Impact of electronic warnings on online personality scores and test-taker reactions in an applicant simulation

Gary N. Burns *, Jenna N. Fillipowski, Megan B. Morris, Elizabeth A. Shoda

Wright State University, 3640 Colonel Glenn Hwy, Dayton, OH 45435, United States

ARTICLE INFO

ABSTRACT

This study investigates the impact of different types of electronic warnings during a simulated job application assessment. Results indicated that negatively worded warnings and accusations worked better than positively worded warnings at blunting faking. Although there was some evidence that test-takers engaging in higher levels of faking heeded warnings more, warnings tended to decrease scores for all test-takers. While positive warnings motivated test-takers to perform well, negative warnings and accusations increased test-taking anxiety. Whereas past research has failed to find a relationship between warnings and perceived fairness, current results suggest that this relationship depends on the level of test-taker distortion.

1. Introduction

Despite the popular use of personality inventories, a concern with applied personality measurement is that job applicants will distort, or fake, their test behavior, resulting in an overly-positive score. Research widely supports this fear, with estimates indicating that somewhere between a quarter to half of surveyed job applicants distort their responses to personality measures (Donovan, Dwight, & Hurtz, 2003; Griffith, Chmielowski, & Yoshita, 2007; Landers, Sackett, & Tuzinski, 2011). While mixed results have led some to question if this type of faking is a concern (Hogan, Barrett, & Hogan, 2007; Hough, 1998; Ones, Viswesvaran, & Reiss, 1996), the evidence that faking is a threat to hiring decisions and construct validity is becoming more difficult to dismiss (Marcus, 2006; Mueller-Hanson, Heggestad, & Thornton, 2003; Rosse, Stecher, Miller, & Levin, 1998).

The assertion that faking might cause adverse outcomes is supported by the argument that when selection ratios are low, people who fake have a disproportionately greater chance of being hired, even if the criterion-related validity is unaffected (Marcus, 2006). Several studies have illustrated that faking can also have implications for hiring decisions (Christiansen, Goffin, Johnston, & Rothstein, 1994). Furthermore, claims that faking has no impact on predictive validity are weakened by the fact that scales designed to measure faking (e.g., social desirability) are correlated

to honest personality scores of interest (e.g., conscientiousness, agreeableness, and emotional stability) (McCrae & Costa, 1983; Ones et al., 1996). This makes it difficult to tell whether applicants who score high on social desirability scales are faking, or if they truly have elevated personality traits (Goffin & Christiansen, 2003), thus giving less credibility to the viewpoint that faking has little to no impact on criterion-related validity, and reigniting an interest in further exploring the antecedents and outcomes of response distortion.

The concern about faking has inspired a wide range of techniques to counteract applicant distortion. These techniques can generally be broken down into two broad categories, proactive or reactive, and a third that combines both. First, there are proactive attempts to keep applicants from engaging in faking behavior. Such methods include the use of subtle or forced-choice item formats (Christiansen, Burns, & Montgomery, 2005; Lautenschlager, 1994) or warning applicants not to fake (Dwight & Donovan, 2003; Landers et al., 2011; Schrader & Osburn, 1977). Second, there has been a wide range of reactive procedures, sometimes referred to as validity scales, to detect and correct for faking after the fact (Bing, Kluemper, Davison, Taylor, & Novicevic, 2011; Hough, Eaton, Dunnette, Kamp, & McCloy, 1990; Wales & Seeman, 1968). Unfortunately, the research to date on correcting for faking is not promising (Barrick & Mount, 1996; Christiansen et al., 1994; Ellingson, Sackett, & Hough, 1999). The use of social desirability scores as a proxy of faking behavior is strongly discouraged (Burns & Christiansen, 2006; Peterson, Griffith, Isaacson, O'Connell, & Mangos, 2011), but other techniques might have promise (i.e., Bing et al.'s, 2011).

* Corresponding author. Tel.: +1 (937) 775 4643.
E-mail addresses: gary.burns@wright.edu (G.N. Burns), jennafilipkowski@gmail.com (J.N. Fillipowski), morris.156@wright.edu (M.B. Morris), shoda.2@wright.edu (E.A. Shoda).

Of the proactive strategies, the easiest one to implement with existing procedures is warnings. Unfortunately, researchers have also expressed concerns about warnings. Several researchers have cautioned that warnings might negatively impact test-taker reactions (Tett et al., 2006; Zickar & Gibby, 2006). Tett et al. (2006) also cautioned that warnings might cue risk-takers to attempt faking whereas Zickar and Gibby (2006) warned that, over time, test-takers would become aware of this method and the limited capability to detect fakers accurately. These concerns led Tett et al. (2006) to recommend that these warnings might be best used as a supplement to other types of faking detection measures. Taking advantage of such pairing, it is possible to target warnings only to those test-takers that are believed to be engaging in faking behavior (see Fan et al., 2012; Landers et al., 2011).

A third technique combines proactive and reactive strategies to target interventions to test-takers that are believed to be engaging in faking behavior (see Landers et al., 2011). Fan et al. (2012) illustrates the impact of specifically targeting such electronic warnings to test-takers identified early in the testing process. While such targeted warnings can easily be delivered, our current understanding and measurement of faking limits the accuracy in the administration of these warnings. While research on alternative faking measures continues to develop, the large body of extant research indicates that such electronic warnings will be rife with false positive and false negatives. As such, it is necessary to understand the impact of electronic warnings at multiple levels of various indicators of distortion. In an attempt to increase knowledge in this area, we explored the use of three different types of electronically delivered warnings in a laboratory experiment. Taking advantage of the laboratory setting we were able to conduct a true experiment to examine the impact of warnings across participants engaging in different levels of faking without ethically compromising personnel processes.

1.1. Effects of warnings on test scores

Many studies have shown the effectiveness of introducing a warning statement during a personality test administration, with Landers et al. (2011) providing strong evidence that warnings reduced the amount of blatant extreme responding in real world settings. In a meta-analytic review of the warning literature, Dwight and Donovan (2003) examined the effectiveness of warning applicants not to fake on personality measures among different types of warnings. Overall, test-takers who received warnings had test scores that were .23 standard deviations lower on average than unwarned test-takers. They also found that the type of warning moderated the efficacy of the warnings. Warnings regarding identification (e.g., this test contains items to identify fakers) had very little effect on test-taking behavior, with only a .01 standard deviation difference between groups. Consequence based warnings (e.g., dishonesty will not be tolerated) were the most effective type of warning, reducing warned test-takers' scores by .30 standard deviations.

Whereas most warnings used in published research have been mild, there is also the capability of accusing applicants of faking (e.g., Landers et al., 2011). Although such an accusation is an extreme reaction, its use is not unprecedented in the field of personality testing. For example, it is recommended that practitioners interpreting MMPI profiles marked as invalid (i.e., a profile with an elevated *L*, *K*, or *F* scale) refrain from making interpretation conclusions (Butcher, 1977). One recommended solution is to inform applicants that their scores are uninterpretable and to have them complete the test again (Butcher, Morfitt, Rouse, & Holden, 1997; Cigrang & Staal, 2001). Whereas researchers with the MMPI have primarily focused on retesting due to defensiveness, Ellingson, Heggestad, and Makarius (2012) found that similar effects were observed when asking test-takers to retest due to intentional distortion. These results indicate that when applicants are accused of distorting their responses, a second test administration typically results in more accurate personality scores. It should be noted that while research on these types of accusations is rare, that this research has emerged indicates that some organizations have already begun exploring such procedures.

1.2. Potential negative effects of warnings

Although warnings might offer a method of controlling or suppressing faking behavior, there are two potential negative consequences of warnings. First, warnings might have a negative effect on test-takers' reactions to the testing process. Despite these concerns, warnings appear to have very little effect on test-takers reactions and attitudes toward employment screening measures that measure personality traits. Examining procedural justice perceptions, McFarland (2003) found no significant difference between warned and unwarned test-takers. Similarly, Converse et al. (2008) found that neither positive nor negative warnings influenced perceived test ease, test-taker expectations, satisfaction with the testing process, or belief in the tests. However, Converse et al. (2008) did find that the negative warnings resulted in a higher level of test-taker anxiety. In our study, we wanted to focus on the effects of an extreme negative warning, an accusation of faking, on test anxiety and perceptions of fairness.

Second, warnings might also distort the nature of the personality data gathered. Robson, Jones, and Abraham (2008) found that warning statements reduced the convergent validity of the personality dimensions. This is also consistent with results reported by McFarland (2003), who found that correlations among personality variables were stronger in unwarned conditions compared to warned conditions. Perhaps most concerning, recent experiments by Fan et al. (2012) and Ellingson et al. (2012) indicate that warnings or retesting could cause non-faking test-takers to reduce their scores, especially when these individuals have been falsely identified as distorting their responses on computer administered tests. Both Fan et al. (2012) and Ellingson et al. (2012) provide convincing evidence that when presented with warnings of intentional distortion, even test-takers not actively distorting their responses will reduce their scores. Although such applicant distortion is different than the phenomenon of faking, it is a form of response distortion that practitioners should be concerned with.

Based on these results, practitioners should carefully consider the potential trade-offs of using warning statements. Strong evidence suggests that blatant distortion will be reduced (Landers et al., 2011) and the overall impact of the situational pressure to fake will be blunted (Dwight & Donovan, 2003); however, this might not influence criterion related validity coefficients (Converse et al., 2008). Furthermore, use of these statements might also make test users more anxious and might result in the degradation of the construct validity of the personality measures (Robson et al., 2008; Vasilopoulos, Cucina, & McElreath, 2005) and inaccurate scores for honest test-takers (Ellingson et al., 2012; Fan et al., 2012). One goal of the current study is to examine how these warnings influence responses to personality screening instruments across test takers responding in different ways.

1.3. Effects of warning at different levels of distortion

Additionally, warnings might not be effective for all respondents. Tett et al. (2006) cautions that warnings might cue risk-takers to engage in faking behavior, whereas other applicants might not believe the message. In addition to these concerns, there is also reason to believe that warnings will have differential effects on applicants engaging in different levels of faking.