

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/10702144>

A Field Study of Frame-of-Reference Effects on Personality Test Validity

Article in *Journal of Applied Psychology* · July 2003

DOI: 10.1037/0021-9010.88.3.545 · Source: PubMed

CITATIONS

77

READS

437

4 authors, including:



[Talya N. Bauer](#)

Portland State University

112 PUBLICATIONS 4,244 CITATIONS

SEE PROFILE



[Leslie B Hammer](#)

Oregon Health and Science University; Port...

91 PUBLICATIONS 2,674 CITATIONS

SEE PROFILE

A Field Study of Frame-of-Reference Effects on Personality Test Validity

John M. Hunthausen, Donald M. Truxillo, Talya N. Bauer, and Leslie B. Hammer
Portland State University

As part of a test validation study at a major U.S.-based airline, the authors tested the effects of providing an “at work” frame-of-reference on the validity of the NEO Five-Factor Inventory among a sample of customer service supervisors ($N = 206$). Frame-of-reference moderated the validity of the Extraversion and Openness to Experience subscales after controlling for cognitive ability. In addition, the frame-of-reference personality test showed incremental validity over cognitive ability ($\Delta R^2 = .16$), but the standard personality test did not ($\Delta R^2 = .05$). The authors’ discussion focuses on implications for personality theory and research and on implications for increasing the validity of personality tests in organizational settings.

Interest in personality as a predictor of job performance has grown in the past decade. Measures of normal personality (e.g., Costa & McCrae, 1992) have been shown to predict a wide range of performance criteria (e.g., Barrick & Mount, 1991, 1993; Hurtz & Donovan, 2000). In accordance, recent research has explored a number of factors that may affect the validity of personality tests, such as social desirability (e.g., Ones, Viswesvaran, & Reiss, 1996) and the frame-of-reference (FOR; e.g., at school, at work, in general) given to test takers when completing a personality questionnaire (e.g., Schmit, Ryan, Stierwalt, & Powell, 1995). For example, Schmit et al. (1995) found that providing college students with an “at school” FOR increased the validity of a conscientiousness measure in predicting grade point average (GPA).

Although research has explored the effects of FOR on the factor structure of personnel selection tests (e.g., Smith, Hanges, & Dickson, 2001), the effects of FOR on personality test validity have received little scrutiny, and the Schmit et al. (1995) study remains the primary published research on this issue. However, Schmit et al. only explored the effects of FOR on the validity of a Conscientiousness measure. Although Schmit et al. called for field research on this issue, the effects of FOR on validity remain unexplored in a field setting with actual performance ratings as the criterion.

The present research addresses these gaps in the literature in three specific ways. First, as part of a concurrent validation study at a major airline, we explored the effects of providing an at-work FOR on the validity of a personality measure (NEO Five-Factor

Inventory [NEO-FFI]; Costa & McCrae, 1992) in an organizational setting. Second, we extended the work of the Schmit et al. (1995) study by using a measure of actual work performance as the criterion. Third, we extended past research by exploring the effects of FOR in predicting work performance after controlling for cognitive ability, a variable shown to be predictive of job performance across settings (Hunter & Hunter, 1984).

FOR Effects and the Validity of Personality Measures

The development of the five-factor (“Big Five”) model (FFM) of normal adult personality (Digman, 1990; Goldberg, 1993; John, 1990), consisting of Agreeableness, Conscientiousness, Extraversion, Neuroticism, and Openness to Experience, has led to the application of Big Five measures to the field of personnel selection. Barrick and Mount’s (1991) meta-analytic study of the predictive validity of the Big Five found that Conscientiousness predicted performance across the five job categories represented. In addition, Openness to Experience and Extraversion were predictors of training performance, whereas Extraversion was a valid predictor of managerial performance. Further meta-analytic studies have confirmed the usefulness of the FFM for predicting performance for various jobs and criteria (e.g., Hurtz & Donovan, 2000; Salgado, 1997; Tett, Jackson, & Rothstein, 1991).

Some moderators of the relationship between the FFM and job performance have been explored. For example, Barrick and Mount (1993) provided evidence that the degree of autonomy that individuals perceive in a particular job (a situational variable) moderates the relationship between Conscientiousness and job performance. Relevant to the present study, Schmit et al. (1995) explored FOR as a moderator of personality test validity in a lab setting.

It could be argued from two standpoints that providing a FOR could actually lead to lower criterion-related validities. First, the powerful situation approach (see Gatewood & Feild, 2001) argues that individual differences such as personality will have relatively little effect on behavior in powerful situations, whereas behavior will be guided primarily by the context, and that providing a FOR for a strong situation such as work could lead to reduced variance and lower validity. Second, providing a FOR could lead to test transparency and greater faking, increasing systematic measurement error and again reducing validity.

John M. Hunthausen, Donald M. Truxillo, and Leslie B. Hammer, Department of Psychology, Portland State University; Talya N. Bauer, School of Business Administration, Portland State University.

This research was completed as part of John M. Hunthausen’s doctoral dissertation, which was chaired by Donald M. Truxillo, at the Department of Psychology at Portland State University. Talya N. Bauer and Leslie B. Hammer served as committee members. An earlier version of this article was presented at the Academy of Management Conference, Denver, Colorado, August 2002.

Correspondence concerning this article should be addressed to John M. Hunthausen, who is now at P.O. Box 6890–SEAHE, Alaska Airlines, Seattle, Washington 98168-0900. E-mail: john.hunthausen@alaskaair.com

In contrast, theory and some empirical evidence suggest that higher predictive validities are obtained if personality inventories elicit how individuals behave, feel, and think specifically at work (e.g., Hogan, 1991; Schmit et al., 1995). For example, Wright and Mischel (1987) suggested that stable patterns of behavior are contingent on situational conditions (i.e., a Person \times Situation interaction), which they labeled *conditional dispositions*. That is, a person's underlying personality system leads not only to average levels of behavior across situations but also to variable patterns of behavior across situations, a Person \times Situation interaction ("if . . . then" situation-behavior relationships; e.g., Mendoza-Denton, Ayduk, Mischel, Shoda, & Testa, 2001; Mischel & Shoda, 1995). Similarly, McCrae and Costa (1996) suggested that to make inferences about underlying traits, personality questions should tap manifestations of behavioral tendencies that are influenced by external factors. Thus, behavioral tendencies may be situation-specific, even though the underlying trait they reflect is generally constant across time and situations.

In the context of the FFM, Schmit et al. (1995) compared the validities of an altered FOR personality test and a standard personality test by using a student sample and GPA as the criteria. Schmit et al. manipulated the FOR by modifying the items to reflect the school context. Specifically, Schmit et al. appended a reference to school to the beginning or end of each personality test item. As predicted, the criterion-related validity of the Conscientiousness dimension of the test with the "at school" FOR was greater in magnitude than that of the unaltered personality test. Although specifying a FOR could lead to reductions in variance and hence to reduced validity, particularly if the context provided is a powerful situation (see Gatewood & Feild, 2001), the study of Schmit et al. found the opposite to be the case.

The present study significantly extends the Schmit et al. (1995) study in three ways. First, whereas Schmit et al. used a college student sample, we used a concurrent validation design among customer service employees to examine FOR effects on personality-predictor validity. Second, we used actual job performance as the criterion for examining personality test validity (i.e., rather than GPA). Third, we explored FOR effects on all FFM dimensions (i.e., not just Conscientiousness), an important area for research in understanding the relationship between the FFM and job performance (Gatewood & Feild, 2001). Job analytic data for the customer service manager job that we chose for this study supported the investigation of the relationship between each of the FFM dimensions and job performance. The major responsibilities of this job include handling customer complaints and requests, ensuring that baggage gets loaded on the correct planes, ensuring on-time departure of planes, and supervising customer service agents and baggage handlers. We believed that Extraversion would be relevant to the supervisory aspects of the job (cf. Barrick & Mount, 1991), Agreeableness to the customer service dimensions, Openness to Experience to the ability to handle complaints creatively, and Neuroticism would be negatively related to work performance. We believed that Conscientiousness would be important for most aspects of job performance (cf. Barrick & Mount, 1991).

In summary, we hypothesized that there would be a stronger relationship between each of the FFM variables and job performance when using an at-work FOR than when using a standard FOR. In addition, because cognitive ability is also an antecedent of

performance and recent research has controlled for both cognitive ability and personality when exploring the incremental validity of predictors (e.g., Cortina, Goldstein, Payne, Davison, & Gilliland, 2000; Mount, Witt, & Barrick, 2000), we tested the effects of FOR on the relationship between the FFM and job performance after controlling for cognitive ability.

Method

Participants

Participants were 214 entry-level customer service managers employed with a major U.S. airline at airports within the United States. Most participants (86%) had worked for the airline for over 5 years and had held their current position for at least 1 year (87%). The sample was 54% male and 46% female, 55% White, 23% Hispanic, 12% Black, 4% Asian-Pacific Islander, 2% Native American, and 4% defined themselves under the category of "other." Individuals for whom job performance data were unavailable ($n = 2$) and whose test scores were more than 3 *SDs* beyond the mean ($n = 6$) were excluded from the study, resulting in a final sample of 206. Approximately half of the employees were randomly assigned to each of the two FOR conditions, the standard personality test ($n = 102$), and the FOR personality test ($n = 104$).

Because we were concerned about the motivation of this incumbent sample versus a sample of actual applicants, we assessed test-taking motivation by using seven items adapted from Arvey, Strickland, Drauden, and Martin's Test Attitude Survey (1990; $\alpha = .94$) prior to participants taking the two tests. The mean for the sample was 4.37 (on a scale of 1–5; $SD = .67$), suggesting reasonably high test-taking motivation. However, the test-taking motivation means for the two groups differed slightly (FOR condition: $M = 4.26$, standard condition: $M = 4.49$), $t(203) = 2.49$, $p < .05$. Therefore test-taking motivation was used as a control variable in all analyses.

Measures

Personality. We measured personality by using the NEO-FFI (Costa & McCrae, 1992). The NEO-FFI is based on the FFM of personality (Digman, 1990) and includes scales of Neuroticism, Extraversion, Openness to Experience, Agreeableness, and Conscientiousness. Each of the five factors consists of 12 items that combine to yield a total of 60 items. Responses to each item are recorded on a 5-point scale ranging from 1 (*strongly disagree*) to 5 (*strongly agree*). Alpha internal consistency reliabilities for the five factors commonly range from .68 to .86 (Costa & McCrae, 1992).

Cognitive ability. The general cognitive ability measure used in the present study was the Wonderlic Personnel Test (WPT; Wonderlic Personnel Test, 1992). The WPT is a 12-min timed test consisting of 50 multiple-choice and fill-in-the-blank questions. Studies have shown the WPT to be reliable (Dodrill, 1983; McKelvie, 1989). The overall general population mean for the WPT is 21.06 ($SD = 7.12$; Wonderlic, 2000). In the present study the mean for the WPT was 23.63 ($SD = 5.87$).

Job performance. Supervisory ratings of overall job performance were measured using a three-item measure from Motowidlo and Van Scotter (1994). The single stem for the three items read "Please rate this supervisor's overall job performance." The first item ranged from 5 (high; *exceeds standards for job performance*) to 3 (moderate; *meets standards for job performance*) to 1 (low; *does not meet standards for job performance*). The second item ranged from 5 (high; *performs at a high level compared with others in the same job*) to 3 (moderate; *performs at an average level compared with others in the same job*) to 1 (low; *performs at a low level compared with others in the same job*). The third item ranged from 5 (high; *contributes more to the station [i.e., airport] effectiveness than most*), to 3 (moderate; *makes an average contribution to station effectiveness*) to 1 (low; *contributes less to station effectiveness than most*). The three items

had similar means (3.38–3.52) and standard deviations (.72–.75), and we created the overall performance score from the mean of these three variables. The alpha reliability across the three items of this scale was .88 ($M = 3.51, SD = 0.71$).

FOR Manipulation

The two conditions (standard FOR and at-work FOR) were defined by the instructions participants were given for completing the NEO-FFI. Specifically, rather than appending “at work” to each item as was done by Schmit et al. (1995), the FOR form of the NEO-FFI was established by adding the reference to work before and during the administration. (This was a simplified procedure suggested by Schmit et al.) Participants were instructed orally and via an instruction sheet to think about how they behave at work when responding to each statement. In addition, a reminder (“Remember, think about how you are AT WORK in general when responding to these questions”) was attached to the top of the pages of the NEO-FFI test booklet. As a manipulation check, an item was added at the end of the test, tapping whether they considered how they are at work when responding to the items. All participants in the FOR condition indicated they thought about how they are at work when responding to the items.

Procedure

Testing sessions (roughly 45 min) were conducted at 16 airports, with multiple testing sessions in some locations. Half of the testing sessions offered the FOR personality inventory administration, and the other half of the testing sessions offered the standard personality inventory administration. In addition, testing sessions were counterbalanced such that for half of the participants the NEO was administered first, and for half the participants the WPT was administered first. Sessions were conducted on company time. Participation was voluntary.

At the beginning of each testing session, participants were given a brief overview of the study as described in a cover letter and informed consent form. Participants were then administered the pretest test-taking motivation measure. Next they were administered the NEO-FFI (FOR or standard) and the WPT. Within 2 days after the testing session, participants’ supervisors were provided with the performance evaluation forms and cover letters describing the study, noting that ratings would be used for research purposes only.

Results

Means, standard deviations, and intercorrelations of study variables are given in Table 1. The correlations indicate that cognitive ability ($r = .24$) and conscientiousness ($r = .21$) were correlated with job performance. None of the other personality variables were correlated with job performance. Tables 2 and 3 show means, standard deviations, and intercorrelations for study variables for the standard and FOR conditions, respectively. Conscientiousness, Extraversion, and Openness to Experience were significantly correlated with job performance in the FOR condition but not in the standard condition. Although cognitive ability was correlated with job performance, $r = .17, p < .05$, one-tailed, in the FOR condition versus $r = .31, p < .01$, two-tailed, in the standard condition, this was not significantly different.

We first explored whether using an at-work FOR would increase the validity of the personality measures by comparing the correlation between the personality variables and job performance in each condition (see Schmit et al., 1995). There was a significant correlation ($p < .01$) between job performance and Conscientiousness ($r = .31$), Extraversion ($r = .26$), and Openness to Experience

Table 1
Means, Standard Deviations, and Intercorrelations of Primary Study Variables

Variable	M	SD	1	2	3	4	5	6	7	8	9	10	11
1. Gender	0.47	0.50	—										
2. Job tenure	2.89	1.24	-.11	—									
3. Cognitive ability	23.63	5.87	-.13	.07	—								
4. Test-taking motivation	4.37	0.67	-.04	-.12	-.01	(.94)							
5. FOR condition	0.50	0.50	-.02	.01	-.04	-.17*	—						
6. Conscientiousness	37.88	5.55	.12	-.19**	-.07	.25**	-.05	—					
7. Extraversion	33.87	5.53	.04	-.24**	-.11	.22**	.11	.39*	—				
8. Openness to Experience	27.98	5.08	.09	-.15*	.09	.11	-.10	.22**	.28**	—			
9. Neuroticism	13.61	6.07	.07	.04	-.10	-.12	.00	-.31**	-.40**	-.30**	—		
10. Agreeableness	34.04	5.34	.20*	-.22**	-.15*	.25**	-.01	.27**	.33**	.19**	-.42**	—	
11. Job performance	3.51	0.71	.11	.12	.24*	-.03	-.12	.21**	.06	.09	-.09	-.03	(.88)

Note. $N = 208$. FOR was coded as follows: 0 = standard administration, 1 = at-work FOR. Gender was coded as follows: 0 = male, 1 = female. Job tenure was coded as follows: 1 = less than 1 year, 2 = 1–3 years, 3 = 3–5 years, 4 = 5–10 years, 5 = greater than 10 years. FOR = Frame-of-reference.
* $p < .05$. ** $p < .01$.

Table 2
Means, Standard Deviations, and Intercorrelations of Primary Study Variables in the Standard NEO-FFI Administration

Variable	M	SD	1	2	3	4	5	6	7	8	9	10
1. Gender	0.48	0.50	—									
2. Job tenure	2.87	1.15	-.19	—								
3. Cognitive ability	23.87	5.92	-.06	.07	—							
4. Test-taking motivation	4.49	0.55	-.08	-.21*	.04	—						
5. Conscientiousness	38.14	5.57	.14	-.17	-.06	.27**	—					
6. Extraversion	33.23	5.78	.08	-.25*	-.15	.30**	.44**	—				
7. Openness to Experience	28.50	4.99	.07	-.17	.07	.23*	.20*	.34**	—			
8. Neuroticism	13.61	6.34	-.02	.05	-.03	-.25*	-.46**	-.49**	-.44**	—		
9. Agreeableness	34.08	5.54	.26**	-.24*	-.09	.26**	.26**	.36**	.28**	-.51**	—	
10. Job performance	3.59	0.70	.14	.18	.17	-.15	.10	-.10	-.12	-.06	-.03	—

Note. N = 102. Gender was coded as follows: 0 = male, 1 = female. Job tenure was coded as follows: 1 = less than 1 year, 2 = 1–3 years, 3 = 3–5 years, 4 = 5–10 years, 5 = greater than 10 years. NEO-FFI = NEO Five-Factor Inventory (Costa & McRae, 1992). *p < .05. **p < .01.

(r = .27) in the FOR condition but not in the standard condition (rs = .10, -.10, -.12, respectively, ns). Furthermore, we found differences between the validity coefficients for the two samples for Extraversion (z = -2.60, p < .01) and Openness to Experience (z = -2.81, p < .01), and a marginally significant difference for conscientiousness (z = -1.56, p = .06, one-tailed). The differences for Neuroticism and Agreeableness were not significant (zs = .42 and -.06, respectively, ns).

Next we tested the moderating effects of FOR on personality test validity by using a series of four regression equations created in a hierarchical fashion. With overall job performance as the dependent measure, we entered the control variables (test-taking motivation, job tenure, and gender) in Step 1, cognitive ability in Step 2, the centered personality variables and FOR condition in Step 3, and the Personality × FOR Interaction (product) terms for each of the FFM dimensions in Step 4. The results are shown in Table 4. There was a significant increase in R² in Step 4 with the addition of the interaction terms, ΔR² = .05, F(5, 184) = 2.60, p < .05. Examination of the betas indicated that the interaction terms for Extraversion (β = .14, p < .05) and Openness to Experience (β = .15, p < .05) were significant in Step 4. Specifically, these betas indicate a stronger relationship between each of these personality measures and job performance for those in the FOR condition. The Conscientiousness × FOR Interaction was not

significant in the final equation, although there was a main effect for Conscientiousness (β = .22, p < .01). In summary, these regression analyses provide support for FOR effects on the validity of Extraversion and Openness to Experience after controlling for cognitive ability.

Accordingly, we also compared the incremental validity of the personality measures over cognitive ability for each condition using hierarchical regression. That is, for those in each condition, we entered cognitive ability and test-taking motivation in Step 1, and the three personality dimensions of Conscientious, Extraversion, and Openness to Experience in Step 2, with overall job performance as the dependent variable. Although the addition of the personality variables did not show incremental validity over cognitive ability for the standard personality test, ΔR² = .05, F(3, 94) = 1.75, ns, they did for the FOR personality test, ΔR² = .16, F(3, 97) = 7.08, p < .01. Examination of the betas indicated that for the FOR condition, Conscientiousness (β = .26, p < .01), Extraversion (β = .17, p < .05), and Openness to Experience (β = .14, p < .10) were each positively related to overall job performance.

Discussion

The present study found that an at-work FOR moderated the validity of a personality test in a work context after controlling for

Table 3
Means, Standard Deviations, and Intercorrelations of Primary Study Variables in the At-Work FOR NEO-FFI Administration

Variable	M	SD	1	2	3	4	5	6	7	8	9	10
1. Gender	0.46	0.50	—									
2. Job tenure	2.90	1.33	-.05	—								
3. Cognitive ability	23.40	5.85	-.20*	.07	—							
4. Test-taking motivation	4.26	0.75	-.02	-.06	-.05	—						
5. Conscientiousness	37.63	5.55	.09	-.21*	-.08	.24*	—					
6. Extraversion	34.49	5.22	-.01	-.23*	-.06	.22*	.36**	—				
7. Openness to Experience	27.47	5.14	.11	-.14	.10	.00	.23*	.24*	—			
8. Neuroticism	13.60	5.82	.16	.02	-.18	-.02	-.16	-.30**	-.15	—		
9. Agreeableness	34.01	5.15	.14	-.21*	-.22*	.24*	.27**	.30**	.09	-.31**	—	
10. Job performance	3.42	0.71	.08	.07	.31**	.02	.31**	.26**	.27**	-.12	-.02	—

Note. N = 104. Gender was coded as follows: 0 = male, 1 = female. Job tenure was coded as follows: 1 = less than 1 year, 2 = 1–3 years, 3 = 3–5 years, 4 = 5–10 years, 5 = greater than 10 years. FOR = Frame-of-reference; NEO-FFI = NEO Five-Factor Inventory (Costa & McRae, 1992). *p < .05. **p < .01.

Table 4
Hierarchical Regression of Cognitive Ability, Personality Variables, and FOR on Job Performance

Step and variable	Job performance		
	R^2	ΔR^2	β
Step 1	.03*		
Test-taking motivation			-.01
Job tenure			.13*
Gender			.14*
Step 2	.10†	.07†	
Cognitive ability			.26†
Step 3	.18†	.08†	
Conscientiousness			.24†
Extraversion			.10
Openness to Experience			.01
Neuroticism			.03
Agreeableness			-.04
FOR condition			-.11
Step 4	.23†	.05*	
Conscientiousness \times FOR			.06
Extraversion \times FOR			.14*
Openness to Experience \times FOR			.15*
Neuroticism \times FOR			.04
Agreeableness \times FOR			-.04
<i>F</i> for overall equation		5.27**	

Note. $N = 206$. Betas are from the final equation. FOR was coded as follows: 0 = standard administration, 1 = at-work FOR. Gender was coded as follows: 0 = male, 1 = female. FOR = Frame-of-reference.

* $p < .05$, two-tailed. † $p < .01$, one-tailed.

cognitive ability. Specifically, we found that when providing an at-work FOR to participants, Extraversion and Openness to Experience scales of the NEO-FFI (Costa & McRae, 1992) showed increased concurrent validity in a sample of customer service managers for a major airline. In addition, the correlational analyses revealed a marginally significant difference for the conscientiousness scale. These findings are consistent with those of Schmit et al. (1995) who found that providing test takers with a specific FOR (i.e., an at-work context) for a personality test led to higher criterion validities. Such a FOR should direct applicants to focus on how they would behave in a work context (i.e., a conditional disposition; e.g., Wright & Mischel, 1987) when completing personality tests. The present study was the first field research to show that administering a personality selection test in a work-related context can lead to increased test validity.

These results suggest that the at-work FOR increased the predictive validity of measures of Extraversion and Openness to Experience. Extraversion should be related to job performance given the interpersonal nature of this customer service job. In accordance, willingness to interact with people may vary considerably across contexts such that people show more extraversion given different contexts (e.g., work vs. social gatherings). Thus, the FOR manipulation may have captured this difference, leading to increased validity. Similarly, the effects for Openness to Experience may have been found because being interested in learning new things at work may be particularly important for this job. Marginally significant FOR effects were found for Conscientiousness in the correlational analysis, but not in the moderated regression analysis. Although this may be because interaction effects can

be hard to detect with linear regression (Pedhazur, 1982), it is also possible that Conscientiousness is less susceptible to FOR effects.

We found no FOR effects for Agreeableness and Neuroticism. This seemed particularly surprising for Agreeableness in this customer service job. However, it may be that being overly agreeable (i.e., overly accommodating) at work may not be good in a supervisory job such as this one, leading to overaccommodation of customer demands. Alternatively, Agreeableness and Neuroticism may not be important to this type of work. Future research should explore this issue.

This research advances the literature in four ways. First, it extends the published research on personality test validity and FOR (e.g., Schmit et al., 1995) by demonstrating the effects of FOR in an actual field validation study. Moreover, unlike past studies that used college student samples and GPA as the criterion, the present study used actual employees and job performance as the criterion. Second, although past research on the FFM has focused primarily on Conscientiousness, the present study explored the effects of FOR on test validity for all of the FFM factors. Thus, the results of this study suggest that other personality dimensions may be valid when an at-work FOR is used. Third, the present study explored the effects of FOR on personality test validity after controlling for cognitive ability. Because cognitive ability is considered the other primary determinant of job performance in addition to personality (e.g., Murphy & Shiarella, 1997), these results give a more complete picture of the incremental validity of using a specific FOR with personality tests when other key predictors are controlled. Fourth, because providing a FOR did not lower validity, it challenges the powerful situation (decreased variance) and transparency (increased systematic error) arguments. Moreover, it supports Wright and Mischel's (1987) understanding of personality as stable patterns of behavior that are contingent on situational conditions and conditional dispositions.

Future Research

We see several avenues for future studies of FOR and test validity. First, research should explore how different work contexts affect the application of FOR to personality tests. That is, the type of job should impact whether the at-work FOR enhances personality test validity. For example, the results of the present study suggest that providing a FOR increased the validity of Extraversion, Openness to Experience, and, to a lesser extent, Conscientiousness. This makes sense for customer service managers employed at a large airline. Future research should explore FOR effects in other jobs. In addition, jobs may vary in terms of whether they provide individuals with a powerful social context (see Gatewood & Feild, 2001). Providing a FOR for jobs with very strong contexts may reduce variability and hence actually decrease validity. Although the results suggest that providing a FOR did not decrease validity by providing a powerful situation or by facilitating applicant faking, research across a range of different jobs is needed. Finally, future research should explore whether using even an irrelevant FOR (e.g., school, home, with friends) reduces error and increases validity.

Second, future research should extend the test of these issues to applicants. Applicant samples may display greater variance on each of the predictor variables. Although our pretest suggested that these employee participants were sufficiently motivated, providing

a FOR to highly motivated applicants could decrease validity by facilitating applicant faking. Difficulties in conducting this type of research among applicants include the legal issue of random assignment of actual applicants to FOR conditions and the restriction of range on criterion measures given the likelihood that all applicants would not be hired. The results of the present study are compelling because results were found despite the range restriction that is expected among a sample of actual employees. Moreover, in contrast, incumbent responses are not clouded by impression management effects or “faking good.” However, it may be difficult for applicants responding to a FOR for a job in which they have never worked. Therefore, future research using applicant samples should recognize that if work is given as a frame of reference, the work situation previously experienced by that individual may be qualitatively different from the job for which he or she is applying.

Third, future research should explore the effects of FOR on the differential relationships between predictors and specific job performance dimensions. For example, [Stewart \(1999\)](#) found that different subtraits of Conscientiousness were differentially related to performance criteria. Future studies of FOR effects on the relationship between personality traits–subtraits and specific job performance dimensions should lead to the development of a nomological network to increase understanding of the predictors of work performance and their outcomes.

Although some research has shown that the FOR of a personality test does not alter the underlying factor structure (e.g., [Smith, Hanges, & Dickson, 2001](#)), future research should verify this finding, if possible. This approach would provide evidence to determine whether the psychometric characteristics of each of the FFM dimensions were differentially altered by an at-work FOR. Unfortunately, the sample sizes of each condition in the current research were too small ($n = 102$, $n = 104$) to conduct a comparative factor analysis.

As noted, we controlled for test-taking motivation in the regressions because the mean for test-taking motivation was lower in the FOR condition. In addition, there were significant correlations between test-taking motivation and Conscientiousness, Extraversion, and Agreeableness. Perhaps an at-work FOR, at least for job incumbents, acts to suppress the need to “present one’s self favorably.” For example, people may feel they can describe themselves as being not so agreeable at work without suggesting they are disagreeable in general. Future research should explore whether FOR has this effect with job applicants.

Finally, the role of other individual differences should be explored. For example, [Schmit et al. \(1995\)](#) suggested that other moderators, such as self-monitoring, may help explain the effects of FOR (i.e., why applicants use different frames of reference). In a hiring situation, high self-monitors, who rely more on the immediate situation to guide their behaviors, may be more likely to refer to work-related experiences (i.e., an at-work FOR) than low self-monitors when responding to unaltered personality items. Therefore, providing an at-work FOR to test items may provide low self-monitors the context that allows them to more accurately indicate how they would behave at work. These issues should be explored in applicant settings where self-monitoring may be of greater importance. In addition, individual differences may be controlled for by conducting a within-subjects experimental design. That is, research should examine personality predictor valid-

ities for the same set of individuals responding with a FOR and with no FOR.

Implications for Organizations

The potential implications of this research for organizations are substantial. Such implications suggest that personality tests that use an at-work FOR show incremental validity over their standard counterparts. Such validity may be accomplished as easily as it was in the present study by simply altering test instructions. Using a more work-specific personality test appears to increase its validity and utility. Therefore, test developers, personnel psychologists, and human resource professionals should consider the FOR that applicants are using when responding to personality-related selection assessments. Specifically, validity and utility of personality tests for personnel selection appear to be enhanced when applicants are provided with a specific FOR to help them focus on work-related behaviors, such as when instructions and test items refer to a job-related FOR.

Organizational decision makers and researchers should explore the use of at-work FORs when evaluating personality test validity or should consider those tests that already use an at-work FOR. However, we caution that such decisions should be made on an individual basis, and specific personality dimensions and FORs should only be considered after adequate validation work.

References

- [Arvey, R. D., Strickland, W., Drauden, G., & Martin, C. \(1990\).](#) Motivational components of test taking. *Personnel Psychology, 43*, 696–716.
- [Barrick, M. R., & Mount, M. K. \(1991\).](#) The Big Five personality dimensions and job performance: A meta-analysis. *Personnel Psychology, 44*, 1–26.
- [Barrick, M. R., & Mount, M. K. \(1993\).](#) Autonomy as a moderator of the relationships between the Big Five personality dimensions and job performance. *Journal of Applied Psychology, 78*, 111–118.
- [Cortina, J. M., Goldstein, N. B., Payne, S. C., Davison, H. K., & Gilliland, S. W. \(2000\).](#) The incremental validity of interview scores over and above cognitive ability and conscientiousness scores. *Personnel Psychology, 53*, 325–351.
- [Costa, P. T., Jr., & McCrae, R. R. \(1992\).](#) *NEO PI-R: Professional manual. Revised NEO Personality Inventory NEO PR-R and NEO Five-Factor Inventory NEO-FFI*. Odessa, FL: Psychological Assessment Resources.
- [Digman, J. M. \(1990\).](#) Personality structure: Emergence of the five-factor model. *Annual Review of Psychology, 41*, 417–440.
- [Dodrill, C. B. \(1983\).](#) Long-term reliability of the Wonderlic Personnel Test. *Journal of Consulting and Clinical Psychology, 51*, 316–317.
- [Gatewood, R. D., & Feild, H. S. \(2001\).](#) *Human resource selection* (5th ed.). Orlando, FL: Dryden Press.
- [Goldberg, L. R. \(1993\).](#) The structure of phenotypic personality traits. *American Psychologist, 48*, 26–34.
- [Hogan, R. T. \(1991\).](#) Personality and personality measurement. In M. D. Dunnette & L. M. Hough (Eds.), *Handbook of industrial and organizational psychology* (Vol. 2, pp. 873–919). Palo Alto, CA: Consulting Psychologists Press.
- [Hunter, J., & Hunter, R. \(1984\).](#) Validity and utility of alternate measures of job performance. *Psychological Bulletin, 96*, 72–98.
- [Hurtz, G. M., & Donovan, J. J. \(2000\).](#) Personality and job performance: The Big Five revisited. *Journal of Applied Psychology, 85*, 869–879.
- [John, O. P. \(1990\).](#) The “Big Five” factor taxonomy: Dimensions of personality in the natural language and in questionnaires. In L. A. Pervin

- (Ed.), *Handbook of personality: Theory and research* (pp. 66–100). New York: Guilford Press.
- McCrae, R. R., & Costa, P. T., Jr. (1996). Toward a new generation of personality theories: Theoretical contexts for the five-factor model. In J. S. Wiggins (Ed.), *The five-factor model of personality: Theoretical perspectives* (pp. 51–87). New York: Guilford Press.
- McKelvie, S. J. (1989). The Wonderlic Personnel Test: Reliability and validity in an academic setting. *Psychological Reports*, *65*, 161–162.
- Mendoza-Denton, R., Ayduk, O., Mischel, W., Shoda, Y., & Testa, A. (2001). Person \times Situation interactionism in self-encoding (I am . . . when . . .): Implications for affect regulation and social information processing. *Journal of Personality and Social Psychology*, *80*, 533–544.
- Mischel, W., & Shoda, Y. (1995). A cognitive-affective system theory of personality: Reconceptualizing situations, dispositions, dynamics, and invariance in personality structure. *Psychological Review*, *102*, 246–268.
- Motowidlo, W. C., & Van Scotter, J. R. (1994). Evidence that task performance should be distinguished from contextual performance. *Journal of Applied Psychology*, *79*, 475–480.
- Mount, M. K., Witt, L. A., & Barrick, M. R. (2000). Incremental validity of empirically keyed biodata scales over GMA and Five Factor Personality constructs. *Personnel Psychology*, *53*, 299–323.
- Murphy, K. R., & Shirella, A. H. (1997). Implications of the multidimensional nature of job performance for the validity of selection tests: Multivariate frameworks for studying test validity. *Personnel Psychology*, *50*, 823–854.
- Ones, D. S., Viswesvaran, C., & Reiss, A. D. (1996). Role of social desirability in personality testing for personnel selection: The red herring. *Journal of Applied Psychology*, *81*, 660–679.
- Pedhazur, E. J. (1982). *Multiple regression in behavioral research* (2nd ed.). New York: Harcourt Brace Jovanovich.
- Salgado, J. F. (1997). The five factor model of personality and job performance in the European community. *Journal of Applied Psychology*, *82*, 30–43.
- Schmit, M. J., Ryan, A. M., Stierwalt, S. L., & Powell, A. B. (1995). Frame-of-reference effects on personality scale scores and criterion-related validity. *Journal of Applied Psychology*, *80*, 607–620.
- Smith, D. B., Hanges, P. J., & Dickson, M. W. (2001). Personnel selection and the five-factor model: Reexamining the effects of applicants' frame of reference. *Journal of Applied Psychology*, *86*, 304–315.
- Stewart, G. L. (1999). Trait bandwidth and stages of job performance: Assessing differential effects for conscientiousness and its subtraits. *Journal of Applied Psychology*, *84*, 959–968.
- Tett, R. P., Jackson, D. N., & Rothstein, M. (1991). Personality measures as predictors of job performance: A meta-analytic review. *Personnel Psychology*, *44*, 703–742.
- Wonderlic. (2000). *Wonderlic Personnel Test and scholastic level exam user's manual*. Libertyville, IL: Author.
- Wonderlic Personnel Test. (1992). *Wonderlic Personnel Test user's manual*. Libertyville, IL: Author.
- Wright, J. C., & Mischel, W. (1987). A conditional approach to dispositional constructs: The local predictability of social behavior. *Journal of Personality and Social Psychology*, *53*, 1159–1177.

Received April 8, 2002

Revision received July 23, 2002

Accepted August 21, 2002 ■

Low Publication Prices for APA Members and Affiliates

Keeping you up-to-date. All APA Fellows, Members, Associates, and Student Affiliates receive—as part of their annual dues—subscriptions to the *American Psychologist* and *APA Monitor*. High School Teacher and International Affiliates receive subscriptions to the *APA Monitor*, and they may subscribe to the *American Psychologist* at a significantly reduced rate. In addition, all Members and Student Affiliates are eligible for savings of up to 60% (plus a journal credit) on all other APA journals, as well as significant discounts on subscriptions from cooperating societies and publishers (e.g., the American Association for Counseling and Development, Academic Press, and Human Sciences Press).

Essential resources. APA members and affiliates receive special rates for purchases of APA books, including the *Publication Manual of the American Psychological Association*, and on dozens of new topical books each year.

Other benefits of membership. Membership in APA also provides eligibility for competitive insurance plans, continuing education programs, reduced APA convention fees, and specialty divisions.

More information. Write to American Psychological Association, Membership Services, 750 First Street, NE, Washington, DC 20002-4242.