

# Candidate Faking on Personality Assessments

By Jeff Weekley, Ph.D., Kenexa

Over the past decade, personality measures have increased in popularity as predictors of performance. There are many reasons for this, including the emergence of the five-factor framework, research showing personality measures can have very useful levels of validity, and the tendency for such measures to demonstrate less adverse impact than measures of cognitive ability. With this increased use have come concerns about the “fakability” of personality inventories. Specifically, because most personality inventories are relatively transparent, concerns have been raised about the extent to which job applicants answer them honestly. Whether conscious or not, research suggests applicants are motivated to present the image most likely to be viewed positively by decision makers. Incumbents, who already have the job and who may be responding under “research-only” instructions, have less motivation to attempt to manage the impressions they make. The result of this “test-taker motivation” may well be exaggerated levels of conscientiousness, agreeableness and extraversion for many applicants.

To date, a large amount of research has been conducted on faking and personality measures. Researchers have looked at three main issues: mean score differences, measurement equivalence (are faked measures “psychologically the same” as honest measures), and validity. The real issues are the extent to which applicants do actually respond differently from non-applicants and what effects, if any, these differences have on the accuracy of the inferences drawn from the affected measures.

Looking only at studies comparing applicants with incumbents, researchers have consistently found higher mean scores among applicants than among similar groups of incumbents (e.g., Bass, 1957; Kirchner, Dunnette & Mousely, 1960; Michaelis & Eyesnk, 1971; Rosse et al., 1998; Robie, Zickar & Schmit, 2001; Stark et al., 2001; White et al., 2001). In a recent meta-analysis of the issue, Birkeland, Manson,

Kisamore, Brannick, and Liu (2003) examined 29 studies comparing applicants with non-applicants. Their results showed that across all jobs, applicants scored significantly higher on emotional stability, conscientiousness and openness. In general, these mean differences tend to be about one third of a standard deviation favoring applicants (Birkeland, et al., 2003; Hough et al. 1990; Hough, 1998b; White et al., 2001).

Concern has also been raised about the measurement equivalence of scales across applicant and incumbent samples (Stark, Chernyshenko, Chan, Lee, & Drasgow, 2001). If personality measures are not equivalent across applicant and incumbent situations, then relationships between personality measures and other variables (e.g., criterion-related validity) may not be the same across these situations. Studies comparing factor structures across settings have been less consistent than those looking at mean level effects. Where as some have found measures to be equivalent (e.g., Smith, Hanges & Dickson, 2000; Smith & Ellingson, 2002; Robie et al., 2001) others have found differences in factor structure across settings (e.g., Schmit & Ryan, 1993; Stark, et al., 2001). Overall, the research on factor structure suggests that applicant responses may be different from incumbent responses, but the magnitude of the differences is not always large or important.

Research examining the effects of testing situations on criterion-related validity tentatively suggests that predictive validity estimates are slightly lower than concurrent validity estimates. For example, Dunnette, McCartney, Carlson and Kirchner (1962) found applicant responses to a personality measure to have slightly less validity than employee responses. Ones, Viswesvaran and Schmidt (1993), in looking at personality-based integrity tests, found predictive validity coefficients of .31 and concurrent validity coefficients of .37. In her work on Project A, Hough (1998b) examined eight personality dimensions (affiliation, potency, achievement,

dependability, adjustment, agreeableness, intellectance and rugged individualism) and four criteria (job proficiency, educational success, counterproductive behavior and training success) in both predictive and concurrent research designs. Hough (1998b) estimated that "concurrent validity studies produce validity coefficients that are, on average, .07 points higher than predictive validity studies (pg. 143)." It is possible that predictive validities are slightly lower because only a subset of applicants fake, thereby changing the rank-order of those applicants in relation to those who did not fake. It is worth emphasizing that personality measures still offered practically useful levels of validity in predictive settings.

From the foregoing, it can be concluded that applicants' scores on personality measures are usually higher than are incumbents' scores (Birkeland, et al., 2003; Stark, et al., 2001; Robie, et al., 2001; Rosse, et al., 1998; White et al., 2001). The impact of taking a test in applicant settings on scale structure is less clear, with some finding substantial differences (Michaeli & Eysenck, 1971; Schmit & Ryan, 1993; Stark et al., 2001) and others finding small or trivial differences (Robie, et al., 2001; Smith, et al., 2000). The results of studies on the criterion-related validity of personality measures generally indicate that they work slightly better for incumbents than they do for applicants (Dunnette, et al., 1962; Hough, 1998; Ones, et al., 1993). In short, the evidence indicates that applicants do tend to fake personality measures a little, but the impact of this tendency does not significantly detract from their usefulness. ■

[www.kenexa.com](http://www.kenexa.com)  
866.391.9557