

## RETESTING PERSONALITY IN EMPLOYEE SELECTION: IMPLICATIONS OF THE CONTEXT, SAMPLE, AND SETTING<sup>1</sup>

COURTNEY L. HOLLADAY

*MD Anderson Cancer Center  
University of Texas*

EMILY DAVID

*Zayed University*

STEFANIE K. JOHNSON

*University of Colorado Denver*

*Summary.*—The present study sought to assess when and how actual job applicants change their responses when filling out an unproctored personality selection assessment for a second time. It was predicted feedback would be a key contextual motivator associated with how much applicants change their answers during the second administration. Mediation results showed that individuals receiving feedback that showed a low score on the personality assessment was the reason they did not get the job were more likely to employ faking response strategies in the second testing session, predicting the highest change in scores between the first and second testing sessions. Individuals receiving no feedback and those not experimentally motivated to fake (i.e., a comparison group of students) showed less change in responses across administrations.

Research has generally concluded that personality remains relatively stable over time (McCrae & Costa, 1994; Roberts & DelVecchio, 2000; Hampson & Goldberg, 2006; Roberts, Walton, & Viechtbauer, 2006) and that much of the observed variation in personality is due to transient error rather than true change (Chmielewski & Watson, 2009). Indeed, the notion that traits are relatively fixed by adulthood is an important assumption underlying the use of personality in personnel selection (Barrick & Mount, 1991; Barrick, Mount, & Judge, 2001; Ones, Dilchert, Viswesvaran, & Judge, 2007). Despite the ubiquity of personality testing, its use in employee selection is not without opponents (Guion & Gottier, 1965; Morgeson, Campion, Dipboye, Hollenbeck, Murphy, & Schmitt, 2007). One of the key criticisms is that motivated applicants can (and frequently do) fake their responses on personality tests (Rosse, Stecher, Miller, & Levin, 1998; Birkeland, Manson, Kisamore, Brannick, & Smith, 2006; Morgeson, *et al.*, 2007).

Potentially exacerbating this problem, many organizations are moving toward unproctored, online selection testing in cases where test centers may be cost-prohibitive or when applicant tracking systems do not easily integrate with the test center (Guo & Drasgow, 2010). Research-

---

<sup>1</sup>Address correspondence to Courtney Holladay, 7007 Bertner Ave., Houston, TX 77030 or e-mail (CLHolladay@mdanderson.org).

ers have debated whether or not unproctored settings increase the occurrence of faking or cheating (Chapman & Webster, 2003; Young, 2003; Bartram & Brown, 2004; Tippins, Beaty, Drasgow, Gibon, Pearlman, Segall, *et al.*, 2006; Ellingson, Sackett, & Connelly, 2007; Hogan, Barrett, & Hogan, 2007; Nye, Do, Drasgow, & Fine, 2008; Landers, Sackett, & Tuzinski, 2011). Some studies have found little to no difference in scores between the proctored and unproctored testing conditions (Nye, *et al.*, 2008; Templer & Lange, 2008; Arthur, Glaze, Villado, & Taylor, 2009). For example, Nye, *et al.* (2008) found that the factor structures of tests between proctored and unproctored conditions did not differ, and concluded there was no evidence for cheating in the unproctored setting. However, the authors used an Attention-to-Detail Test rather than a personality test. As this examination was more cognitive in nature and was strictly timed, it was likely that participants had little time to alter their answers. In a personality inventory, participants are likely to have at least a vague idea of how the ideal employee would answer (Schmit & Ryan, 1993). Indeed, some researchers have found modest differences between personality scores administered in proctored and unproctored testing conditions ( $d$  ranging from  $-.13$  to  $.14$ ; Bartram & Brown, 2004;  $d$  ranging from  $-.18$  to  $.35$ ; Brown, Bartram, Holtzhausen, Carstairs, & Mylonis, 2005).

The most widely accepted design to address this research question is to retest individuals using a repeated measures design within a selection context. Such studies not only help researchers to understand the effect that changes in applicant scores can have on the selection process, but also provide practitioners with insight into the prevalence of faking responses in an unproctored setting. The present study used this methodology to examine responses on a personality inventory across two testing occasions to examine the extent to which scores changed in response to feedback. Specifically, this study aimed to detect changes in personality scores in an unproctored, high-stakes selection setting that occurred as a result of feedback. In addition, the motivation to fake was tested as an underlying mechanism explaining the relationship between negative feedback and score change.

### *Faking in Personality Testing*

The notion that job applicants can alter their responses on personality inventories has been frequently investigated in the organizational behavior literature (see Viswesvaran & Ones 1999 and Birkeland, *et al.*, 2006 for reviews). The general consensus is that it is possible to fake on non-cognitive self-report tests (Ones, Viswesvaran, & Reiss, 1996; Ellingson, Sackett, & Hough, 1999; Viswesvaran & Ones, 1999; Zickar, & Robie, 1999; McFarland & Ryan, 2000; Vasilopoulos, Reilly, & Leaman, 2000; Ellingson,

Smith, & Sackett, 2001; Morgeson, *et al.*, 2007). For example, Miller and Barrett (2008) provided evidence that police applicants could be effectively coached on how to fake on personality pre-employment tests ( $d$  ranging from .28 to .57). Meta-analyses summarizing the literature on faking on personality inventories have concluded that people completing the measure as part of a job application process tend to score higher on desirable personality traits than those filling out the inventory for other purposes (sample size weighted mean effect sizes ranging from .48 to .65; Viswesvaran & Ones, 1999), and that applicants tend to especially inflate job-relevant traits (i.e., emotional stability  $d = .44$  and conscientiousness  $d = .45$ ; Birkeland, *et al.*, 2006).

The extent to which personality scores can be faked has also been confirmed using laboratory studies that employ instructions to respond in a socially desirable way (e.g., McFarland & Ryan, 2000; Mueller-Hanson, Heggstad, & Thornton, 2003; Komar, Brown, Komar, & Robie, 2008; Peterson, Griffith, & Converse, 2009; Robie, Komar, & Brown, 2010; Hartman & Grubb, 2011). The fact that other applicants can positively distort their responses is particularly problematic for those applicants who have real desirable scores on personality measures (Mueller-Hanson, *et al.*, 2003). Because individuals with the top scores advance to further consideration, a person who is faking on a test may take the place of someone who truly has the desired personality traits (Mueller-Hanson, *et al.*, 2003). Supporting this idea, Hartman and Grubb (2011) noted that faking changes the rank-order of test takers.

Previous faking studies have frequently been couched within a framework of impression management or social desirability (Leary & Kowalski, 1990; Holden, Kroner, Fekken, & Popham, 1992). Responding in a socially desirable manner positively exaggerates one's true attributes (Holden, *et al.*, 1992; Lönnqvist, Paunonen, Tuulio-Henriksson, Lönnqvist, & Verkasalo, 2007). The theory states that test takers with no incentive will not fake on personality tests because there is nothing to be gained from portraying themselves in a different light. Job applicants, however, are highly motivated to exaggerate their good qualities, or respond as they think an employer would want them to in order to get a desired job (Zickar & Robie, 1999; Pauls & Crost, 2005; Robie, Brown, & Beaty, 2007). This supposition has been substantiated by studies showing differences in response patterns between applicants and students (e.g., Schmit & Ryan, 1993; Griffin, Hesketh, & Grayson, 2004) as well as applicants and job incumbents (Rosse, *et al.*, 1998).

#### *Effect of Feedback on Faking in Personality Testing*

Although many researchers have shown that faking on personality in-

ventories can and does occur among job applicants, it is less clear how applicants respond when taking a pre-employment assessment for a second time because they did not get the job after the first attempt. Personality testing is often only one hurdle in the application process, and consequently, a person may be denied a job due to the assessment or a mismatch with another requirement (e.g., not meeting minimum years of experience). In either case, applicants may decide to reapply. The present study examined whether or not applicants changed their responses on a personality measure when applying to an organization for a second time. The authors expected that feedback following their first attempt would play a large role in determining whether individuals fake well in subsequent administrations.

The impact of feedback on performance has traditionally been studied in the leadership, performance management, training, learning, and development literatures (Levy & Thompson, in press). In a study that stands apart, Bauer, Maertz, Jr., Dolen, and Campion (1998) investigated feedback to applicants focusing on organizational justice. The present study aimed to understand the effects of manipulating feedback and its effects in directing behavior in selection testing. The authors proposed that the type of feedback provided to the candidate about their personality test performance would influence the observed change in subsequent administrations. That is, receiving feedback stating that a poor result on the personality assessment was the reason they did not get the job would cause applicants to be more motivated to change their responses during the second attempt.

Empirical evidence, albeit limited, supports the notion that providing feedback in employment testing scenarios leads to testing performance change. For example, Landers, *et al.* (2011) found that applicants who were previously unsuccessful could increase their scores substantially by employing alternative testing strategies, such as blatant extreme responding. They attributed this finding to a coaching rumor, where applicants shared tips for how to maximize success on the test. In another previous study comparing the retesting strategies of passing and non-passing job candidates, Hausknecht (2010) found that whereas passing candidates generally answered identically across the two administrations, failing candidates were much more likely to change their answers. There is a difference between observing a change in responses the second time and deliberately faking responses in an attempt to create an ideal profile. As such, it was proposed in the present study that candidates who knew they failed would be more likely to distort their responses to resemble the ideal profile the second time they took the exam (i.e., fake well).

*Hypothesis 1.* Job applicants receiving failure feedback will demonstrate greater change in test scores between the first and second testing sessions as compared

to applicants receiving vague feedback that does not specify why they were not selected.

*Hypothesis 2.* Job applicants receiving failure feedback will demonstrate greater change in faking between the first and second testing sessions as compared to applicants receiving vague feedback that does not specify why they were not selected.

*Hypothesis 3.* Faking will partially explain the change in test scores between the first and second testing sessions.

## METHOD

### *Participants*

The focal organization, a large hospital in the Southwest, sees approximately 90,000 applicants per year. In the present study, we examined the 1,282 applicants from this pool that retested; of whom 153 failed the pre-employment assessment and 1,129 passed the assessment at Time 1. The organization employed a high selection ratio where the cutoff score differentiated those applicants who passed from those that failed with a resulting 88% passing rate. Of the retesting applicant pool, 849 applicants were female (408 male, 25 declined to identify). They were racially and ethnically diverse (Asian = 16.1%; Black = 37.7%; Hispanic = 12.1%; White = 25.3%; Other = 3.0%; 5.1% declined to identify) and were applying to occupations ranging from clinical (e.g., nurse), to administrative (e.g., secretary), to research (e.g., research data coordinator).

### *Procedure*

The present study used a quasi-experimental design, capitalizing on the applicants involved in the selection process at a large organization. All applicants in this study, whether they had previously met the assessment requirements in the first testing administration or not, were retested for employment in a second administration separated by six months from their original testing date. In this study's sample, those applicants who previously met the assessment requirements had not been hired after their first application and were retested as part of the application process to new position openings. The retesting policy of the organization stated that test scores were only valid for a period of six months regardless of the outcome of meeting or not meeting assessment requirements.

### *Measures*

*Feedback.*—Applicants who did not meet the cutoff score on the personality assessment, i.e., “failing candidates,” received feedback through an automated e-mail, generated by the system the same day that they failed to meet the personality assessment requirements. In contrast, applicants who met the cutoff score on the personality test (“passing candidates”), but did not get the job for other reasons (e.g., not meeting

minimum qualifications), were not given feedback regarding their performance on the personality measure; instead, they received an e-mail, typically sent 30 to 60 days following their test session, stating that they did not meet requirements for the position for which they applied (see Fig. 1).

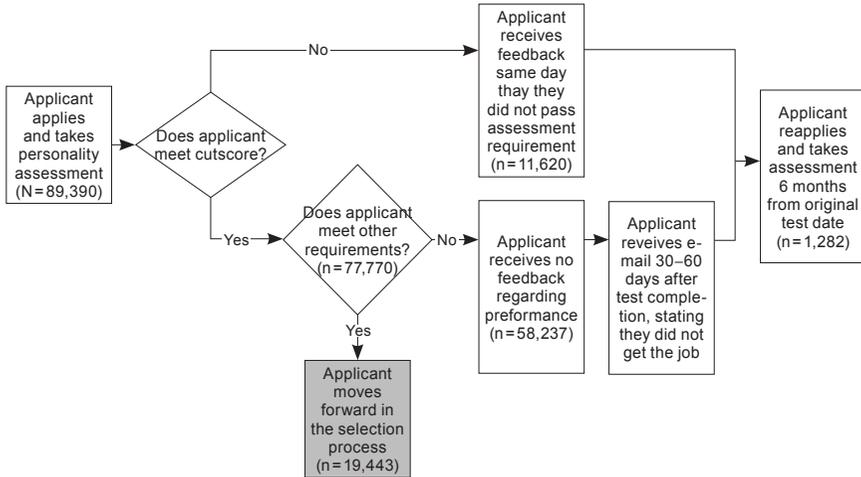


FIG. 1. Study procedure. Applicants within grey activity box were omitted from the present study.

*Personality inventory.*—The personality measure was developed proprietary to the organization by their internal industrial organizational psychologists as a selection tool to assist in hiring employees who are more likely to be team-oriented and empathetic in their approach to people, proactive, logical, and dependable in their approach to work, and tolerant of stress in their approach to change. The test constructs align most closely with conscientiousness (e.g., “Other people have often commented on how responsible I am.”) and emotional stability (e.g., “I feel anxious when my manager or supervisor changes expectations for my work”) subscales of the Big-Five personality traits, as described in more detail below. The 54 items, answered in multiple choice or true/false formats, were keyed based on the empirical data from a development sample that included item crosstab and correlation review. While these items were scored empirically, the coders also reviewed the suggested scoring to ensure the responses were rational. A number of validation studies have been conducted, though unpublished, with non-overlapping samples from the present study. Specifically, a concurrent criterion validation study ( $N = 1,593$ ) showed evidence for the test in predicting performance of current employees ( $r = .36$ ,  $p < .01$ ). Subsequent predictive studies

( $N = 1,166$ ;  $N = 895$ ) following applicants hired into the organization showed a similar relationship of test scores with actual organization performance ratings ( $r = .21$ ,  $r_c = .31$ ,  $r = .24$ ,  $r_c = .31$ ; all  $p$ 's  $< .01$ ), even with the typical skewed distribution of ratings seen in organizations. Further examination of the assessment has shown test-retest reliability ( $r = .77$ ,  $p < .01$ ) and acceptable internal consistency (Table 1; Nunnally, 1978; Robinson, Shaver, & Wrightsman, 1991).

A previous construct validation study ( $N = 150$ ) with items from the organization's proprietary measure interspersed with items from pre-established scales of personality, integrity, and resilience (Goldberg, 1999; Johnsen, Eid, Pallesen, Bartone, & Nissestad, 2009) demonstrated expected relationships: the test's emotional stability scale was positively related to both the pre-established measures of emotional stability ( $r = .51$ ,  $p < .01$ ) and challenge-related resilience ( $r = .35$ ,  $p < .01$ ). Further, the test's conscientiousness scale was positively related to both the pre-established measures of conscientiousness ( $r = .49$ ,  $p < .01$ ) and commitment-related resilience ( $r = .52$ ,  $p < .01$ ). Providing evidence of discriminant validity, no statistically significant relationships were found with either scale and general mental ability (conscientiousness  $r = .04$ , ns; emotional stability  $r = .07$ , ns).

An overall score on the assessment was used in applying the selection requirement. Applicants were given the following instructions: "Your test is in a forced-choice format. You will be presented with several possible answers for each question; please read each question carefully before answering it. Select the best answer from the choices given. If the answer you think would be correct is not offered, select the most correct answer from the available choices." While personality items do not have right or wrong answers, there are traits that are a better fit for organizations based on prediction of performance and retention. In this organization, responses were scored based upon the validation study, with "ideal" responses (those given by top performers in the organization) assigned a score of 1; the responses to the test were then summed for the candidate's score and compared to the cutoff score, the organization's minimum standard, to be classified as a passing or failing score.

*Faking scale.*—Five items (e.g., "I never make a mistake in my work") were used to assess applicants' faking behavior. These items were imbedded within the test and their ratings were summed (scored 0: Not faking or 1: Faking) to create the faking score, with higher scores indicative of an applicant who was faking.

## RESULTS

Means and correlations are presented in Table 1. Hypothesis 1 posited

that applicants receiving failure feedback would have greater changes in test scores between the first and second testing sessions as compared to applicants receiving vague feedback that does not specify why they did not get the job. The extent to which failure feedback resulted in greater increases in test scores was tested using a repeated measures analysis of variance (ANOVA) with Time 1 and Time 2 test scores as the within-subjects variable and failure feedback as the between-subjects variable. The expected interaction was observed such that test scores increased in the failure feedback condition but not in the no-failure feedback condition<sup>2</sup> (Tables 2 and 3).

TABLE 1  
STUDY MEANS, STANDARD DEVIATIONS, AND CORRELATIONS

Test	M	SD	r						
			1	2	3	4	5	6	
1. Failure feedback	1.88	.32							
2. Time 1 faking	3.23	1.18	.17*	<b>.58</b>					
3. Time 2 faking	3.33	1.20	.08*	.65*	<b>.61</b>				
4. Difference score for faking	.10	.99	-.14*	-.30*	-.14*				
5. Time 1 test score	60.09	7.83	.64*	.27*	.18*	-.15*	<b>.66</b>		
6. Time 2 test score	60.77	7.86	.32*	.17*	.22*	-.08*	.52*	<b>.74</b>	
7. Difference score for test score	.69	7.66	-.29*	-.04	.03	.10*	-.20*	-.07*	

Note.— $n = 1,281$ . Failure feedback is coded as 1: Failure feedback, 2: No failure feedback. Cronbach's alphas represented on the diagonal in bold. \* $p < .01$ .

TABLE 2  
STUDY MEANS AND STANDARD DEVIATIONS BY GROUP

Test	Failure Feedback Condition		Non-failure Feedback Condition		Student Sample	
	M	SD	M	SD	M	SD
Time 1 test score	46.41	4.19	61.94	6.20	49.92	12.74
Time 2 test score	54.01	9.11	61.69	7.20	46.48	11.50
Difference score for test score	9.76	6.56	5.49	4.29	6.68	5.48
Time 1 faking	2.71	1.21	3.31	1.15		
Time 2 faking	3.08	1.28	3.37	1.18		
Difference score for faking	0.93	0.91	0.61	0.72		

Hypothesis 2 posited that job applicants receiving failure feedback would demonstrate greater change in faking between the first and second

<sup>2</sup>Given the unequal sample sizes and unequal variances of the test scores between the failure feedback condition and the non-failure feedback condition, we retested our results with the difference scores described below using the Kruskal-Wallis non-parametric test. The results were the same.

TABLE 3  
ANALYSIS OF VARIANCE RESULTS COMPARING TIME 1 AND TIME 1 TEST SCORES AND FAKING SCORES  
FOR FAILURE FEEDBACK CONDITION VS NON-FAILURE FEEDBACK CONDITION

	<i>MS</i>	<i>df</i>	Wilks' $\lambda$	<i>F</i>	<i>p</i>	$\eta_p^2$
ANOVA for test scores						
Time		1, 1,279	.90	139.28*	< .001	0.10
Failure feedback	36,326.57	1, 1,279		556.14	< .001	0.30
Time x failure feedback		1, 1,279	.89	158.80*	< .001	0.11
ANOVA for faking						
Time		1, 1,279	.98	27.18*	< .001	0.02
Failure feedback	52.51	1, 1,279		22.98	< .001	0.02
Time x failure feedback		1, 1,279	.99	14.04*	< .001	0.01

Note.— $n = 1,281$ .  $\eta_p^2$  of .01–.06 indicates a small effect, .06–.14 indicates a medium effect, .14 or greater indicates a large effect (Gray & Kinnear, 2012). \* $p < .01$ .

testing sessions as compared to applicants receiving vague feedback that does not specify why they did not get the job. This study tested the extent to which failure feedback resulted in greater increases in faking scores using a second repeated measures ANOVA with Time 1 and Time 2 faking scores as the within-subjects variable and failure feedback as the between-subjects variable. The expected interaction was observed (Table 2 and 3). Faking scores increased to a greater extent in the failure feedback condition than in the no-failure feedback condition (Tables 2 and 3).

Hypothesis 3 was that faking scores would partially explain the difference in test scores between the first and second testing sessions. To test whether the increased faking explained the increased test scores, difference scores (Time 1–Time 2) were created for faking and test scores. Then the dependent variable (difference in test scores) was regressed on the mediator (difference in faking) and the independent variable (failure feedback). A Sobel test and bootstrap analysis were used to test for the statistical significance of the mediation using the SPSS macro provided by Preacher and Hayes (2008). Bootstrap analysis allows for the testing of indirect effects without the assumption that the indirect effect is normally distributed, avoiding problems related to asymmetric and non-normal sampling distributions that often characterize mediated relationships (MacKinnon, Lockwood, & Williams, 2004). There was a significant indirect effect of failure feedback on difference in test scores through the mediating variable of difference in faking ( $z = 3.05$ ,  $p < .01$ ; point estimate = .35,  $SE = 0.14$ ,  $99\%CI = -.77, -.04$ ). Failure feedback was positively associated with increased faking and increased faking was positively related to increased test performance.

*Supplementary study and analyses.*—In sum, the results support the hypothesis. Failure feedback was associated with increased test scores and increased faking. More importantly, faking served as an important mediator between failure feedback and test scores. One problem with the results is that the participants who received failure feedback also had much lower scores than the participants who did not receive failure feedback (on both faking and test scores). Thus, regression toward the mean might explain the increase in scores. To test for regression toward the mean we used Oldham's (1962) formula in which change in scores over time is correlated with average scores across time. A significant correlation indicates regression toward the mean. Using the average score across time [(Time 1 + Time 2)/2] creates an unbiased estimate of regression toward the mean because measurement error on the least squares slope is independent from measurement error on the mean of one's scores. In the data, there was no evidence of regression toward the mean for either test scores ( $r_{1281} = .00$ ,  $p = .90$ ) or faking ( $r_{1281} = -.02$ ,  $p = .43$ ).

As a further test of regression toward the mean, a second sample was sought out, which would be expected to score low on the test, to see if the scores increased over time as a result of regression toward the mean. Specifically, we recruited 50 students from an undergraduate psychology course at a southern university in the United States. It was expected that the Psychology students would score low on the faking and personality tests because they should be unmotivated to fake good and because their backgrounds (Psychology undergraduate) are unlikely to fit with that found in a hospital setting. Students were given information about the study and offered extra credit in the course as an incentive to participate. Those choosing to participate were given a survey containing pre-established measures of personality (Goldberg, 1999) and the proprietary measure and were instructed to create a unique code as an identifier in order to preserve anonymity. After a five-week interval, the students were asked to complete the survey a second time. Time 1 and Time 2 surveys were linked using the unique codes created by the participants. The students were told to answer truthfully on both occasions and that their responses would be used for research purposes.

The student participants (38 women, 12 men) were between the ages of 18 and 33 years ( $M = 21.2$ ,  $SD = 2.7$ ) and were racially and ethnically diverse (48% Hispanic, 18% White, 18% Black, 12% Asian, and 4% other). All of the students worked at least 20 hours per week in a variety of occupations and industries (e.g., food, healthcare, banking, and retail). The average post-high school education was 2.7 yr. ( $SD = 1.2$ ).

A repeated measures ANOVA with test time as the within-subjects

variable and feedback (applicants with failure feedback vs student sample with no feedback) as the between-subjects variable. The results showed a statistically significant interaction between feedback and testing time (Wilks'  $\lambda = .94$ ,  $F_{1,201} = 13.25$ ,  $p < .001$ ,  $\eta_p^2 = 0.06$ ; Table 2) indicating a medium effect: job applicants receiving failure feedback showed a larger increase in test scores than the student sample, even though both groups began with similar mean Time 1 scores. In the applicant group, test scores increased from Time 1 to Time 2 to a greater extent than in the student sample (Table 2). Therefore, regression toward the mean is unlikely to account for the findings. Interestingly, the student sample also showed a smaller increase in test scores than the applicants who did not receive failure feedback (Wilks'  $\lambda = .99$ ,  $F_{1,1176} = 6.61$ ,  $p < .01$ ,  $\eta_p^2 = 0.01$ ; Table 2) indicating a small effect.

TABLE 4  
TEST OF INDIRECT EFFECTS OF FAILURE FEEDBACK ON TEST SCORES THROUGH FAKING

	Step 1					Step 2				
	B	SE	$\beta$	t	p	B	SE	$\beta$	t	p
Step 1										
Failure feedback	-4.27	.49	-.29	-10.74	< .001	-4.15	.40	-.28	-10.35	<.001
Difference in faking scores						.39	.17	.06	2.19	.03
R <sup>2</sup>			.08					.09		
AdjR <sup>2</sup>			.08					.09		
SE			4.62					4.61		
F			115.34*					60.30*		
(df <sub>n</sub> , df <sub>d</sub> )			(1, 1,279)					(2, 1,278)		

Note.— $n = 1,281$ . \* $p < .01$ .

#### DISCUSSION

The literature on personality retesting in the selection setting has only begun to explore the situations and factors that may lead to changes in applicant scores over time. Understanding the implications of retesting is critical in an environment where most organizations allow applicants to retest after an elapsed period of time. The present study employed a quasi-experimental design (i.e., within-subjects design with applicants in a high stakes situation) allowing for a true test of the retesting effect in an unproctored assessment setting. With this design, the present study explored the contextual factor of feedback on faking personality scores.

Although personality test scores were positively correlated from Time 1 to Time 2, results revealed that the feedback provided to applicants did

affect the extent to which individuals changed their scores in the second administration. That is, the greatest difference between Time 1 and 2 scores was observed for those receiving failure feedback, whereas those receiving no feedback about why they were not selected for the job showed relatively less change across administrations. In addition, the results showed a statistically significant mediating effect of faking that explained the relationship between failure feedback and score change. The additional study and analyses further underscored the role of the feedback-based motivation of the test taker. Those who received feedback that they had done poorly on the personality inventory had the lowest correlation between Time 1 and Time 2 scores. In contrast, the results provide evidence that personality scores remained unchanged for applicants receiving no feedback about how they fared on the first personality assessment.

In situations where improvement over time is desired (e.g., in job performance), it makes sense to encourage the provision of timely and specific feedback. However, the question remains whether the goal of organizations is to improve the testing performance of applicants who did not meet minimum expectations in their initial application. This is particularly true of personality tests, given the fact that personality is not expected to actually change or "improve" from one time to the next (Chmielewski & Watson, 2009). The results of this study indicate job candidates, particularly when given specific feedback that they did not meet assessment requirements, are likely to change their responses the second time they attempt the test and do so by faking (as opposed to choosing random responses).

Given the likelihood that applicants will employ faking strategies on personality-based selection assessments, organizations have several options to deal with this problem. For example, McFarland, Ryan, and Ellis (2002) found that simply randomizing the order of personality items (rather than having items grouped by trait) made the inventory harder to fake. Others have found that a verification warning helps to reduce faking behavior on exams (Vasilopoulos, Cucina, & McElreath, 2005; Robie, Taggar, & Brown, 2009). Informing the employee that former employers or other sources will be used to verify personality scores may be enough to elicit honest answers. There are also statistical means of dealing with faking, including adjusting scores (Hough, 1998; Ellingson, *et al.*, 1999), cut-score strategies (Berry & Sackett, 2009), and the use of response latencies (Holden & Hibbs, 1995). Careful test construction incorporating forced-choice items (Jackson, Wroblewski, & Ashton, 2000; Christiansen, Burns, & Montgomery, 2005) and empirical scoring (Sisco & Reilly, 2007) has also proven to be a viable alternative. The results of the present study suggest that delaying or minimizing the specificity of the feedback provided to applicants could serve as an additional strategy that should be added to the above practices.

## REFERENCES

- ARTHUR, W. R., GLAZE, R. M., VILLADO, A. J., & TAYLOR, J. E. (2009) Unproctored Internet-based tests of cognitive ability and personality: magnitude of cheating and response distortion. *Industrial and Organizational Psychology: Perspectives on Science and Practice*, 2(1), 39-45.
- 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., & JUDGE, T. A. (2001) Personality and performance at the beginning of the new millennium: what do we know and where do we go next? *Personality and Performance*, 9, 9-30.
- BARTRAM, D., & BROWN, A. (2004) Information exchange article: online testing: mode of administration and the stability of OPQ 32i scores. *International Journal of Selection and Assessment*, 12(3), 278-284.
- BAUER, T. N., MAERTZ, C. P., JR., DOLEN, M. R., & CAMPION, M. A. (1998) Longitudinal assessment of applicant reactions to employment testing and test outcome feedback. *Journal of Applied Psychology*, 83(6), 892-903.
- BERRY, C. M., & SACKETT, P. R. (2009) Faking in personnel selection: tradeoffs in performance versus fairness resulting from two cut-score strategies. *Personnel Psychology*, 62(4), 835-863.
- BIRKELAND, S. A., MANSON, T. M., KISAMORE, J. L., BRANNICK, M. T., & SMITH, M. A. (2006) A meta-analytic investigation of job applicant faking on personality measures. *International Journal of Selection and Assessment*, 14(4), 317-335.
- BROWN, A., BARTRAM, D., HOLTZHAUSEN, G., CARSTAIRS, J., & MYLONIS, G. (2005, April) Online personality and motivation testing: is unsupervised administration an issue? Paper presented at the 20th annual conference of the Society for Industrial and Organizational Psychology, Los Angeles, CA.
- CHAPMAN, D. S., & WEBSTER, J. (2003) The use of technologies in the recruiting, screening, and selection processes for job candidates. *International Journal of Selection and Assessment*, 11(2-3), 113-120.
- CHMIELEWSKI, M., & WATSON, D. (2009) What is being assessed and why it matters: the impact of transient error on trait research. *Journal of Personality & Social Psychology*, 97(1), 186-202.
- CHRISTIANSEN, N. D., BURNS, G. N., & MONTGOMERY, G. E. (2005) Reconsidering forced-choice item formats for applicant personality assessment. *Human Performance*, 18(3), 267-307.
- ELLINGSON, J. E., SACKETT, P. R., & CONNELLY, B. S. (2007) Personality assessment across selection and development contexts: insights into response distortion. *Journal of Applied Psychology*, 92(2), 386-395.
- ELLINGSON, J. E., SACKETT, P. R., & HOUGH, L. M. (1999) Social desirability corrections in personality measurement: issues of applicant comparison and construct validity. *Journal of Applied Psychology*, 84, 155-166.
- ELLINGSON, J. E., SMITH, D., & SACKETT, P. R. (2001) Investigating the influence of social desirability on personality factor structure. *Journal of Applied Psychology*, 86(1), 122-133.

- GOLDBERG, L. R. (1999) A broad-bandwidth, public domain, personality inventory measuring the lower-level facets of several five-factor models. In I. Mervielde, I. Deary, F. D. Fruyt, & F. Ostendorf (Eds.), *Personality Psychology in Europe*. Tilburg, The Netherlands: University Press. Pp. 7-28.
- GRAY, C. D., & KINNEAR, P. R. (2012) *IBM SPSS Statistics 19 Made Simple*. Psychology Press: New York.
- GRIFFIN, B., HESKETH, B., & GRAYSON, D. (2004) Applicants faking good: evidence of item bias in the NEO PI-R. *Personality and Individual Differences*, 36(7), 1545-1558.
- GUION, R. M., & GOTTIER, R. F. (1965) Validity of personality measures in personnel selection. *Personnel Psychology*, 18, 135-164.
- GUO, J., & DRASGOW, F. (2010) Identifying cheating on unproctored Internet tests: the Z-test and the Likelihood Ratio Test. *International Journal of Selection and Assessment*, 18(4), 351-364.
- HAMPSON, S. E., & GOLDBERG, L. R. (2006) A first large cohort study of personality trait stability over the 40 years between elementary school and midlife. *Journal of Personality & Social Psychology*, 91(4), 763-779.
- HARTMAN, N. S., & GRUBB, W. (2011) Deliberate faking on personality and emotional intelligence measures. *Psychological Reports*, 108(1), 120-138.
- HAUSKNECHT, J. P. (2010) Candidate persistence and personality test practice effects: implications for staffing system management. *Personnel Psychology*, 63(2), 299-324.
- HOGAN, J., BARRETT, P., & HOGAN, R. (2007) Personality measurement, faking, and employment selection. *Journal of Applied Psychology*, 92(5), 1270-1285.
- HOLDEN, R. R., & HIBBS, N. (1995) Incremental validity of response latencies for detecting fakers on a personality test. *Journal of Research in Personality*, 29(3), 362-372.
- HOLDEN, R. R., KRONER, D. G., FEKKEN, G., & POPHAM, S. M. (1992) A model of personality test item response dissimulation. *Journal of Personality and Social Psychology*, 63(2), 272-279.
- HOUGH, L. M. (1998) The millennium for personality psychology: new horizons or good ole daze. *Applied Psychology: An International Review*, 47(2), 233-261.
- JACKSON, D. N., WROBLEWSKI, V. R., & ASHTON, M. C. (2000) The impact of faking on employment tests: does forced choice offer a solution? *Human Performance*, 13(4), 371-388.
- JOHNSEN, B., EID, J., PALLESEN, S., BARTONE, P. T., & NISSESTAD, O. A. (2009) Predicting transformational leadership in Naval cadets: effects of personality hardiness and training. *Journal of Applied Social Psychology*, 39(9), 2213-2235.
- KOMAR, S., BROWN, D. J., KOMAR, J. A., & ROBIE, C. (2008) Faking and the validity of conscientiousness: a Monte Carlo investigation. *Journal of Applied Psychology*, 93(1), 140-154.
- LANDERS, R. N., SACKETT, P. R., & TUZINSKI, K. A. (2011) Retesting after initial failure, coaching rumors, and warnings against faking in online personality measures for selection. *Journal of Applied Psychology*, 96(1), 202-210.
- LEARY, M. R., & KOWALSKI, R. M. (1990) Impression management: a literature review and two-component model. *Psychological Bulletin*, 107(1), 34-47.

- LEVY, P. E., & THOMPSON, D. J. (in press) The role of feedback in organizations. In R. Sutton, M. Hornsby, & K. Douglas (Eds.), *Feedback: the handbook of praise, criticism, and advice*. New York: Peter Lang.
- LÖNNQVIST, J., PAUNONEN, S., TUULIO-HENRIKSSON, A., LÖNNQVIST, J., & VERKASALO, M. (2007) Substance and style in socially desirable responding. *Journal of Personality*, 75(2), 291-322.
- MACKINNON, D. P., LOCKWOOD, C. M., & WILLIAMS, J. (2004) Confidence limits for the indirect effect: distribution of the product and resampling methods. *Multivariate Behavioral Research*, 39(1), 99-128.
- MCCRAE, R. R., & COSTA, P. T. (1994) The stability of personality: observation and evaluations. *Current Directions in Psychological Science*, 3(6), 173-175.
- McFARLAND, L. A., & RYAN, A. (2000) Variance in faking across noncognitive measures. *Journal of Applied Psychology*, 85(5), 812-821.
- McFARLAND, L. A., RYAN, A., & ELLIS, A. (2002) Item placement on a personality measure: effects on faking behavior and test measurement properties. *Journal of Personality Assessment*, 78(2), 348-369.
- MILLER, C. E., & BARRETT, G. V. (2008) The coachability and fakability of personality-based selection tests used for police selection. *Public Personnel Management*, 37(3), 339-351.
- MORGESON, F. P., CAMPION, M. A., DIPBOYE, R. L., HOLLENBECK, J. R., MURPHY, K., & SCHMITT, N. (2007) Reconsidering the use of personality tests in personnel selection contexts. *Personnel Psychology*, 60(3), 683-729.
- MUELLER-HANSON, R., HEGGESTAD, E. D., & THORNTON, G. (2003) Faking and selection: considering the use of personality from select-in and select-out perspectives. *Journal of Applied Psychology*, 88(2), 348-355.
- NUNNALLY, J. C. (1978) *Psychometric theory*. (2nd ed.) New York: McGraw Hill.
- NYE, C. D., DO, B., DRASGOW, F., & FINE, S. (2008) Two-step testing in employee selection: is score inflation a problem? *International Journal of Selection and Assessment*, 16(2), 112-120.
- OLDHAM, P. D. (1962) A note on the analysis of repeated measurements of the same subjects. *Journal of Chronic Diseases*, 15(10), 969-977.
- ONES, D. S., DILCHERT, S., VISWESVARAN, C., & JUDGE, T. A. (2007) In support of personality assessment in organizational settings. *Personnel Psychology*, 60(4), 995-1027.
- 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(6), 660-679.
- PAULS, C. A., & CROST, N. W. (2005) Effects of different instructional sets on the construct validity of the NEO-PI-R. *Personality and Individual Differences*, 39(2), 297-308.
- PETERSON, M. H., GRIFFITH, R. L., & CONVERSE, P. D. (2009) Examining the role of applicant faking in hiring decisions: percentage of fakers hired and hiring discrepancies in single- and multiple-predictor selection. *Journal of Business and Psychology*, 24(4), 373-386.
- PREACHER, K. J., & HAYES, A. F. (2008) Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. *Behavior Research Methods*, 40, 879-891.

- ROBERTS, B. W., & DELVECCHIO, W. F. (2000) The rank-order consistency of personality traits from childhood to old age: a quantitative review of longitudinal studies. *Psychological Bulletin*, 126(1), 3-25.
- ROBERTS, B. W., WALTON, K. E., & VIECHTBAUER, W. (2006) Patterns of mean-level change in personality traits across the life course: a meta-analysis of longitudinal studies. *Psychological Bulletin*, 132(1), 1-25.
- ROBIE, C., BROWN, D. J., & BEATY, J. C. (2007) Do people fake on personality inventories? A verbal protocol analysis. *Journal of Business and Psychology*, 21(4), 489-509.
- ROBIE, C., KOMAR, S., & BROWN, D. J. (2010) The effects of coaching and speeding on Big-five and Impression Management Scale scores. *Human Performance*, 23(5), 446-467.
- ROBIE, C., TAGGAR, S., & BROWN, D. J. (2009) The effects of warnings and speeding on scale scores and convergent validity of conscientiousness. *Human Performance*, 22(4), 340-354.
- ROBINSON, J. P., SHAVER, P. R., & WRIGHTSMAN, L. S. (EDS.) (1991) *Measures of personality and social psychological attitudes*. San Diego: Academic Press.
- ROSSE, J. G., STECHER, M. D., MILLER, J. L., & LEVIN, R. A. (1998) The impact of response distortion on pre-employment personality testing and hiring decisions. *Journal of Applied Psychology*, 83(4), 634-644.
- SCHMIT, M. J., & RYAN, A. M. (1993) The Big-five in personnel selection: factor structure in applicant and nonapplicant populations. *Journal of Applied Psychology*, 78(6), 966-974.
- SISCO, H., & REILLY, R. R. (2007) Five-factor Biodata Inventory: resistance to faking. *Psychological Reports*, 101(1), 3-17.
- TEMPLER, K. J., & LANGE, S. R. (2008) Internet testing: equivalence between proctored lab and unproctored field conditions. *Computers in Human Behavior*, 24(3), 1216-1228.
- TIPPINS, N. T., BEATY, J., DRASGOW, F., GIBSON, W. M., PEARLMAN, K., SEGALL, D. O., & SHEPHERD, W. (2006) Unproctored Internet testing in employment settings. *Personnel Psychology*, 59(1), 189-225.
- VASILOPOULOS, N. L., CUCINA, J. M., & McELREATH, J. M. (2005) Do warnings of response verification moderate the relationship between personality and cognitive ability? *Journal of Applied Psychology*, 90(2), 306-322.
- VASILOPOULOS, N. L., REILLY, R. R., & LEAMAN, J. A. (2000) The influence of job familiarity and impression management on self-report measure scale scores and response latencies. *Journal of Applied Psychology*, 85(1), 50-64.
- VISWESVARAN, C., & ONES, D. S. (1999) Meta-analyses of fakability estimates: implications for personality measurement. *Educational and Psychological Measurement*, 59(2), 197-210.
- YOUNG, M. C. (2003) Effects of retesting on a new army measure of motivational attributes: implications for response distortion, test validity, and operational use. Paper presented at Annual Meeting of the International Public Management Association Assessment Council, Baltimore, MD.
- ZICKAR, M. J., & ROBIE, C. (1999) Modeling faking good on personality items: an item-level analysis. *Journal of Applied Psychology*, 84(4), 551-563.