journal of Applied Social Psychology, 36,* 979–1016. http://dx.doi.org/10.1111/j.0021-9029.2006.00052.x

Meiring, D., Van de Vijver, F.J.R., Rothmann, S., & Barrick, M.R. (2005). Construct, item, and method bias of cognitive and personality measures in South Africa. *SA Journal of Industrial Psychology, 31,* 1–8. http://dx.doi.org/10.4102/sajip.v31i1.182

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,* 683–729. http://dx.doi.org/10.1111/j.1744-6570.2007.00089.x

Mouton, J. (2001). *How to succeed in your Master's and Doctoral studies: A South African guide and resource book.* Pretoria, South Africa: Van Schaik Publishers.

Mueller-Hanson, R.A., Heggestad, E.D., & Thornton, G.C. III. (2006). Individual differences in impression management: An exploration of the psychological processes underlying faking. *Psychology Science, 48*(3), 288–312.

Nunnally, J.C., & Bernstein, I.H. (1994). *Psychometric theory.* (3rd edn.). New York, NY: McGraw-Hill.

O'Connell, M.S., Kung, M., & Tristan, E. (2011). Beyond impression management: Evaluating three measures of response distortion and their relationship to job performance. *International Journal of Selection and Assessment, 19,* 340–351. http://dx.doi.org/10.1111/j.1468-2389.2011.00563.x

Odendaal, A. (2013). *The measurement of social desirability: A cross-cultural perspective.* Unpublished Doctoral Thesis, University of Johannesburg, South Africa.

Ones, D.S., Dilchert, S., & Visvesvaran, C. (2012). Cognitive ability. In N. Schmitt (Ed.), *The Oxford handbook of personnel assessment and selection.* New York: Oxford University Press.

Ones, D.S., Dilchert, S., Visweswaran, C., & Judge, T.A. (2007). In support of personality assessment in organizational settings, *Personnel Psychology, 60,* 995–1027. http://dx.doi.org/10.1111/j.1744-6570.2007.00099.x

Ones, D.S., Visweswaran, C., & Reiss, A.D. (1996). Role of social desirability in personality testing for personnel selection: The red herring. *Journal of Applied Psychology, 81,* 660–679. http://dx.doi.org/10.1037/0021-9010.81.6.660

Paulhus, D.L. (1984). Two-component models of social desirable responding. *Journal of Personality and Social Psychology, 46,* 598–609. http://dx.doi.org/10.1037/0022-3514.46.3.598

Paulhus, D.L. (2002). Socially desirable responding: The evolution of a construct. In H.I. Braun, D.N. Jackson, & D.E. Wiley (Eds.), *The role of constructs in psychological and educational measurement* (pp. 49–69). Mahwah, NJ: Erlbaum.

Rossė, J.G., Stecher, M.D., Miller, J.L., & Levin, R.A. (1998). The impact of response distortion on pre-employment personality testing and hiring decisions. *Journal of Applied Psychology, 83,* 634–644. http://dx.doi.org/10.1037/0021-9010.83.4.634

Roth, P.L., Bevier, C.A., Bobko, P., Switzer, F.S., & Tyler, P. (2001). Ethnic group differences in cognitive ability in employment and educational settings: A meta-analysis. *Personnel Psychology, 54,* 297–330. http://dx.doi.org/10.1111/j.1744-6570.2001.tb00094.x

Rothstein, M.G., & Goffin, R.D. (2006). The use of personality measures in personnel selection: What does current research support? *Human Resources Management Review, 16,* 155–180.

Rushton, J.P., & Jensen, A.R. (2010). The rise and fall of the Flynn effect as a reason to expect a narrowing of the Black–White IQ gap. *Intelligence, 38,* 213–219. http://dx.doi.org/10.1016/j.intell.2009.12.002

Sackett, P.R. (2011). Integrating and prioritizing theoretical perspectives on applicant faking of personality measures. *Human Performance, 24,* 379–385. http://dx.doi.org/10.1080/08959285.2011.597478

Schaap, P., & Vermeulen, T. (2008). The construct equivalence and item bias of the PIP/SPEEX conceptualization-ability test for members of five language groups in South Africa. *SA Journal of Industrial Psychology, 34*(3), 29–38. http://dx.doi.org/10.4102/sajip.v34i3.749

Schmidt, F.L., & Hunter, J.E. (2004). General mental ability in the world of work: Occupational attainment and job performance. *Journal of Personality and Social Psychology, 86*(1), 162–173. http://dx.doi.org/10.1037/0022-3514.86.1.162

Schmit, M.J., & Ryan, A.M. (1993). The big five in personnel selection: Factor structure in applicant and nonapplicant populations. *Journal of Applied Psychology, 78,* 966–974. http://dx.doi.org/10.1037/0021-9010.78.6.966

Shuttleworth-Jordan, A.B. (1996). On not reinventing the wheel: A clinical perspective on culturally relevant test usage in South Africa. *South African Journal of Psychology, 26*(2), 96–102. http://dx.doi.org/10.1177/008124639602600205

Snell, A.F., Sydell, E.J., & Lueke, S.B. (1999). Towards a theory of applicant faking: Integrating studies of deception. *Human Resource Management Review, 9,* 219–242. http://dx.doi.org/10.1016/S1053-4822(99)00019-4

Stephen, D.F., Welman, J.C., & Jordaan, W.J. (2004). English language proficiency as an indicator of academic performance at a tertiary institution. *SA Journal of Human Resource Management, 2*(3), 42–53. http://dx.doi.org/10.4102/sajhrm.v2i3.48

Taylor, T.R. (1994). A review of three approaches to cognitive assessment and a proposed integrated approach based on a unifying theoretical framework. *South African Journal of Psychology, 26,* 184–191. http://dx.doi.org/10.1177/008124639402400403

Tett, R.P., Freund, K.A., Christiansen, N.D., Fox, K.E., & Coaster, J. (2011). Faking on self-report emotional intelligence and personality test: Effects of faking opportunity, cognitive ability and job type. *Personality and Individual Differences, 52,* 195–201. http://dx.doi.org/10.1016/j.paid.2011.10.017

Tett, R.P., & Simonet, D.V. (2011). Faking in personality assessment: A multi-saturation perspective on faking as performance. *Human Performance, 24,* 302–321. http://dx.doi.org/10.1080/08959285.2011.597472

Theron, C. (2007). Confessions, scapegoats and flying pigs: Psychometric testing and the law, *SA Journal of Industrial Psychology, 33*(1), 102–117. http://dx.doi.org/10.4102/sajip.v33i1.260

Van de Vijver, F.J.R., & Rothmann, S. (2004). Assessment in multicultural groups: The South African case. *SA Journal of Industrial Psychology, 30,* 1–7. http://dx.doi.org/10.4102/sajip.v30i4.169

Van Zyl, E., & Visser, D. (1998). Differential item functioning in the Figure Classification test. *SA Journal of Industrial Psychology, 24*(2), 25–33. http://dx.doi.org/10.4102/sajip.v24i2.650

Viswesvaran, C., Deller, J., & Ones, D.S. (2007). Personality measures in personnel selection: Some new contributions. *International Journal of Selection and Assessment, 15,* 354–358. http://dx.doi.org/10.1111/j.1468-2389.2007.00394.x

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

Viswesvaran, C., Ones, D.S., Cullen, M. J., Drees, S.A., & Langkamp, K. (2003, April). Appropriate police officer norms on personality scales. In S. Spilberg & D.S. Ones, *Personality and work behaviours of police officers,* Symposium conducted at the Annual Conference of the Society of Industrial and Organizational Psychology, Orlando, FL.

Watkins, M.N., & Elliot, A. (1997). The relation between cognitive ability and psychological readiness of semi-skilled employees in the South African mining industry. *SA Journal of Industrial Psychology, 23*(1), 1–5. http://dx.doi.org/10.4102/sajip.v23i1.615

Wrensen, L.B., & Biderman, M.D. (2005, April). *Factors related to faking ability: A structural equation model application.* Paper presented at the 20th Annual Society for Industrial and Organisational Psychology Conference, Los Angeles, CA.

Ziegler, M. (2012). *Applicant faking: A look into the black box.* Matthias Humboldt-Universität zu Berlin, Germany. Retrieved January 03, 2012, from http://www.siop.org/tip/july11/06ziegler.aspx

Ziegler, M., & Bühner, M. (2009). Modeling socially desirable responding and its effects. *Educational and Psychological Measurement, 69,* 548. http://dx.doi.org/10.1177/0013164408324469

Ziegler, M., MacCann, C., & Roberts, R.D. (2011). 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). New York, NY: Oxford University Press. http://dx.doi.org/10.1093/acprof:oso/9780195387476.001.0001