

USING WEB-BASED FRAME-OF-REFERENCE TRAINING TO DECREASE BIASES IN PERSONALITY-BASED JOB ANALYSIS: AN EXPERIMENTAL FIELD STUDY

HERMAN AGUINIS
The Business School
University of Colorado Denver

MARK D. MAZURKIEWICZ
Department of Psychology
Colorado State University

ERIC D. HEGGESTAD
Department of Psychology
University of North Carolina Charlotte

We identify sources of biases in personality-based job analysis (PBJA) ratings and offer a Web-based frame-of-reference (FOR) training program to mitigate these biases. Given the use of job analysis data for the development of staffing, performance management, and many other human resource management systems, using biased PBJA ratings is likely to lead to a workforce that is increasingly homogenous in terms of personality but not necessarily a workforce with improved levels of performance. We conducted a field experiment (i.e., full random assignment) using 2 independent samples of employees in a city government and found evidence in support of the presence of biases as well as the effectiveness of the proposed solution. Specifically, FOR training was successful at decreasing the average correlation between job incumbents' self-reported personality and PBJA ratings from .27 to .07 (administrative support assistants) and from .30 to .09 (supervisors). Also, FOR training was successful at decreasing mean PBJA ratings by $d = .44$ (administrative support assistants) and by $d = .68$ (supervisors). We offer the entire set of Web-based FOR training materials for use in future research and applications.

This research was conducted, in part, while Herman Aguinis was on sabbatical leave from the University of Colorado Denver and holding visiting appointments at the University of Salamanca (Spain) and University of Puerto Rico.

A previous version of this manuscript was presented at the 21st Annual Conference of the Society for Industrial and Organizational Psychology, Dallas, 2006. The data reported in this article were used, in part, by Mark D. Mazurkiewicz in completing his master's degree in industrial and organizational psychology at Colorado State University. We thank Mark J. Schmit for comments on a previous draft.

Correspondence and requests for reprints should be addressed to Herman Aguinis, Dean's Research Professor, Department of Management and Entrepreneurship, Kelley School of Business, Indiana University, 1309 E. 10th Street, Bloomington, IN 47405-1701; haguinis@indiana.edu, <http://mypage.iu.edu/~haguinis>.

Job analysis is a fundamental tool that can be used in every phase of employment research and administration; in fact, job analysis is to the human resources professional what the wrench is to the plumber

(Cascio & Aguinis, 2005, p. 111)

Job analysis is a fundamental tool in human resource management (HRM). Information gathered through job analysis is used for staffing, training, performance management, and many other HRM activities (Aguinis, 2009; Cascio & Aguinis, 2005). Accordingly, the accuracy of the information gathered during the job analysis process is a key determinant of the effectiveness of the HRM function. As such, there is an ongoing interest in the development of improved job analysis tools. One recent addition to the job analysis toolkit is the use of personality-based job analysis (PBJA; Hogan & Rybicki, 1998; Meyer & Foster, 2007; Meyer, Foster, & Anderson, 2006; Raymark, Schmit, & Guion, 1997). In contrast to more traditional job analysis methods that focus on tasks or behaviors, PBJA provides explicit links between a job and the personality characteristics required for that job. Following a general trend in the globalization of HRM practices (Bjorkman & Stahl, 2006; Cascio & Aguinis, 2008b; Myers et al., 2008), the use of PBJA is becoming increasingly popular in the United States (Sackett & Laczko, 2001) as well as in other countries around the world, including France (e.g., Touzé & Steiner, 2002) and Turkey (e.g., Sümer, Sümer, Demirutku, & Çifci, 2001).

The first purpose of this article is to derive and test theory-based hypotheses suggesting that PBJA ratings are vulnerable to cognitive biases (i.e., self-serving bias, implicit trait policies, social projection, false consensus, and self-presentation). Ignoring the operation of these biases has important implications because, overall, it may lead to a workforce that is increasingly homogenous in terms of personality but not necessarily a workforce with improved levels of performance. For example, if biased PBJA ratings lead to the conclusion that a trait is related to job performance when, in fact, this trait is not an important determinant, then a selection system created on the basis of the PBJA ratings would be expected to result in the hiring of individuals who would perform suboptimally.

The second purpose of this article is to use established principles derived from frame-of-reference (FOR) training to implement a Web-based intervention that mitigates the impact of these biases. To enhance our confidence in the obtained results, we implemented a true field experimental design including complete random assignment of participants to conditions. Also, to enhance confidence in the generalizability of our conclusions, our study included two independent samples (i.e., administrative support assistants and supervisors working in a city government).

Although we found some differences between the samples, results were fairly consistent in showing the effectiveness of FOR training. We describe the FOR training program in detail and make it available upon request so that it can be used in future PBJA research and applications.

Benefits of Using PBJA

There are several potential benefits associated with the use of PBJA. One of these benefits is related to the increased emphasis on customer service and emotional labor and the concomitant need to include personality characteristics in the job analysis process (Sanchez & Levine, 2000b). When employees engage in emotional labor, they facilitate positive customer interactions by showing appropriate emotions they may not feel, creating appropriate emotions within the self, or suppressing inappropriate emotions during the transaction (Ashforth & Humphrey, 1993; Morris & Feldman, 1996). Therefore, the selection of individuals who can successfully engage in emotional labor has become relevant for many organizations. Initial research linking personality and emotional labor has shown that personality characteristics are correlated with effective performance of emotional labor (Diefendorff, Croyle, & Gosserand, 2005) and therefore should be included in job analysis. In short, PBJA can allow for an identification of which personality characteristics may facilitate customer interactions.

A second potential benefit of using PBJA is based on the positive relationship between some personality characteristics and job performance in some contexts (Ones, Dilchert, Viswesvaran, & Judge, 2007; Tett & Christiansen, 2007). Although there is controversy regarding the value-added contribution of personality relative to other predictors of performance (Morgeson, et al., 2007), some personality traits, such as Conscientiousness, can be used as valid predictors for many different types of occupations (Ones et al., 2007). Using PBJA allows for a more clear identification of which traits may be better predictors of which facets of performance for various types of jobs.

There are additional potential advantages of using PBJA. For example, conducting a PBJA may lead to not only improved predictive validity but also improved face validity of personality assessments. Jenkins and Griffith (2004) found that using a PBJA instrument led to the development of more valid selection instruments and that job applicants perceived these instruments to be more job related compared to more general selection instruments with less perceived job relatedness. Second, it can be used to develop selection instruments for a variety of jobs regardless of their hierarchical position in the organization. For example, Sümer et al. (2001) conducted PBJA interviews and surveys to identify

personality characteristics needed for various officer positions in the Turkish Army, Navy, and Gendarme. Results indicated that there were some personality traits (e.g., Conscientiousness–self-discipline, Agreeableness–Extraversion) that were common to all officer positions. Third, PBJA may be particularly useful for cross-functional and difficult-to-define jobs that cannot be described in terms of simple tasks or discrete knowledge, skills, and abilities (Brannick, Levine, & Morgeson, 2007). These types of jobs are becoming increasingly pervasive in the 21st-century organizations given that a large number of jobs are departing from traditional conceptualizations of fixed jobs (Cascio & Aguinis, 2008b; Shippmann et al., 2000). In sum, there seems to be a compelling business as well as HRM best practices case for using PBJA.

Although there is considerable conceptual (e.g., Morgeson & Campion, 1997) and some empirical (e.g., Dierdorff & Rubin, 2007; Morgeson, Delaney-Klinger, Mayfield, Ferrara, & Campion, 2004) research on the factors that can affect job analysis ratings in general, little is known about factors affecting PBJA ratings. Thus, although PBJA is an attractive tool to practitioners, caution must be exercised in the use of that tool until more is known about the possible biases in the resulting ratings.

One explanation for the lack of systematic research on PBJA, and ways to improve its accuracy, is the widely documented gap between science and practice in HRM, industrial and organizational (I-O) psychology, and related fields (Aguinis & Pierce, 2008; Cascio & Aguinis, 2008a; McHenry, 2007; Rynes, Colbert, & Brown, 2002; Rynes, Giluk, & Brown, 2007). For example, Muchinsky (2004) noted that researchers, in general, are not necessarily concerned about how their theories, principles, and methods are put into practice outside of academic study. In fact, Latham (2007) recently issued a severe warning that “we, as applied scientists, exist largely for the purpose of communicating knowledge to one another. One might shudder if this were also true of another applied science, medicine” (p. 1031). On the other hand, Muchinsky (2004) noted that practitioners, in general, are deeply concerned with matters of implementation. The lack of systematic research on PBJA may be another indicator of the gap between science and practice in HRM and I-O psychology in general and in the area of job analysis in particular.

The Personality-Related Personnel Requirements Form (PPRF)

Raymark et al. (1997) developed the PPRF as a supplement to more traditional nonpersonality-based methods of analyzing jobs. The Raymark et al. (1997) study followed best practices, and although we are aware of one other PBJA instrument (i.e., Performance Improvement Characteristics by Hogan & Rybicki, 1998), the PPRF is arguably the

most credible peer-reviewed PBJA tool available in the public domain. In addition to the PPRF development team, the project included the participation of “44 psychologists with extensive knowledge of psychological aspects of work, personality theory, or both” (Raymark et al., 1997, p. 725). The sample used to gather evidence in support of the usefulness of the PPRF included job incumbents who had held their positions for more than 6 months, who were working more than 20 hours per week, and who were holding 260 demonstrably different jobs. Finally, evidence supporting the usefulness of the PPRF included its ability to differentiate among occupational categories and its reliability in describing jobs.

The PPRF is a worker-oriented job analysis method but goes beyond other worker- and task-oriented approaches in that it assesses the extent to which each of the Big Five personality traits is needed for a particular job. The Big Five is the most established and thoroughly researched personality taxonomy in work settings (Barrick & Mount, 2003; Ones et al., 2007), and it includes factors that are fairly stable over time and are considered to be fundamental characteristics of personality (Goldberg, 1993). Someone who is high in *Extraversion* is likely to be talkative, outgoing, affectionate, and social. *Conscientiousness* describes a person’s tendency to be organized, responsible, and reliable. *Emotional Stability* represents a person’s tendency to not worry excessively, be anxious, or insecure. *Openness to Experience* describes a person’s preference for the creative, artistic, and novel. *Agreeableness* describes an individual’s tendency to be trusting, modest, and generally good natured.

Prior to the development of the PPRF, those who wanted to generate hypotheses about the extent to which each of the Big Five personality traits was needed for a particular job had to make inferences based on data from other job analysis methods (Raymark et al., 1997). For example, if a task-based job analysis procedure indicated that incumbents spend most of their time operating equipment in the absence of others, one could make the inference that Extraversion was not a job-relevant or job-necessary trait. This process of linking personality characteristics to job tasks is haphazard, unsystematic, and dependent upon the job analyst’s knowledge of personality theory. Accordingly, Raymark et al. (1997) created the PPRF to directly assess Big Five personality traits relevant to a particular job. The PPRF consists of sets of behavioral indicators associated with the five personality traits. Respondents indicate the extent to which each behavioral indicator is relevant to the job under consideration. Like most job analysis instruments, the PPRF utilizes survey methodology as the data collection procedure and typically uses job incumbents as the source of data. After data collection is complete, averaged scores across respondents

indicate the extent to which each trait (or subdimension of each trait) is relevant to the job.

Biases in PBJA

As noted earlier, there is a body of literature on factors that can affect the validity of traditional (i.e., nonpersonality-based) job analysis ratings. In their conceptual article, Morgeson and Campion (1997) classified such potential factors into social sources (i.e., due to norms in the social environment) or cognitive sources (i.e., due to limitations of people as information processors). In an empirical test of some of the Morgeson and Campion (1997) propositions, Morgeson et al. (2004) investigated the potential impact of one particular source of bias: self-presentation. Self-presentation is a type of social source by which “individuals attempt to control the impressions others form of them” (Leary & Kowalski, 1990, p. 34). Morgeson et al. (2004) found evidence that self-presentation may be responsible for rating inflation, particularly in the case of ability statements. Although Morgeson and Campion (1997) and Morgeson et al. (2004) did not discuss the specific case of PBJA or the relationship between rater personality and PBJA ratings, we rely on and expand upon their work to argue that there are four cognitive processes that are likely to affect the accuracy of PBJA ratings: self-serving bias, implicit trait policies (ITPs), social projection, and false consensus. We discuss each of these next.

Self-Serving Bias and ITPs

PBJA ratings may be biased due to self-serving bias and ITPs. Self-serving bias is a tendency for individuals to assume that successful performance is due to personal, internal characteristics whereas failure is attributed to factors outside the control of the individual. ITPs are individually held beliefs about the causal effect of certain personality characteristics on effective performance.

Regarding the effect of self-serving bias, a typical finding is that participants assume greater personal responsibility for success when they are randomly assigned to a “success” group than failure when randomly assigned to a “failure” group (Duval & Silvia, 2002; Urban & Witt, 1990). Specifically related to the study of personality at work, Cucina, Vasilopoulos, and Sehgal (2005) found that students rating the job of “student” had a tendency to report that self-descriptive personality traits were necessary for successful academic performance. Thus, if individuals make the assumption that they are good performers on a particular job, they may take

this notion further and assume that their own personal traits are the best or even the only traits that are necessary to do a job correctly.

Regarding the biasing effect of ITPs, Motowidlo, Hooper, and Jackson (2006) provided empirical evidence that individuals who were agreeable and extraverted believed the relationship between these two traits and effectiveness was more strongly positive than those who were low on the two traits. These researchers also showed that individual differences in ITPs were related to trait-relevant behavior at work. These results suggest that individuals who were high on Extraversion would: (a) engage in extraverted behaviors at work, and (b) hold the implicit theory that extraverted behaviors are related to performance across a number of situations. If one of these individuals were asked to rate the relevance of extraverted behaviors to job performance, it seems likely that he or she would endorse a large number of these behaviors relative to another job incumbent who is not as high on the Extraversion trait. Thus, the ITPs demonstrated to date appear to be consistent with self-serving bias.

Social Projection and False Consensus

Social projection is a cognitive bias that leads individuals to expect others to be similar to themselves (Robbins & Krueger, 2005). In other words, individuals reference their own characteristics when making predictions of other people's behavior and use projection as a judgmental heuristic. In the context of PBJA, due to social projection, individuals completing the questionnaire may assume a greater degree of similarity in personality traits with others than warranted. Thus, PBJA ratings are likely to be more reflective of an individual's particular personality as compared to the traits needed for the job in general because the respondent would assume that his or her personality is similar to that of others.

The false consensus effect is similar to social projection because it involves a process through which people assume that their views, traits, and behaviors are indicative of the views, traits, and behaviors of others (Mullen et al., 1985). In a review of false consensus research, Marks and Miller (1987) noted that selective exposure to similar others or the availability of information in memory may lead an individual to conclude that his or her views, opinions, and styles are similar to those of his or her peers. The operation of the false consensus effect was confirmed by a study demonstrating that individuals tend to project their own job characteristics when rating jobs held by other people using the Job Diagnostic Survey (Oliver, Bakker, Demerouti, & de Jong, 2005).

Summarizing the preceding section on biases in PBJA, individuals are likely to believe that their personality traits lead to successful performance due to the operation of self-serving bias and ITPs, and that others are

similar to them in terms of personality traits due to social projection and false consensus effect. Note that self-presentation is a different mechanism compared to each of these four cognitive processes. Self-presentation takes place when a rater attempts to control the impressions of others, for example, by providing inflated ratings of the need for the same personality traits that he or she possesses. On the other hand, self-serving bias, ITPs, social projection, and false consensus may or may not involve an attempt to control the impressions of others. In addition, self-presentation appears to be mostly conscious and less relevant when ratings are private (Tedeschi, 1981), whereas each of the four cognitive mechanisms we have described is mostly unconscious and, in our study, participants received several reassurances that their ratings were strictly confidential. However, each of these five sources of bias appears to have complementary and somewhat overlapping effects because the predicted end result is similar in terms of the resulting bias: There is an expected inflation in the relationship between a rater's personality and the personality characteristics believed to be needed for his or her job.

In addition to PBJA ratings being potentially biased due to a relationship with the respondents' own personality characteristics, we argue that self-serving bias, ITPs, social projection, false consensus, and self-presentation can also have an additional effect: inflation of PBJA ratings. PBJA ratings may be upwardly biased by job incumbents' own standing on the Big Five personality traits. Specifically, because of the influence of the biasing factors described earlier (e.g., ITPs), raters may be more likely to endorse a large number of behavioral indicators related to their own traits. So, overall, and given that it is reasonable to assume that at least some of the job incumbents will identify with each of the Big Five, PBJA ratings should, on average, provide an exaggerated (i.e., upwardly biased) view of the personality traits needed for the job. In addition, even if individuals do not possess a particular trait, it is likely that they would attempt to protect their self-concepts by providing favorable information about their jobs because such information is likely to lead to others' positive impressions (Morgeson et al., 2004). In support of this logic, Morgeson et al. (2004) found an upward bias in people's endorsements of the abilities required for their jobs.

Thus far, our discussion has focused on the theory-based reasons why PBJA ratings may be biased in two different ways: (a) the relationship between PBJA ratings and raters' personality may be greater than it should be (i.e., inflation in correlations), and (b) PBJA ratings may be higher than they should be (i.e., inflation in means). These biases have important implications for practice. For example, practitioners using PBJA information to develop a selection test would end up selecting individuals who are similar to the current workforce in terms of personality but not necessarily

the best possible performers. Similarly, if PBJA information is used as input in creating standards and objectives in a performance management system (cf. Aguinis, 2009), employees would receive feedback suggesting that they should display behaviors that would make them more similar in terms of personality to other coworkers but not necessarily more effective on the job. In addition, they may have an exaggerated (i.e., upwardly biased) view of the extent to which personality traits are related to job performance. If this information were to be used in the creation of a selection system, then otherwise qualified individuals could be screened out for having personality scores that are “too low,” leading to false negative errors (cf. Aguinis & Smith, 2007).

Web-Based FOR Training

Job analysis best practices suggest that all individuals involved in the process should receive some form of training (Aguinis & Kraiger, 2009; Gael, 1988; McCormick & Jeanneret, 1988). Regarding job incumbents, who are the ones typically filling out the job analysis questionnaire, it is commonly recommended that training be offered regarding the kind of information the organization is seeking about the job and how to fill out the questionnaire correctly (Gael, 1988; Hakel, Stalder, & Van De Voort, 1988). Our intervention consists of the design and delivery of a Web-based FOR training program specifically created to mitigate the impact of biases hypothesized to affect PBJA ratings.

FOR training was originally developed for use in performance appraisal (e.g., Aguinis, 2009; Bernardin & Buckley, 1981; Pulakos, 1984). FOR training seeks to minimize rater biases by including the following steps: (a) providing raters with a definition of each rating dimension, (b) defining the scale anchors, (c) describing what behaviors were indicative of each dimension, (d) allowing judges to practice their rating skills, and (e) providing feedback on the practice (Cascio & Aguinis, 2005). FOR training participants learn which job-relevant behaviors are indicative of good or poor performance. Thus, FOR training provides a common frame of reference and a system for the raters to use. Several narrative and meta-analytic reviews suggest that FOR training is one of the most effective methods for increasing rater accuracy (Aguinis, 2009; Woehr & Huffcutt, 1994).

We are not aware of any FOR training application in the context of job analysis. However, given the evidence regarding the effectiveness of FOR training in the context of performance appraisal, we decided to adopt this approach in an attempt to mitigate the effects of biases in PBJA ratings. FOR training is likely to be effective because its goal is to impose a common mental framework on all raters. This solves a problem that PBJA and

performance appraisal share in common: the use of idiosyncratic standards during a rating task. In the context of performance appraisal, individuals in FOR training are provided with instructions regarding: (a) which dimensions constitute performance, (b) which behaviors are indicative of those dimensions, and (c) how to map those behaviors on to a rating scale. FOR training for a PBJA rating task would only differ from a performance appraisal FOR training in terms of the first step. In PBJA FOR training, individuals are not necessarily told which dimensions constitute performance because this is the purpose of the job analysis process (respondents are rating the job on these dimensions). Instead, respondents must be instructed on what constitutes the performance domain for a given job. In short, they are told that job performance consists of what *all* individuals do, on average, to perform their jobs successfully; they are given an opportunity to practice; and then receive feedback so they can improve the accuracy of their ratings. One key point is that job incumbents should not rely exclusively on their own experiences when assigning PBJA ratings. In short, if our intervention is successful, we should find evidence in support of the following hypotheses:

Hypothesis 1: The relationship between job incumbents' PBJA ratings and their own personality traits will be weaker for participants in a PBJA FOR training condition in comparison to those who receive the standard instructions.

Hypothesis 2: PBJA ratings will be lower for participants in a PBJA FOR training condition in comparison to those who receive the standard instructions.

Method

Participants

To enhance the confidence in our results in terms of internal validity and generalizability, we used two independent samples of employees working for a large city government in the western United States: an administrative support assistant sample (including three job categories) and a supervisor sample (also including three job categories). The administrative support assistant sample included 96 individuals (80 women, 83.33%; 2 individuals did not indicate their gender). Within the administrative support assistant job family, the first job category involves standard/intermediate-performance-level office support work ($n = 13$), the second job category involves performing a variety of full-performance-level office support work ($n = 37$), and the third one involves performing specialized and/or technical office support work that required detailed knowledge of the

specialized/technical area ($n = 46$). Mean job tenure for these employees was more than 6 years (i.e., 77.71 months, $Mdn = 60$ months) with a range of 3 to 365 months.

The supervisor sample included 95 individuals (51 women, 53.7%). Mean job tenure for these individuals was more than 13 years (i.e., 159.64 months, $Mdn = 141$ months) with a range of 4 to 520 months. The three job categories within the supervisor job family were supervisors of professionals ($n = 45$), supervisors of support ($n = 31$), and supervisors of labor/trade ($n = 19$). Note that the city government includes a wide variety of functional areas including park management, financial management, public safety, customer service, legal services, business development, and many others. Although members of the same job family, individuals in the supervisor sample varied in terms of hierarchy and included first-line supervisors as well as higher-level supervisors.

Measures

Job incumbent personality. We used the 50-item sample questionnaire included on the Web site for the International Personality Item Pool (IPIP; <http://ipip.ori.org/ipip/>) to gather self-assessments of the participants' personality. This set of IPIP items was designed to measure the Big Five personality traits and is publicly available for use in academic research. Each trait is assessed by 10 items. Participants responded to the items using Likert-type scales ranging from 1 = *very inaccurate* to 7 = *very accurate*. Sample items include the following: "I am the life of the party," "I feel little concern for others," and "I am always prepared." Scale scores were calculated as an additive composite of responses across the items for each scale. We used the standard instructions for administering the IPIP:

On the following pages, there are phrases describing people's behaviors. Please use the rating scale below to describe how accurately each statement describes you. Describe yourself as you generally are now, not as you wish to be in the future. Describe yourself as you honestly see yourself, in relation to other people you know of the same sex as you are, and roughly your same age. So that you can describe yourself in an honest manner, your responses will be kept in absolute confidence. *There is no way for us to identify individual respondents.* Please read each statement carefully, and then click on the bubble that corresponds to the response on the scale. *Using the following scale, select the response that best represents the accuracy level of each item.*

There is substantial evidence in support of the reliability and construct validity of the IPIP (Goldberg, 1999; Goldberg et al., 2006; Lim & Ployhart, 2006; Mihura, Meyer, Bel-Bahar, & Gunderson, 2003). In addition, in

this study, internal consistency reliability coefficients (i.e., Cronbach's α) were as follows for the supervisor and administrative support assistant samples, respectively: Emotional Stability, .84 and .82; Extraversion, .88 and .84; Openness, .81 and .80; Agreeableness, .81 and .75; and Conscientiousness, .76 and .73. As a further check on the psychometric properties of the personality scales, we conducted a confirmatory factor analysis at the item level using data from both samples to compare a five-factor model to a single-factor model. To assess the fit of each model, we examined three different fit indexes: comparative fit index (CFI), root mean square error of approximation (RMSEA), and the expected cross-validation index (ECVI). Because the models were not nested, our comparison of the five-factor and one-factor models was based primarily on a comparison of the RMSEA and the ECVI statistics. The five-factor model can be said to fit better to the extent that the RMSEA and the ECVI are smaller than for the single-factor model (Levy & Hancock, 2007). As expected, the five-factor model had an excellent fit (CFI = .97; RMSEA = .06; ECVI = 11.99). Moreover, the five-factor model had a better fit compared to a competing single-factor model (CFI = .93; RMSEA = .09; ECVI = 18.06).

PBJA tool. We used the PPRF (Raymark et al., 1997) to conduct the PBJA. As noted earlier, the development of the PPRF followed best scientific practices, it has been peer reviewed, and it reliably differentiates between occupations. The PPRF includes 12 sets of items that are broken down as follows: sets 1, 2, and 3 include 28 items that describe job behaviors related to Extraversion; sets 4, 5, and 6 include 24 items for Agreeableness; sets 7, 8, and 9 include 29 items for Conscientiousness; set 10 includes 9 items for Neuroticism (polar opposite to Emotional Stability); and sets 11 and 12 include 17 items for Openness to Experience. Therefore, the measure includes a total of 107 items meant to measure each of the Big Five factors. Sample items include the following: "persuade coworkers or subordinates to take actions (at first they may not want to take) to maintain work effectiveness," "start conversations with strangers easily," and "work until the task is done rather than stopping at quitting time." As suggested by Raymark et al. (1997), participants responded to each item on a Likert-type scale ranging from 0 = *not required* to 2 = *essential*. Scale scores were calculated as the mean item response across the items of each of the five trait scales.

In this study, internal consistency reliability coefficients (i.e., Cronbach's α) were as follows for the supervisor and administrative support assistant samples, respectively: Emotional Stability, .82 and .80; Extraversion, .90 and .90; Openness, .93 and .90; Agreeableness, .87 and .86; and Conscientiousness, .85 and .86. As a further check on the psychometric properties of the PBJA tool, we conducted confirmatory factor analyses

at the item level using data from both samples to compare a five-factor model to a single-factor model. As expected, the five-factor model (CFI = .92; RMSEA = .07; ECVI = 34.47) had a superior fit than the competing single-factor model (CFI = .87; RMSEA = .09; ECVI = 44.14).

Manipulation check. After completing the PPRF, all participants were asked about the reference they used when completing the questionnaire. Specifically, participants were asked “When filling out questions about your job, which of these did you think of?” and were presented with two response options: “mostly my own experiences” and “people in general who work this job.”

Demographic information. Participants were asked to indicate the length of time they had worked for the organization and their gender. The organization was concerned about confidentiality and privacy issues so we were not able to ask questions regarding other demographic characteristics. On the other hand, the organization was not concerned about us collecting tenure and gender information. As a consequence of not collecting any additional demographic information, participants were reassured that we would not be able to track their individual responses. Although initially not intentional on our part, we see the exclusion of any additional demographic information as an advantage of our procedures given concerns about intentional distortion in completing personality measures (Ellingson, Sackett, & Connelly, 2007).

Procedure and Experimental Design

Participants from both samples were recruited via an e-mail invitation sent by the organization’s head of human resources. The e-mail directed individuals to our study’s Web site and made them aware that their participation would enter them in a lottery where they could win \$100 (for supervisors) or free admission to a regional professional conference (for administrative support assistants). A follow-up e-mail was sent 2 weeks later. As an additional means of recruiting from the administrative support assistant population, flyers announcing the study were distributed at a work-related conference sponsored by the organization. Individuals were informed that they could participate in the study during work hours from their office computers or, alternatively, they could also participate from home or any other Internet-enabled location.

All study materials were presented and responded to over the Internet. When a participant visited the study Web site, he or she was given a brief introduction to the study. They then were given the option to continue, thereby giving their consent to participate in the study, or to exit and not participate in the study. Individuals choosing to participate began by indicating their job category. At this point, participants were

randomly assigned to the FOR training or standard instructions group. Participants completed the IPIP and the PPRF, the order of which was counterbalanced to eliminate possible order effects. Thus, our study was a true field experimental study including complete random assignment to conditions.

Note that for participants in the FOR training condition, the training procedure was always administered immediately prior to the completion of the PPRF regardless of whether the PPRF was administered before or after the IPIP. All participants completed the manipulation check question immediately after completing the PPRF. The demographic questions regarding tenure and gender were presented last. Following the demographic questions, participants were thanked for their participation and were directed to a follow-up Web page that assigned them a random lottery number that they could print. We announced the winning numbers after the study was completed and the holders of these numbers were able to claim their prizes using the number they had printed previously.

Standard instructions condition. Participants who were randomly assigned to the standard instructions condition received the usual instructions that accompany the administration of the PPRF. Specifically, participants read the following information before completing the PBJA tool:

This inventory is a list of statements used to describe jobs or individual positions. It is an inventory of “general” position requirements. These position requirements are general in that they are things most people can do; most of them can be done without special training or unique abilities. Even so, some of them are things that can, if done well, add to success or effectiveness in the position or job. Some of them may be things that should be left for others to do—not part of this position’s requirements.

Each item in this inventory begins with the words, “Effective performance in this position requires the person to . . .” Each item is one way to finish the sentence. The finished sentences describe things some people, on some jobs, should do. An item may be true for the position or job being described, or it may not be. For each item, decide which of these statements best describes the accuracy of the item for the position being analyzed:

Doing this is not a requirement for this position (Not Required)

Doing this helps one perform successfully in this position (Helpful)

Doing this is essential for successful performance in this position (Essential)

Show which of these describes the importance of the statement for your position by selecting the response under “Not Required,” “Helpful,” or “Essential.”

FOR training condition. In the FOR training condition, participants were provided with a Web-based interactive training session on how to

respond to the items on the PPRF. We make the entire set of FOR training materials available free of charge upon request in html, text, or Microsoft Word format so that users can upload them on their own Web sites for future research or applications.

As noted earlier, our training program was based on theory and was guided by the principles of FOR training. Specifically, our Web-based training program defined the scale anchors clearly, included examples of which behaviors (of all people who do the job successfully) would be indicative of each item, allowed participants to practice providing ratings, and gave them feedback regarding their practice ratings.

The first series of Web pages introduced participants to the PPRF response scale and provided definitions for each of the response options. On the first of these pages, an example of a PPRF item (“Effective performance in this position requires the person to take control in group situations”) was presented. Participants then read a short passage above the sample item to introduce the response options. Specifically, participants read the following instructions:

The first response option is “Not Required.” Checking this response means this behavior is not necessary for satisfactory performance because it’s not really relevant for people in general doing your job. In the example below, let’s assume the person filling out this questionnaire works on an assembly line. Taking control in group situations doesn’t really apply to anyone doing this particular job. Therefore, “Not Required” would be a good answer.

The other two response options (i.e., “Helpful” and “Essential”) were explained in a similar manner.

Next, participants were given the following instructions:

Make sure you answer based on what everyone **MUST DO** and not just yourself. We’re trying to study the job in general, and not your specific style.

On this page also, participants were shown a sample PPRF item (“Effective performance in this position requires the person to keep your work area as organized as possible”) and were given the following information:

Consider the example item below: While I may be a top performer at the organization and I keep my work area organized, most people don’t always do this and they perform acceptably. Therefore, instead of rating this as “essential,” I will rate it as “helpful.”

On the next page, participants were given the following information:

You should not rely only on your personal experiences when responding to the items. Rather, you should think about everyone who does your job.

Consider the item below: While you may be an outstanding performer and you have a clean work area, it does not mean that having a clean work area is 'essential' for doing the job well. If your work area were not tidy, would you still be able to do the job effectively?

On the next series of Web pages, participants were given a chance to answer an item and receive feedback. The first of these pages read as follows:

Try it out. Click on one of the bubbles below to answer the question and then press "submit" for feedback!

Participants were then shown another sample PPRF item ("Effective performance in this position requires the person to: . . . develop new ideas") and asked to provide a response. Depending upon which response anchor the participant selected, he or she was provided with appropriate feedback. For example, if the participant selected the option "Not Required," the next page displayed the following text:

You selected "Not Required." This means for your job, developing new ideas is not essential for effective job performance. In other words, you do not need to develop new ideas to be considered a satisfactory employee."

The other two response options were followed with similar information. Finally, participants read the following statement:

This concludes our training! Thank you for your participation. Please click on the "Submit" button to begin filling out the actual survey.

Results

First, we describe analyses and results regarding the similarity of members in the FOR training and standard instructions groups on key individual characteristics and the extent to which FOR training was effective at changing the reference point from self to people in general in completing the PPRF (i.e., manipulation check). These analyses are important because they provide evidence regarding whether our intervention was successful and whether differences between groups can be attributed to our intervention or other extraneous factors including job incumbents' personality traits, gender, or tenure with the organization. Second, we provide descriptive statistics and results of an exploratory interrater reliability analysis. Finally, we describe analyses and results regarding the test of each of our substantive hypotheses.

Base-Line Similarity of FOR Training and Standard Instructions Groups

Our first set of analyses involved gathering evidence to rule out that substantive study results are due to extraneous factors and not our study's use of FOR training. Specifically, we conducted a multivariate analyses of variance (MANOVA) with each of the two independent samples to evaluate the similarity of the FOR training (i.e., experimental) and the standard instructions (i.e., control) groups. For each of the two MANOVAs, we used group (i.e., experimental vs. control) as the independent variable and scores on the five IPIP scales, months of tenure, and respondent gender as the seven dependent variables.

As expected, given that individuals were randomly assigned to conditions, the multivariate main effect was not statistically significant in the administrative support assistant sample ($F_{6,88} = .33, p > .05$, partial $\eta^2 = .02$). Nevertheless, to further evaluate the similarity of the groups, we conducted post hoc univariate analyses (i.e., ANOVAs). Each of the seven tests assessing possible differences between the groups was not statistically significant ($p > .05$).

Similarly, the multivariate main effect was not statistically significant in the supervisor sample ($F_{6,88} = 1.21, p > .05$, partial $\eta^2 = .08$). We conducted the seven follow-up ANOVAs and found a statistically significant main effect for Extraversion ($F_{1,93} = 4.82, p < .05$, partial $\eta^2 = .05$). However, given that the initial omnibus test was not statistically significant and that we conducted seven post hoc tests each using $\alpha = .05$, the presence of only one statistically significant result could be explained by chance alone. Specifically, applying a simple Bonferroni correction would lead to $\alpha = .05/7 = .007$. Using this corrected type I error rate leads to the conclusion that the effect for Extraversion is also not statistically significant.

In sum, results indicate that possible differences between FOR training and standard instructions groups are not explained by differences in job incumbents' personality traits, gender, or tenure with the organization.

Manipulation Check

Although we did not obtain the ideal result (i.e., 100% of raters in the FOR training condition used people in general as the referent point), an examination of the responses to the manipulation check question suggests that FOR training was successful in changing the reference point used by respondents in completing the PPRF. For the administrative support assistant sample, 64.8% of participants who received FOR training indicated they thought of people in general when responding to the PPRF items. On the other hand, only 22.0% of participants in the standard instructions

group indicated that they thought of people in general. Results of a formal statistical test indicated that the proportion of participants endorsing the “people in general” option was statistically significantly different across the two conditions ($\chi^2_1 = 17.22, p < .01$). We found a similar result in the supervisor sample. Specifically, 62.5% of participants who received FOR training indicated they thought of people in general when responding to the PPRF items, whereas only 29.1% of participants in the standard instructions group indicated that they thought of people in general. This difference in the proportion of participants endorsing the “people in general” option across the two conditions was also statistically significant ($\chi^2_1 = 10.54, p < .01$).

We computed effect sizes in the form of odds ratios to get further insight into the practical significance of the results regarding the manipulation check (Cohen, 2000). The odds ratio is a good indicator of the size of the effect of an intervention, particularly with dichotomous dependent variables. Based on the percentages reported above, the odds ratio for the administrative support assistant sample is $64.8/22.0 = 2.95$. This means that the odds of a participant in the FOR training group thinking about people in general when providing job analysis ratings was 2.95 times greater than the odds of a participant in the standard instructions condition thinking about people in general. The odds ratio for the supervisor sample was 2.15, meaning that the odds of a supervisor in the FOR training group thinking about people in general when providing job analysis ratings was 2.15 times greater than the odds of a participant in the standard instructions condition thinking about people in general. In sum, results based on descriptive statistics (i.e., percentages), tests of significance (i.e., χ^2 s), and effect sizes (i.e., odds ratios) support the effectiveness of the experimental manipulation.

Descriptive Statistics and Interrater Reliability Analysis

Means, standard deviations, and correlations for all study variables for the administrative support assistant and supervisor samples are included in Tables 1 and 2, respectively. In each of these tables, the correlations below the main diagonal are those for the standard instructions group, and the correlations above the main diagonal are those for the FOR training group. The main diagonals include internal consistency reliability (i.e., alpha) estimates for each sample combining the control and experimental groups.

We computed intraclass correlations (ICCs) to examine the degree of interrater agreement on the job analysis ratings. Our study included a total of six jobs grouped into two categories (i.e., administrative support assistants and supervisors). Thus, this was an exploratory analysis

TABLE 1
Means, Standard Deviations, Correlations, and Reliabilities for the Administrative Support Assistant Sample

Variable	M	SD	1	2	3	4	5	6	7	8	9	10	11	12
1. Tenure	77.71	71.38	^a	.06	.01	-.10	-.14	-.17	-.01	.16	-.01	.06	.21	.12
2. Gender	1.86	.35	.29	^a	-.25	-.24	.01	-.23	-.07	-.18	-.24	-.02	-.23	-.15
IPIP scales														
3. E	33.03	7.80	-.04	.07	.88	.37**	.07	.40**	.03	.05	-.06	-.12	.01	-.01
4. A	42.15	4.77	.18	.27	.21	.75	.21	.23	.13	-.03	.06	.10	.27*	-.07
5. C	41.44	5.02	.04	-.11	.06	-.02	.73	.41**	.10	.10	.15	.30*	.00	.12
6. ES	36.86	6.88	.03	.01	.48**	.00	.43**	.84	.14	-.18	-.16	-.14	-.13	-.24
7. O	37.36	5.88	-.24	-.58**	.08	-.06	.26	.15	.80	-.08	.04	.10	-.02	.05
PPRF scales														
8. E	1.01	.34	-.06	-.07	.27	.24	.15	.14	.36*	.90	.67**	.43**	.33*	.69**
9. A	1.19	.33	-.07	-.15	.34*	.30	.09	.14	.22	.63**	.86	.62**	.52**	.69**
10. C	1.44	.29	-.01	-.18	.15	.11	.29	.04	.23	.37**	.61**	.86	.57**	.68**
11. ES	1.41	.42	-.08	-.11	.28	.33	.03	.05	.23	.32*	.40**	.54**	.80	.50**
12. O	1.04	.38	-.20	-.10	.23	.20	.07	.09	.45**	.60**	.61**	.54**	.67**	.90

Notes. Correlations below main diagonal are for the standard instructions group ($n = 41$) and correlations above main diagonal are for the FOR training group ($n = 54$). Means, standard deviations, and internal consistency reliability coefficients are shown for entire sample. IPIP = international personality item pool, PPRF = personality-related personnel requirements form, E = Extraversion, A = Agreeableness, C = Conscientiousness, ES = Emotional Stability, O = Openness. Gender was coded as 1 = male, 2 = female; tenure is expressed in months.

^aSingle-item measure.
* $p < .05$, ** $p < .01$.

TABLE 2
Means, Standard Deviations, Correlations, and Reliabilities for the Supervisor Sample

Variable	<i>M</i>	<i>SD</i>	1	2	3	4	5	6	7	8	9	10	11	12
1. Tenure	159.64	108.80	^a	-.10	-.34*	-.12	.03	.10	-.03	.21	.29	.33*	.24	.26
2. Gender	1.54	.50	.05	^a	.36*	.53**	.35*	-.04	.29	.05	-.08	.07	.08	-.16
IPIP scales														
3. E	33.61	6.62	-.07	.10	.84	.52**	-.05	-.29	.20	-.00	.05	-.15	-.12	.02
4. A	40.14	5.74	-.08	.41**	.30*	.81	.37*	-.23	.34*	.01	.11	-.14	.05	-.22
5. C	39.48	5.62	-.14	.25	.16	.13	.76	.12	.26	-.08	-.19	.06	.06	-.29
6. ES	36.82	6.19	-.05	.14	.31*	.09	.32*	.82	-.09	.16	-.10	.29	.09	.09
7. O	38.48	5.74	-.21	-.01	.24	.21	.19	.03	.81	.05	.06	.12	.36*	.17
PPRF scales														
8. E	1.46	.31	.04	-.15	.13	.04	.31*	.29*	.51**	.90	.45**	.52**	.53**	.67**
9. A	1.24	.32	.01	.16	.19	.20	.21	.41**	.42**	.60**	.87	.58**	.63**	.41**
10. C	1.49	.26	-.13	.06	.09	.06	.40**	.38**	.07	.36**	.55**	.85	.63**	.52**
11. ES	1.46	.38	-.23	.04	.20	.08	.24	.27*	.19	.46**	.62**	.50**	.82	.47**
12. O	1.35	.41	-.18	.00	.20	.15	.39**	.33*	.51**	.73**	.63**	.42**	.55**	.93

Notes. Correlations below main diagonal are for the standard instructions group ($n = 55$) and correlations above main diagonal are for the FOR training group ($n = 40$). Means, standard deviations, and internal consistency reliability coefficients are shown for the entire sample. IPIP = international personality item pool, PPRF = personality-related personnel requirements form, E = Extraversion, A = Agreeableness, C = Conscientiousness, ES = Emotional Stability, O = Openness. Gender was coded as 1 = male, 2 = female; tenure is expressed in months.

^aSingle-item measure.

* $p < .05$, ** $p < .01$.

TABLE 3
*Intraclass Correlations (ICC) for PPRF (Personality Traits Needed for the Job)
 Ratings by Job and Training Condition*

	Standard instructions group			FOR training		
	<i>n</i>	ICC	Adjusted ICC	<i>n</i>	ICC	Adjusted ICC
ASAs (standard/intermediate performance-level office support)	5	.37	.37	8	.72	.62
ASAs (full-performance-level office support)	18	.88	.67	18	.80	.53
ASAs (specialized and/or technical office support work)	18	.86	.62	27	.91	.64
Supervisors of professionals	19	.69	.37	26	.76	.38
Supervisors of support	22	.79	.46	9	.66	.51
Supervisors of labor/trade	14	.74	.50	5	.29	.29

Notes. ASA = administrative support assistant, *n* = number of raters. Adjusted ICCs were computed using the Spearman–Brown formula to a case of *n* = 5.

because it would be easier to determine whether FOR training increases the reliability of ratings if more jobs were included in the analysis. Nevertheless, we expected that the degree of interrater agreement would be at least as high for the FOR training as compared to the standard instructions condition. Stated differently, if FOR training is effective, trained raters should be more interchangeable (Morgeson & Campion, 1997) than raters who have not received training and are being differentially affected by biasing factors (Voskuijl & van Sliedregt, 2002). Given that we had more than one rater in each job and the raters were considered a random set of possible raters, our situation is what has been labeled Case 2 (ICC2, *k*; Aguinis, Henle, & Ostroff, 2001; Shrout & Fleiss, 1979). The ICCs for the experimental and control groups for each of the six positions are included in Table 3. This table also shows that the number of raters within each position varied from as few as five to as many as 27. Because reliability estimates increase as the number of raters increases, the 12 ICCs are not directly comparable. Accordingly, we adjusted each of the ICCs using the Spearman–Brown formula to estimate the reliability of a mean rating based on five raters. We chose *n* = 5 as the adjustment factor because this is the smallest *n* across jobs. As shown in Table 3, the adjusted ICCs

TABLE 4
Correlations Between IPIP (Job Incumbents' Self-Reported Personality) and PPRF (Personality Traits Needed for the Job) Ratings

	Standard instructions group	FOR training group	Difference (Standard - FOR)	Test of difference (<i>z</i> or <i>t</i>)
Administrative support				
assistant sample				
Extraversion	.27	.05	.22	1.08
Agreeableness	.30	.06	.24	1.12
Conscientiousness	.29	.30*	-.01	-0.03
Emotional Stability	.05	-.13	.18	0.85
Openness	.45**	.05	.40	2.00*
Mean	.27	.07	.20	2.20*
Supervisor sample				
Extraversion	.13	-.00	.13	0.61
Agreeableness	.20	.11	.09	0.41
Conscientiousness	.40**	.06	.34	1.69*
Emotional Stability	.27*	.09	.18	0.87
Openness	.51**	.17	.34	1.82*
Mean	.30	.09	.21	2.92**

Notes. Tests of the difference between correlations are independent-sample *z* tests for individual traits and *t* tests (Neter, Wasserman, & Whitmore, 1988, p. 402) for mean correlations. IPIP = international personality item pool, PPRF = personality-related personnel requirements form.

For administrative support assistants, $n = 41$ (standard instructions), and $n = 54$ (FOR training); for supervisors, $n = 55$ (standard instructions), and $n = 40$ (FOR training).

* $p < .05$, ** $p < .01$.

were as large or larger in the FOR training condition than in the control condition for four of the six positions.

Tests of Substantive Hypotheses

Hypothesis 1 predicted that the FOR training program would be effective at decreasing the positive relationship between job incumbents' PBJA ratings and their self-reported ratings of their own personality traits. Table 4 displays results relevant to this hypothesis. For the administrative support assistant sample, the self-job correlation for Openness decreased by .40, the correlation for Agreeableness decreased by .24, the correlation for Extraversion decreased by .22, the correlation for Emotional Stability decreased by .18, and the correlation for Conscientiousness remained virtually identical (i.e., increased by .01). Across all of the personality traits, the overall correlation for self (i.e., IPIP) and job (i.e., PPRF) ratings was $\bar{r} = .27$ for the standard instructions condition and $\bar{r} = .07$ for the FOR

training condition ($t_8 = 2.20, p < .05$ for the difference between these correlations). This represents an average decrease of $\Delta\bar{r} = .20$. Note that all of the correlations in both conditions are positive except for Emotional Stability, which decreased from .05 in the standard instructions condition to $-.13$ in the FOR training condition. Although negative in value, this is not necessarily inconsistent with the theory-based prediction that, after being exposed to FOR training, participants will report a less positive relationship between their self-reported traits and the traits reported as needed for their jobs.

For the supervisor sample, the self-job correlations for Openness and Conscientiousness decreased by .34, the correlation for Emotional Stability decreased by .18, the correlation for Extraversion decreased by .13, and the correlation for Agreeableness decreased by .09. The mean correlations across the five traits were $\bar{r} = .30$ for the standard instructions condition and $\bar{r} = .09$ for the FOR training condition for the supervisor sample ($t_8 = 2.92, p < .01$ for the difference between these correlations). This represents an average decrease of $\Delta\bar{r} = .21$.

We squared the correlations to gain a better understanding of the meaning of the average decrease in each of the two samples. For the administrative support assistant sample, self-reported personality explained 7.29% (i.e., $.27^2$) of variance in PPRF ratings under standard administration conditions but only 0.49% (i.e., $.07^2$) of variance in PPRF ratings in the FOR training condition. For the supervisor sample, self-reported personality explained 9.0% (i.e., $.30^2$) of variance in PPRF ratings under standard administration conditions but only 0.81% (i.e., $.09^2$) of variance in PPRF ratings when our proposed FOR training program is used. Taken together, these results provide support for Hypothesis 1.

As is predicted by Schneider's (1987) attraction-selection-attrition model (ASA; see also De Fruyt & Mervielde, 1999; Ployhart, Weekley, & Baughman, 2006; Schaubroeck, Ganster, & Jones, 1998), as well as research on the gravitation hypothesis (Ones & Viswesvaran, 2003; Wilk, Desmarais, & Sackett, 1995), PBJA ratings should not be completely uncorrelated with incumbent personality, even for the FOR training group. Although the two job families (i.e., supervisors and administrative support assistants) include three job categories each, the two job groupings are separate job families, which means that job requirements are considered to be similar (but not identical) for the jobs within each family. Within a single homogeneous job, the correlation between personality and PBJA ratings should be close to zero in the FOR training condition. But, given that each of the two job families includes heterogeneous (yet similar) jobs, we would expect that FOR training would decrease correlations but not completely eliminate them. Results are consistent overall with this expectation.

Hypothesis 2 predicted that the FOR training intervention would decrease the PBJA ratings observed in the standard instructions condition. Results of analyses conducted to test this hypothesis are displayed in Table 5. This table shows that the mean ratings for the individual traits are higher for the standard instructions condition for each of the five traits in each of the two samples. Also, the standardized mean differences between the two conditions are $d = .44$ for administrative support assistants and $d = .68$ for supervisors.

In addition to the means and standardized mean differences effect sizes between conditions (i.e., d scores), Table 5 displays common language effect size statistics (CLs; McGraw & Wong, 1992). CLs, which are expressed in percentages, indicate the probability that a randomly selected score from one population will be greater than a randomly sampled score from the other population. To compute CLs, we used the d s displayed in Table 5 to obtain normal standard scores using the equation $z = d/\sqrt{2}$ and then located the probability of obtaining a z less than the computed value. For example, for the administrative support assistant sample, the overall CL across the five personality traits is 62%. This means that there is a 62% chance that an individual completing the PPRF using the standard instructions would provide a higher rating than if she or he were exposed to the FOR training prior to completing the PPRF. In terms of the other sample, Table 5 shows that there is a 68% chance that a supervisor will provide higher PPRF scores under the usual administration procedure compared to participating in our FOR training program prior to providing his or her PPRF ratings. Table 5 shows that, across the five personality traits, CLs are greater than 50% in each case and range from 57% to 67% for the administrative support assistant sample and from 66% to 74% in the supervisor sample. Table 5 also shows that, for the administrative support assistant sample, differences between means across conditions were statistically significant for Agreeableness ($t_{93} = 1.99, p < .05$), Conscientiousness ($t_{93} = 3.06, p < .01$), and Openness ($t_{93} = 2.73, p < .01$). For the supervisor sample, differences between means were statistically significant for each of the individual traits: Extraversion ($t_{93} = 2.92, p < .01$), Agreeableness ($t_{93} = 4.38, p < .01$), Conscientiousness ($t_{93} = 3.69, p < .01$), Emotional Stability ($t_{93} = 2.75, p < .01$), and Openness ($t_{93} = 2.69, p < .01$). Thus, we found support for Hypothesis 2.

Discussion

Given the central role of job analysis for most HRM activities and what appears to be a trend toward an increased use of PBJA tools, the purpose of our article was to describe a novel and practical application that solves a business problem. In terms of identifying the problem, we offered

TABLE 5
Means, Standard Deviations, and Effect Sizes for PPRF Scores by Condition Within Each Sample

	Administrative support assistant sample				Supervisor sample				<i>d</i>	<i>CL</i>		
	Standard instructions group		FOR training group		Standard instructions group		FOR training group					
	Mean	<i>SD</i>	Mean	<i>SD</i>	Mean	<i>SD</i>	Mean	<i>SD</i>				
Extraversion	1.07	.31	0.96	.36	.32	.59	1.54	.30	1.36	.31	.59**	.66
Agreeableness	1.27	.31	1.13	.34	.43*	.62	1.34	.28	1.08	.30	.90**	.74
Conscientiousness	1.54	.24	1.36	.31	.64**	.67	1.57	.22	1.38	.28	.77**	.71
Emotional Stability	1.47	.39	1.37	.44	.24	.57	1.55	.33	1.34	.42	.57**	.66
Openness	1.15	.41	0.95	.32	.55**	.65	1.44	.40	1.21	.39	.58**	.66
Mean	1.30		1.15		.44 ^a	.62	1.49		1.27		.68 ^a	.68

Notes. *SD* = standard deviation, *d* = standardized mean difference effect size, *CL* = common language effect size (expressed in percentages).

p* < .05, *p* < .01.

^aWe did not compute tests of statistical significance for the mean *d* scores because the number of *d*s used to compute each \bar{d} is only 5. For administrative support assistants, *n* = 41 (standard instructions) and *n* = 54 (FOR training); for supervisors, *n* = 55 (standard instructions) and *n* = 40 (FOR training).

theory-based hypotheses regarding biases operating in the process of conducting a PBJA. In terms of the solution, we designed and administered a FOR training program that weakened the relationship between incumbents' self-reported personality traits and the personality traits reported to be necessary for the job (Hypothesis 1). The FOR training program was also effective at lowering the PBJA scores (Hypothesis 2). Our study implemented a field experimental design including complete random assignment of participants to conditions to increase the confidence that our proposed intervention actually caused the intended outcomes. Also, we used two independent samples to increase the confidence that our proposed intervention can be used with a variety of jobs.

Implications for Practice

As noted in the opening quote of our article, "job analysis is to the human resources professional what the wrench is to the plumber" (Cascio & Aguinis, 2005, p. 111). Accordingly, if there is a problem with the data gathered using job analysis, then it is likely that there will be a problem with the many uses of these data ranging from the development and implementation of selection tools to the design and use of performance management and succession planning systems.

PBJA is a very promising tool. However, as noted by the authors of the instrument, "the PPRF is offered to both researchers and practitioners for use, refinement, and further testing of its technical merits and intended purposes" (Raymark et al., 1997, p. 723). This comment must also be considered within the broader context of what some consider an overemphasis on personality in staffing decisions (Morgeson et al., 2007). Consistent with Raymark et al.'s offer, our study focused on the technical merits of conducting a PBJA and identified two types of problems. First, ratings of personality traits deemed necessary for a particular job are overall related to the personality traits of the incumbents providing the ratings. In a sample of supervisors, their own personalities accounted for 9% of the variance in the PBJA ratings. In a separate sample of administrative support assistants, their own personalities accounted for a similar percentage of variance in the PBJA ratings (i.e., 7.29%). In terms of practice, this means that HRM interventions using data gathered via a PBJA may not produce the anticipated results in terms of their utility because organizations may make staffing decisions based on traits that may not be essential for the job (cf. Cascio & Boudreau, 2008). In addition, perhaps even more important at the organizational level, such HRM interventions are likely to produce a workforce similar to this workforce rather than a workforce that is more competent than this one. For example, using the resulting PBJA ratings to create a selection system would lead to selecting individuals

based on the personality traits of this workforce. Similarly, using PBJA ratings to design a 360-degree feedback system would lead to individuals receiving feedback that they should behave in ways that reflect similar personality traits as those of the current workforce. In short, using PBJA ratings is likely to lead to a workforce that is increasingly homogenous in terms of personality but not necessarily a workforce with improved levels of performance. We also hypothesized and found support for a second type of problem, that PBJA ratings may be exaggerated (i.e., upwardly biased). In terms of practice, this means that the increased workforce homogeneity problem is exacerbated.

Pointing to a problem in the absence of a proposed solution would be unhelpful. Accordingly, our proposed Web-based FOR training program helped to mitigate each of the problems described above. Addressing the first problem, the FOR training program decreased the relationships between self- and job ratings of personality to .07 for the administrative support assistant sample and .09 for the supervisor sample. The decrease in the average correlation across the five personality traits was similar across samples: .20 (administrative support assistants) and .21 (supervisors). In terms of the second problem, using our proposed FOR training decreased the mean PBJA rating score across the five personality traits by $d = .44$ for administrative support assistants and by $d = .68$ for supervisors. Another way to describe these results is to compute common language effect sizes, which indicated that, across the five personality traits, an individual providing PBJA under the typical administration condition would be 62% (administrative support assistants) and 68% (supervisors) more likely to provide a higher rating than if the same individual provided the PBJA ratings after participating in our FOR training. The FOR training effect was also similar across the two occupational types and, when considering individual traits, ranged from 57% to 67% for administrative support assistants and from 66% to 74% for supervisors.

Limitations and Research Needs

We note three potential limitations of our study as well as some directions for future research. First, although in the expected direction, some of the IPIP-PPRF correlations at the trait level were not statistically significant in the standard instructions condition. These results are possibly due to low statistical power given that we had a sample size of about 50 per cell but also to the fact that some of the traits may not be as needed for some types of jobs as compared to others. Related to this issue, results indicate some differences across samples in terms of the self-job correlations for the FOR training conditions. Although we

do not have data to test this possibility directly, these differences could also be due to actual differences regarding job requirements. To test this possibility more thoroughly, future research could include a research design like ours but expand the types and number of jobs to include several clearly distinct jobs ideally across different organizations and even types of industries. Accordingly, future research could test specific hypotheses regarding which traits are likely to show the largest correlations depending on the moderating effect of specific occupations and types of occupations and traits for which the implementation of our FOR training intervention is most effective. On a related issue, future research could also examine extensions of our FOR training intervention to other types of job analysis, including task-oriented (e.g., focused on knowledge, skills, and abilities) as well as physical-ability job analysis, for which it may seem especially difficult for raters to focus on people in general doing the job instead of themselves.

Second, focusing on one's own position or everyone doing the job is not intrinsic to the standard instructions or FOR training. In other words, the standard instructions could be written such that raters are asked to focus on how everyone does their job in general. However, the fact is that the standard instructions that accompany the PPRF and are used by anyone using the PPRF do not include this type of language. Until the publication of this article, users of the PPRF had no reason to believe PPRF ratings may be biased and, hence, also had no reason to believe that the instructions should be changed in any way. Thus, it is quite likely that every past use of the PPRF included the standard instructions and resulted in biased ratings. Nevertheless, is it possible that revising the standard instructions in an attempt to change the referent point may result in less biased ratings? Yes, this is certainly possible. However, note that the FOR training intervention involves more than merely changing the referent point from the rater to people in general. Specifically, FOR training goes beyond a mere change in instructions because it first defines the scale anchors clearly, then it includes examples of which behaviors (of all people who do the job successfully) would be indicative of each item; it also allows participants to practice providing ratings and, finally, gives them feedback regarding their practice ratings. The inclusion of all of these components serves the purpose of creating a common frame of reference among raters. We readily acknowledge that the manipulation check we used was limited and focused only on whether ratings were based on the raters' own experiences or people in general who work their jobs. However, in spite of the limited scope of our manipulation check, the preponderance of the evidence supports the conclusion that FOR training worked as intended and resulted in more accurate ratings, as is described next.

Third, we have argued that FOR training minimizes biases and, hence, produces more accurate PBJA ratings. The topic of rating accuracy is a central yet unresolved issue in the job analysis literature. For example, Morgeson and Campion (2000) noted that “The entire job analysis domain has struggled with questions about what constitutes accuracy. That is, how do we know that job analysis information is accurate or ‘true?’” (p. 819). Similarly, Sanchez and Levine (2000a) argued that “we must caution that a basic assumption of any attempt to assess JA [job analysis] accuracy is that there is some underlying ‘gold standard’ or unquestionably correct depiction of the job. This assumption is problematic at best, for any depiction of a complex set of behaviors, tasks, or actions subsumed under the label of the term job is of necessity a social construction” (p. 810). Given the absence of a gold standard or true score, Sanchez and Levine (2000a) and Morgeson and Campion (2000) argued for the use of an inference-based approach for understanding the extent to which job analysis ratings are accurate. Morgeson et al. (2004) used this approach to understand whether ability ratings were inflated compared to task ratings because, as they noted, “absent some true score, however, it is difficult to definitively establish whether these [ability] ratings are truly inflated” (p. 684). This inference-based approach is similar to the approach used in gathering validity evidence regarding a selection test (Binning & Barrett, 1989), which is based on the more general idea of validation as hypothesis testing (Landy, 1986). In addition, although not mentioned in the job analysis literature, the inference-based approach is also related to the concept of triangulation, which occurs when a similar conclusion is reached through different conceptualizations or operationalizations of the same research question (Scandura & Williams, 2000). Essentially, an inference-based approach to understanding the accuracy of job analysis ratings entails deriving theory-based expectations about how scores should behave under various conditions and assessing the extent to which these expectations receive support.

In our study, we had seven different expectations about the data that we collected. First, participants’ personality, job tenure, and gender should be unrelated to their assignment to either the standard instructions or FOR training conditions. Second, a majority of participants in the FOR training condition were expected to provide ratings using “other people” as opposed to “self” as the reference point. Third, self–job personality correlations were expected to be more strongly positive for the standard instructions condition compared to the FOR training condition. Fourth, mean ratings were expected to be higher for the standard instructions condition compared to the FOR training condition. Fifth, we expected that, overall, interrater agreement would be at least as high among trained raters as among untrained raters. Sixth, based on the discussion of the ASA

framework and the gravitation hypothesis, self–job correlations in the FOR training condition were not expected to be uniformly zero. Finally, also based on the ASA framework and the gravitation hypothesis, self–job correlations were not expected to be identical across job families (i.e., supervisors vs. administrative support assistants). The data conformed to each of these seven expectations. As an additional and more indirect type of evidence, a review of meta-analyses by Morgeson et al. (2007) found that the uncorrected correlations between Big Five personality traits and performance range from $-.02$ to $.15$. Note that in our study the average self–job correlations were $.27$ (administrative support assistants) and $.30$ (supervisors) in the standard instructions conditions. On the other hand, in the FOR training conditions, these correlations were $.07$ and $.09$, respectively, which puts the correlations virtually in the center of the range of trait–performance correlations described by Morgeson et al. (2007). In short, similar to Morgeson et al. (2004), we cannot make a definitive statement about the accuracy of ratings. However, taken together, the evidence gathered points to the effectiveness of the FOR training intervention.

Closing Comments

The overall purpose of our article was to address a contemporary issue in practice, a problem that practitioners face in applying research and theory in the real world, and also to present solutions, insights, tools, and methods for addressing problems faced by practitioners (cf. Hollenbeck & Smither, 1998). Our FOR training program is easy to implement and takes less than 15 minutes to administer online and proved effective at decreasing biases in PBJA ratings. We are making the entire set of FOR training materials available upon request free of charge so that future PBJA research and applications can benefit from it. In closing, given the central role of job analysis in HRM and I-O psychology practice, we hope our article will stimulate further research and applications in the area of PBJA.

REFERENCES

- Aguinis H. (2009). *Performance management* (2nd ed.). Upper Saddle River, NJ: Pearson-Prentice Hall.
- Aguinis H, Henle CA, Ostroff C. (2001). Measurement in work and organizational psychology. In Anderson N, Ones DS, Sinangil HK, Viswesvaran C (Eds.), *Handbook of industrial, work and organizational psychology* (Vol. 1, pp. 27–50). London: Sage
- Aguinis H, Kraiger K. (2009). Benefits of training and development for individuals and teams, organizations, and society. *Annual Review of Psychology*, 60, 451–474.

- Aguinis H, Pierce CA. (2008). Enhancing the relevance of organizational behavior by embracing performance management research. *Journal of Organizational Behavior*, 29, 139–145.
- Aguinis H, Smith MA. (2007). Understanding the impact of test validity and bias on selection errors and adverse impact in human resource selection. *PERSONNEL PSYCHOLOGY*, 60, 165–199.
- Ashforth BE, Humphrey RH. (1993). Emotional labor in service roles: The influence of identity. *Academy of Management Review*, 18, 88–115.
- Barrick MR, Mount MK. (2003). Impact of meta-analysis methods on understanding personality-performance relations. In Murphy KR (Ed.), *Validity generalization: A critical review* (pp. 197–221). Mahwah, NJ: Erlbaum.
- Bernardin HJ, Buckley MR. (1981). A consideration of strategies in rater training. *Academy of Management Review*, 6, 205–212.
- Binning JF, Barrett, GV. (1989). Validity of personnel decisions: A conceptual analysis of the inferential and evidential bases. *Journal of Applied Psychology*, 74, 478–494.
- Bjorkman I, Stahl G. (Eds.) (2006). *Handbook of research in international human resource management*. London: Edward Elgar.
- Brannick MT, Levine EL, Morgeson FP. (2007). *Job and work analysis: Methods, research, and applications for human resource management* (2nd ed.). Thousand Oaks, CA: Sage.
- Cascio WF, Aguinis H. (2005). *Applied psychology in human resource management* (6th ed.). Upper Saddle River, NJ: Pearson-Prentice Hall.
- Cascio WF, Aguinis H. (2008a). Research in industrial and organizational psychology from 1963 to 2007: Changes, choices, and trends. *Journal of Applied Psychology*, 93, 1062–1081.
- Cascio WF, Aguinis H. (2008b). Staffing 21st-century organizations. *Academy of Management Annals*, 2, 133–165.
- Cascio WF, Boudreau JW. (2008). *Investing in people: Financial impact of human resource initiatives*. Upper Saddle River, NJ: Pearson Education/Financial Times.
- Cohen MP. (2000). Note on the odds ratio and the probability ratio. *Journal of Educational and Behavioral Statistics*, 25, 249–252.
- Cucina JM, Vasilopoulos NL, Sehgal KG. (2005). Personality-based job analysis and the self-serving bias. *Journal of Business and Psychology*, 20, 275–290.
- De Fruyt F, Mervielde I. (1999). RIASEC types and Big Five traits as predictors of employment status and nature of employment. *PERSONNEL PSYCHOLOGY*, 52, 701–727.
- Diefendorff JM, Croyle MH, Gosserand RH. (2005). The dimensionality and antecedents of emotional labor strategies. *Journal of Vocational Behavior*, 66, 339–357.
- Dierdorff EC, Rubin RS. (2007). Carelessness and discriminability in work role requirement judgments: Influences of role ambiguity and cognitive complexity. *PERSONNEL PSYCHOLOGY*, 60, 597–625.
- Duval TS, Silvia PJ. (2002). Self-awareness, probability of improvement, and the self-serving bias. *Journal of Personality and Social Psychology*, 82, 49–61.
- Ellingson JE, Sackett PR, Connelly BS. (2007). Personality assessment across selection and development contexts: Insights into response distortion. *Journal of Applied Psychology*, 92, 386–395.
- Gael S. (1988). Interviews, questionnaires, and checklists. In Gale S (Ed.), *The job analysis handbook for business, industry and government* (Vol. 1, pp. 391–414). New York: Wiley.
- Goldberg LR. (1993). The structure of phenotypic personality traits. *American Psychologist*, 48, 26–34.

- Goldberg LR. (1999). A broad-bandwidth, public domain, personality inventory measuring the lower-level facets of several five-factor models. In Mervielde I, Deary I, De Fruyt F, Ostendorf F (Eds.), *Personality psychology in Europe* (Vol. 7, pp. 7–28). Tilburg, The Netherlands: Tilburg University Press.
- Goldberg LR, Johnson JA, Eber HW, Hogan R, Ashton MC, Cloninger CR, et al. (2006). The International Personality Item Pool and the future of public-domain personality measures. *Journal of Research in Personality*, 40, 84–96.
- Hakel MD, Stalder BK, Van De Voort DM. (1988). Obtaining and maintaining acceptance of job analysis. In Gale S (Ed.), *The job analysis handbook for business, industry and government* (Vol. 1, pp. 329–338). New York: Wiley.
- Hogan J, Rybicki S. (1998). *Performance improvement characteristics job analysis manual*. Tulsa, OK: Hogan Assessment Systems.
- Hollenbeck JH, Smither JW. (1998). A letter from the editor and associate editor of Personnel Psychology. *The Industrial-Organizational Psychologist*, 36(1). Available online at: <http://www.siop.org/TIP/backissues/TIPJuly98/hollenbeck.aspx>
- Jenkins M, Griffith R. (2004). Using personality constructs to predict performance: Narrow or broad bandwidth. *Journal of Business and Psychology*, 19, 255–269.
- Landy FJ. (1986). Stamp collecting versus science: Validation as hypothesis testing. *American Psychologist*, 41, 1183–1192.
- Latham GP. (2007). A speculative perspective on the transfer of behavior science findings to the workplace: “The times they are a-changin’.” *Academy of Management Journal*, 50, 1027–1032.
- Leary MR, Kowalski RM. (1990). Impression management: A literature review and two-component model. *Psychological Bulletin*, 107, 34–47.
- Levy R, Hancock GR. (2007). A framework of statistical tests for comparing mean and covariance structure models. *Multivariate Behavioral Research*, 42, 33–66.
- Lim BC, Ployhart RE. (2006). Assessing the convergent and discriminant validity of Goldberg’s international personality item pool: A multitrait-multimethod examination. *Organizational Research Methods*, 9, 29–54.
- Marks G, Miller N. (1987). Ten years of research on the false-consensus effect: An empirical and theoretical review. *Psychological Bulletin*, 102, 72–90.
- McCormick EJ, Jeanneret PR. (1988). Position analyses questionnaire (PAQ). In Gale S (Ed.), *The job analysis handbook for business, industry and government* (Vol. 2, pp. 825–842). New York: Wiley.
- McGraw KO, Wong SP. (1992). A common language effect size statistic. *Psychological Bulletin*, 111, 361–365.
- McHenry J. (2007, April). *We are the very model*. Presidential address delivered at the 22nd Annual Conference of the Society for Industrial and Organizational Psychology, New York.
- Meyer KD, Foster JL. (2007, April). Exploring the utility of three approaches to validating a job analysis tool. Paper presented in M. Anderson (Chair), *Worker-oriented job analysis tools: Development and validation*. Symposium conducted at the 22nd Annual Conference of the Society for Industrial and Organizational Psychology, New York, New York.
- Meyer KD, Foster JL, Anderson MG. (2006, April). *Assessing the predictive validity of the performance improvement characteristics job analysis tool*. Paper presented at the 21st Annual Conference of the Society of Industrial and Organizational Psychology, Dallas, TX.
- Mihura JL, Meyer GJ, Bel-Bahar T, Gunderson J. (2003). Correspondence among observer ratings of Rorschach, Big Five Model, and DSM-IV personality disorder constructs. *Journal of Personality Assessment*, 81, 20–39.

- Morgeson FP, Campion MA. (1997). Social and cognitive sources of potential inaccuracy in job analysis. *Journal of Applied Psychology*, 82, 627–655.
- Morgeson FP, Campion MA. (2000). Accuracy in job analysis: Toward an inference-based model. *Journal of Organizational Behavior*, 21, 819–827.
- Morgeson FP, Campion MA, Dipboye RL, Hollenbeck JR, Murphy K, Schmitt N. (2007). Are we getting fooled again? Coming to terms with limitations in the use of personality tests in personnel selection. *PERSONNEL PSYCHOLOGY*, 60, 1029–1049.
- Morgeson FP, Delaney-Klinger KA, Mayfield MS, Ferrara P, Campion MA. (2004). Self-presentation processes in job analysis: A field experiment investigating inflation in abilities, tasks, and competencies. *Journal of Applied Psychology*, 89, 674–686.
- Morris JA, Feldman DC. (1996). The dimensions, antecedents, and consequences of emotional labor. *Academy of Management Review*, 21, 986–1010.
- Motowidlo SJ, Hooper AC, Jackson HL. (2006). Implicit policies about relations between personality traits and behavioral effectiveness in situational judgment items. *Journal of Applied Psychology*, 91, 749–761.
- Muchinsky PM. (2004). When the psychometrics of test development meets organizational realities: A conceptual framework for organizational change, examples, and recommendations. *PERSONNEL PSYCHOLOGY*, 57, 175–209.
- Mullen B, Atkins JL, Champion DS, Edwards C, Handy D, Story JE, et al. (1985). The false consensus effect: A meta-analysis of 115 hypothesis tests. *Journal of Experimental Social Psychology*, 21, 262–283.
- Myors B, Lievens F, Schollaert E, Van Hoye G, Cronshaw SF, Mladinic A, et al. (2008). International perspectives on the legal environment for selection. *Industrial and Organizational Psychology: Perspectives on Science and Practice*, 1, 206–246.
- Neter J, Wasserman W, Whitmore GA. (1988). *Applied statistics* (3rd ed.). Newton, MA: Allyn and Bacon.
- Oliver J, Bakker AB, Demerouti E, de Jong RD. (2005). Projection of own on others' job characteristics: Evidence for the false consensus effect in job characteristics information. *International Journal of Selection and Assessment*, 13, 63–74.
- Ones DS, Dilchert S, Viswesvaran C, Judge TA. (2007). In support of personality assessment in organizational settings. *PERSONNEL PSYCHOLOGY*, 60, 995–1027.
- Ones DS, Viswesvaran C. (2003). Job-specific applicant pools and national norms for personality scales: Implications for range-restriction corrections in validation research. *Journal of Applied Psychology*, 88, 570–577.
- Ployhart RE, Weekley JA, Baughman K. (2006). The structure and function of human capital emergence: A multilevel examination of the attraction-selection-attrition model. *Academy of Management Journal*, 49, 661–677.
- Pulakos ED. (1984). A comparison of rater training programs: Error training and accuracy training. *Journal of Applied Psychology*, 69, 581–588.
- Raymark PH, Schmit MJ, Guion RM. (1997). Identifying potentially useful personality constructs for employee selection. *PERSONNEL PSYCHOLOGY*, 50, 723–736.
- Robbins JM, Krueger JI. (2005). Social projection to ingroups and outgroups: A review and meta-analysis. *Personality and Social Psychology Review*, 9, 32–47.
- Rynes SL, Colbert AE, Brown KG. (2002). HR professionals' beliefs about effective human resource practices: Correspondence between research and practice. *Human Resource Management*, 41, 149–174.
- Rynes SL, Giluk TL, Brown KG. (2007). The very separate worlds of academic and practitioner periodicals in human resource management: Implications for evidence-based management. *Academy of Management Journal*, 50, 987–1008.
- Sackett PR, Laczko RM. (2001). Job and work analysis. In Borman WC, Ilgen DR, Klimoski RJ (Eds.), *Handbook of psychology: Industrial and organizational psychology* (Vol. 12, pp. 21–37). New York: Wiley.

- Sanchez JI, Levine EL. (2000a). Accuracy or consequential validity: Which is the better standard for job analysis data? *Journal of Organizational Behavior*, *21*, 809–818.
- Sanchez JI, Levine EL. (2000b). The analysis of work in the 20th and 21st centuries. In Anderson N, Ones DS, Sinangil HK, Viswesvaran C (Eds.), *Handbook of industrial, work and organizational psychology* (Vol. 1, pp. 71–89). Thousand Oaks, CA: Sage.
- Scandura TA, Williams EA. (2000). Research methodology in management: Current practices, trends, and implications for future research. *Academy of Management Journal*, *43*, 1248–1264.
- Schaubroeck J, Ganster DC, Jones JR. (1998). Organization and occupation influences in the attraction-selection-attrition process. *Journal of Applied Psychology*, *83*, 869–891.
- Schneider B. (1987). The people make the place. *PERSONNEL PSYCHOLOGY*, *40*, 437–453.
- Shippmann JS, Ash RA, Battista M, Carr L, Eyde LD, Hesketh B, et al. (2000). The practice of competency modeling. *PERSONNEL PSYCHOLOGY*, *53*, 703–740.
- Shrout PE, Fleiss JL. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin*, *86*, 420–428.
- Sümer HC, Sümer N, Demirutku K, Çifci OS. (2001). Using a personality-oriented job analysis to identify attributes to be assessed in officer selection. *Military Psychology*, *13*, 129–146.
- Tedeschi JT. (Ed.). (1981). *Impression management theory and social psychological research*. New York: Academic Press.
- Tett RP, Christiansen ND. (2007). Personality tests at the crossroads: A response to Morgeson, Campion, Dipboye, Hollenbeck, Murphy, and Schmitt (2007). *PERSONNEL PSYCHOLOGY*, *60*, 967–993.
- Touzé PA, Steiner DD. (2002). L'adaptation Française du P.P.R., un outil identifiant les facteurs de personnalité susceptibles de prédire la performance au travail. *Orientation Scolaire et Professionnelle*, *31*, 443–466.
- Urban MS, Witt LA. (1990). Self-serving bias in group member attributions of success and failure. *Journal of Social Psychology*, *130*, 417–418.
- Voskuil OF, van Sliedregt T. (2002). Determinants of interrater reliability of job analysis: A meta-analysis. *European Journal of Psychological Assessment*, *18*, 52–62.
- Wilk SL, Desmarais LB, Sackett PR. (1995). Gravitation to jobs commensurate with ability: Longitudinal and cross-sectional tests. *Journal of Applied Psychology*, *80*, 79–85.
- Woehr DJ, Huffcutt AI. (1994). Rater training for performance appraisal: A quantitative review. *Journal of Occupational and Organizational Psychology*, *67*, 189–205.