Organizational Behavior and Human Decision Processes 116 (2011) 148–162

Contents lists available at ScienceDirect

Organizational Behavior and Human Decision Processes

journal homepage: www.elsevier.com/locate/obhdp

Overclaiming as a measure of faking

Mark N. Bing [a,*], Don Kluemper [b], H. Kristl Davison [a], Shannon Taylor [b], Milorad Novicevic [a]

[a] Department of Management, University of Mississippi, United States
[b] Department of Management, Northern Illinois University, United States

ARTICLE INFO

Article history:
Received 29 March 2010
Accepted 24 May 2011

Accepted by Paul Levy

Keywords:
Socially desirable responding
Response distortion
Impression management
Self-deception
Test faking
Test validity
Personality measurement
Suppression effect
Overclaiming

ABSTRACT

Researchers have recently asserted that popular measures of response distortion (i.e., socially desirable responding scales) lack construct validity (i.e., measure traits rather than test faking) and that applicant faking on personality tests remains a serious concern (Griffith & Peterson, 2008; Holden, 2008). Thus, although researchers and human resource (HR) selection specialists have been attempting to find measures which readily capture individual differences in faking that increase personality test validity, to date such attempts have rarely, if ever succeeded. The current study, however, finds that the overclaiming technique captures individual differences in faking and subsequently increases personality test score validity via suppressing unwanted error variance in personality test scores. Implications of this research on the overclaiming technique for improving HR selection decisions are illustrated and discussed.

Researchers are increasingly recognizing that measures of individual differences in response distortion, namely socially desirable responding scales such as impression management and self-deception scales, do not do what they were designed to do—they do not effectively measure either individual differences in intentional faking or in self-deception (Bing, LeBreton, Davison, Migetz, & James, 2007; Ellingson, Sackett, & Connelly, 2007; Griffith & Peterson, 2008; Holden, 2007, 2008; Paulhus, 2002). Specifically, evidence suggests that responses to social desirability scale items are largely a function of substantive personality traits, such as neuroticism, extraversion, or conscientiousness (Block, 1965; Cattell, Eber, & Tatsuoka, 1970; Lönnqvist, Paunonen, Tuulio-Henriksson, Lönnqvist, & Verkasalo, 2007; McCrae & Costa, 1983; Nicholson & Hogan, 1990; Smith & Ellingson, 2002). Social desirability scale scores have also been found to be related to interpersonal sensitivity (Holden & Fekken, 1989), or alternatively to other personality traits such as narcissism (Johnson & Hogan, 2006), depending upon the particular measure of socially desirable responding.

In the domain of applicant screening, research indicates that socially desirable responding scales do not serve to improve the validity of personality test scores because such scales measure substantive personality traits and are thus overly redundant with personality measures (e.g., Barrick & Mount, 1996; Ellingson, Sackett, & Hough, 1999; Hough & Oswald, 2008; McCrae & Costa, 1983; Ones, Viswesvaran, & Reiss, 1996). Particularly important is the fact that socially desirable responding scales do not suppress unwanted systematic error variance in personality test scores due to intentional response distortion (i.e., faking), and thus do not serve as viable suppressors (Borkenau & Ostendorf, 1992; Ones et al., 1996; Smith & Ellingson, 2002). Therefore, faking continues to be a concern as recent research has indicated that faking occurs in up to 50% of cases involving actual job applicants (Griffith, Chmielowski, & Yoshita, 2007). Furthermore, studies of situations that can motivate some test-takers to engage in faking (e.g., applicant settings) have indicated that these factors do harm personality test scores in terms of (a) their construct validity (Griffin, Hesketh, & Grayson, 2004; Schmit & Ryan, 1993; Stark, Chernyshenko, Chan, Lee, & Drasgow, 2001), (b) their criterion-related validity (Bing, Whanger, Davison, & VanHook, 2004; Douglas, McDaniel, & Snell, 1996; Holden, 2007, 2008), and (c) the hiring decisions they generate (Griffith et al., 2007; Mueller-Hanson, Heggestad, & Thornton, 2003; Rosse, Stecher, Miller, & Levin, 1998). Although researchers measuring faking via socially desirable responding scales have attempted to counter these assertions (e.g., Ellingson et al., 1999; Ones et al., 1996), their conclusions about the impact of faking on test validity are suspect when based on the use of these scales (Bing et al., 2007; Griffith & Peterson, 2008; Smith & Ellingson, 2002). Therefore, there is a nascent need to examine alternative measures of individual differences in personality test faking.

* Corresponding author. Address: Department of Management, School of Business Administration, University of Mississippi, MS 38677, United States.
E-mail address: mbing@bus.olemiss.edu (M.N. Bing).

0749-5978/$ - see front matter © 2011 Elsevier Inc. All rights reserved.
doi:10.1016/j.obhdp.2011.05.006

The purpose of the current study is to investigate whether a relatively new measure of individual differences capable of quantifying response bias and faking, the overclaiming instrument (Paulhus, Harms, Bruce, & Lysy, 2003), can enhance the criterion-related validity of personality test scores via serving as a suppressor. First, we summarize recent studies relevant to personality test faking, particularly those that have drawn differing conclusions about the impact of faking on personality tests. Second, we briefly review research on the overclaiming instrument, describing how foils (i.e., items listing nonexistent topics, authors, books, etc.) in this instrument could be used to capture individual differences in faking because, unlike subjective self-reports of social desirability, overclaiming provides an actuarial, objective measure of faking (i.e., claiming to know that which cannot be known). Third, we explain how overclaiming scale scores from foils could serve as a suppressor variable to enhance personality test score validity. Fourth, we utilize a criterion-related validity study within a between-subjects research design to test the overclaiming instrument's suppressor capabilities in comparison to those of socially desirable responding scales, and to determine if measuring faking with overclaiming enhances the validity of personality test scores. Finally, we conclude by discussing how these results can be used in job applicant settings to enhance the accuracy of personnel selection decisions based on overclaiming and personality test scores.

Recent research on the impact of faking on personality tests

Although a number of researchers agree that faking on self-reports of personality is inevitable (Morgeson et al., 2007) and harms personality test validity (Tett & Christiansen, 2007), other researchers claim that faking on personality tests is very limited (e.g., Hogan, Barrett, & Hogan, 2007). Of late, two studies using repeated measures designs in organizational contexts examined changes in situational factors that might lead to less or more faking on personality tests, and both of these studies concluded that faking or response distortion on such tests is limited (see Ellingson et al., 2007; Hogan et al., 2007). The Ellingson et al. (2007) study noted that changes did occur in some personality test scores across developmental and selection contexts, but these changes were not overly different from changes that occur within the same context across time. They further noted that these changes could largely be attributed to unreliability and test-taker maturation rather than to response distortion.

These findings of Ellingson et al. (2007) are noteworthy but limited for several reasons. First, if test-taker faking tendencies are, in part, stable tendencies within developmental or selection contexts, then when using a repeated measures design faking may occur in both contexts among some test-takers, and yet this would appear as test score stability over time rather than as test score change. Second, although Ellingson et al. (2007) characterized developmental contexts as having a low press for faking in comparison to selection contexts, there remains the possibility that many test-takers would view such developmental contexts as potential threats to organizational advancement if their scores on positive traits were not elevated. Thus, developmental contexts do not entirely remove the motivation to engage in faking. Third, testing participants twice to measure faking, although informative from a scientific perspective, is not very practical for organizations within applied selection contexts. Fourth, as a job performance criterion was not measured in the Ellingson et al. (2007) study, we cannot determine from their results if test faking harms criterion-related validity.

The second study by Hogan et al. (2007) examined changes in personality test scores across two selection contexts using a repeated measures design in which applicants applied for a job, were rejected, and then applied again for that same job 6 months later. Like Ellingson et al. (2007), Hogan et al. (2007) found consistency in the test scores across these occasions. However, given the motivation to fake is nearly identical on both selection occasions, the caveats detailed above with respect to the Ellingson et al. (2007) study are very applicable here. Specifically, those who desired the job when applying the first time, and engaged in faking, were very likely to fake a second time, and this would lead to test score consistency rather than to test score change. Additionally, as Hogan et al. (2007) did not measure a performance criterion, we still cannot determine from their study whether or not faking harms criterion-related validity.

The above conclusions are markedly different from those of other researchers. Both Rosse et al. (1998) and Griffith et al. (2007) investigated personality testing in organizational contexts and concluded that personality test faking does occur, inflates test scores, and substantially alters hiring decisions. Rosse et al. (1998) used a between-subjects research design unlike the repeated measures design used by Ellingson et al. (2007) and Hogan et al. (2007), and thus differences in findings could be attributed to this design factor. However, the differences in findings could also be due to more salient changes in the situational factor (i.e., applicant versus incumbent testing situations) used by Rosse et al. (1998). Further, Griffith et al. (2007) used a repeated measures design like Ellingson et al. (2007) and Hogan et al. (2007) did, and thus differences in findings cannot be attributed to the design factor. Griffith et al. (2007) also used a more salient situational factor, in which real job applicants who had taken a personality test for selection purposes in an organizational setting were subsequently tested with the same test under an honest instructional set in a research setting (i.e., rather than under another organizational setting). Using this repeated measures design, Griffith et al. (2007) found that scores under the organizational applicant setting were highly elevated relative to the honest instructional set in the research setting, and that this faking behavior should significantly impact hiring decisions.

Finally, two very recent studies by Arthur, Glaze, Villado, and Taylor (2010) used repeated measures designs to investigate test-taker faking. In their first study, participants first took a personality test as job applicants (Time 1, high stakes) or as incumbents (Time 1, low stakes), and then took the test again as research participants (Time 2, low stakes). Their results indicated that Time 1 scores were generally higher than Time 2 scores, not only for the participants who were applicants at Time 1 but also for those who were incumbents at Time 1, although the differences were smaller for the latter participants. In their second study, they used a procedure similar to that in their first study, except that the Time 1 participants were all job applicants (high stakes). Again, Time 1 scores were generally higher than Time 2 scores. Together these studies provide further evidence that faking can impact hiring decisions.

Evidently, a consensus does not exist as to the impact of faking on personality tests. However, all five of these particular studies (Arthur et al., 2010; Ellingson et al., 2007; Griffith et al., 2007; Hogan et al., 2007; Rosse et al., 1998) lacked a measure of performance, and therefore the impact of faking on personality test criterion-related validity could not be assessed. Therefore, those who agree that faking can harm personality test validity have continued to search for ways to detect faking and counteract its deleterious effects. For example, recent research using Item Response Theory (IRT) has shown some promise in terms of detecting faking (LaHuis & Copeland, 2009; Zickar, Gibby, & Robie, 2004), which may be used to improve predictions from personality tests (Maij-de Meij, Kelderman, & van der Flier, 2008). Other researchers have used computer-administered tests to acquire item response latencies that may detect faking (e.g., Holden, 1998; Robie et al., 2000).