

Candidate Persistence and Personality Test Practice Effects:  
Implications for Staffing System Management

John P. Hausknecht

Cornell University

In Press: *Personnel Psychology*

## Abstract

Candidates persist in selection settings for numerous reasons, prompting several concerns regarding staffing system management. Predictors of the propensity to retest and personality test practice effects were investigated among a sample of 15,338 candidates who applied for supervisory positions (and 357 who repeated the selection process) over a four-year period with a large organization in the service industry. Results reveal greater likelihood of retesting among internal candidates and overall evidence of small to moderate personality test practice effects. Compared to passing candidates who retested for various reasons, failing candidates pursued alternative response strategies upon retesting and generated dimension-level practice effects that reached .40 to .60, whereas passing candidates generally replicated their initial profiles. For several subscales, low initial scores were associated with practice effects that exceeded a full standard deviation. Implications for research, practice, and policy are discussed.

**KEYWORDS:** personnel selection, practice effects, personality testing, response distortion

Candidate Persistence and Personality Test Practice Effects:  
Implications for Staffing System Management

Many unsuccessful job applicants repeat the selection process, raising numerous questions about how organizations should manage candidate retesting. Although the topic has drawn attention from selection researchers in recent years (e.g., Ellingson, Sackett, & Connelly, 2007; Hausknecht, Halpert, DiPaolo, & Moriarty Gerrard, 2007; Lievens, Reeve, & Heggstad, 2007; Lievens, Buyse, & Sackett, 2005; Raymond, Neustel, & Anderson, 2007), numerous issues continue to lack empirical scrutiny. This is particularly troublesome because practical recommendations concerning appropriate retesting policies are being put forth without the benefit of empirical research (Bourdeau, 2008), suggesting an urgency among staffing specialists who must deal with the complexities that are created when candidates persist.

This study examines two consequential yet unresolved issues concerning selection retesting. One goal is to develop a better understanding of the retest population by examining predictors of the propensity to retest. Such information gives researchers insight into the mechanisms underlying practice effects and informs practitioner discussions concerning retesting policies. The second aim is to extend practice effects research to non-cognitive predictors. Practice effects (or retest effects) refer to score changes associated with completing multiple forms of a selection test (Hausknecht, Trevor, & Farr, 2002; Lievens et al., 2005). Personality test practice effects are assessed along with several theoretically-grounded predictors of score change. These issues are examined using a sample of candidates who applied for supervisory positions with a large service organization over approximately four years.

## Theory and Hypotheses

### *Predicting the Propensity to Retest*

Although researchers have examined what happens to test scores when candidates retest, Lievens and colleagues (2005) called for better understanding of why some failing candidates give up while others persist. No study to date has examined factors that predict who repeats the selection process, although the findings would have implications for research and practice. One possibility is that repeated testing implies greater motivation among those candidates who persist. Because this group has gone to greater lengths to secure an offer, they may be more committed to the job and organization. Consequently, this heightened commitment may translate into enhanced performance and retention for those candidates who are eventually selected (Hausknecht et al., 2002). Thus, to the extent that retesting signifies greater applicant motivation or persistence, staffing policies and practices could be modified accordingly to facilitate applicant retesting, perhaps by adopting shorter retest intervals or proactively marketing retest opportunities.

Conversely, retesting may allow candidates to misrepresent themselves in order to gain entry into the organization (Hogan, Barrett, & Hogan, 2007; McFarland & Ryan, 2000; Landers, Sackett, & Tuzinski, 2009). The organization may inadvertently facilitate this process. For example, some companies provide internal but not external applicants with test-specific feedback for developmental purposes. If internal candidates exploit the information for score improvement rather than development, they would be more confident in their ability to improve scores, and thus more likely to retest. Thus, coupling information about retesting odds with data concerning practice effects would help staffing specialists build a case that particular features of the selection system work at cross-purposes. They may decide, for example, to add a provision to

staffing policies that candidates who receive test-specific feedback and repeat the selection process will be assessed using a different form of assessment. Overall, understanding the propensity to retest helps answer policy questions about whether (or under what conditions) organizations should allow candidate retesting (e.g., Society for Industrial and Organizational Psychology, 2003) and helps researchers understand the meaning of practice effects (e.g., Sackett, Burris, & Ryan, 1989). Several applicant characteristics are examined here as predictors of the propensity to retest<sup>1</sup>.

*Candidate personality.* As mentioned above, retesting behavior may indicate heightened candidate motivation toward securing employment with the organization. Although there are no direct empirical tests concerning predictors of retesting behavior, the motivation-based hypothesis is consistent with rationale found in previous retesting research. In a study of law enforcement candidates, Hausknecht et al. (2002) found that candidates who had to repeat the selection process before gaining entry into the organization demonstrated higher training performance scores and lower turnover risk than those who entered based on a single attempt. The authors argued that greater motivation and commitment among the retest sample explained why retesting was associated with better performance and lower turnover. Although the findings were consistent with this logic, motivation-related constructs were not directly measured. Thus, to build upon this work, this study examines whether persevering and determined individuals are indeed more likely to persist in the face of failure. Most personality inventories contain sub-facets that address elements of persistence, which are often classified under the broader dimension of conscientiousness. For example, Costa, McCrae, and Dye (1991) defined “self-

---

<sup>1</sup> Note that the initial test outcome naturally influences the propensity to retest, as passing candidates rarely retest. (however, as discussed later in the manuscript, there are reasons why passing candidate sometimes retest). Most retest candidates are those who failed the first attempt and were permitted to repeat the selection process after some interval.

discipline” in terms of persistence and as a conscientiousness sub-facet. Concerning the measure studied here, the Gordon Personal Profile and Inventory (GPPI), the “responsibility” facet assesses perseverance and determination, and maps onto the broader conscientiousness construct (Gordon, 1993). To the extent that retesting behavior is indeed indicative of motivation and persistence, high scorers on the responsibility facet should be more likely to repeat the selection process following an unsuccessful first attempt.

*Hypothesis 1:* The odds of retesting will be higher for applicants who score higher on the responsibility scale.

*External/internal status.* A second dimension that may explain retesting likelihood concerns whether the candidate is external or internal to the organization. In a retest context, and when considering organizations that adopt both external and internal sourcing strategies, the *spatial mismatch hypothesis* suggests that external candidates would be disadvantaged relative to their internal counterparts. Researchers in sociology and urban studies developed this perspective to suggest that the spatial separation of people and jobs explains labor market outcomes such as employment and earnings. Building upon seminal work in this area (Kain, 1968), researchers contend that (among other predictions) physical closeness to employment opportunities increases the odds of finding a job, in part because greater distance reduces available information on vacancies, and because distant workers must endure higher search costs (Gobillon, Selod, & Zenou, 2007; Holzer, 1991; Ihlanfeldt, 1997; Wang, 2003). Thus, because of their close proximity to the internal job market and low search costs, internal candidates are expected to take advantage of retest opportunities at a greater rate than external candidates who, since applying initially, have become physically separated from the organization and do not have the same internal networks to leverage for information about future openings.

Another line of reasoning based on the availability of test-related feedback further supports this prediction. As discussed earlier, some organizations provide detailed test feedback only to internal candidates based on the logic that such information could benefit the candidate's development. However, Sackett et al. (1989) noted that test-related feedback reveals the characteristics sought by the organization and allows candidates to adopt alternative response strategies upon retesting. Thus, delivering feedback to internal candidates may instill in them greater self-efficacy concerning future test performance, which would increase their odds of retesting. On the other hand, absent test-specific feedback, external candidates would have little basis for revising responses, and thus should have less motivation to retest. Taken together, the spatial separation and feedback availability arguments both suggest that internal candidates should be more likely to retest.

*Hypothesis 2:* The odds of retesting will be higher for internal candidates versus external candidates.

#### *Personality Test Practice Effects in Selection Contexts*

Although survey data suggest that organizations use cognitive ability and personality tests at nearly the same rate (American Management Association, 2001), most practice effects research involves cognitive measures. Thus, the second goal of this study is to examine personality test practice effects in an actual selection setting. Drawing from several recent studies that addressed personality test practice effects in operational or "high stakes" (vs. experimental/laboratory) selection settings (Ellingson et al., 2007; Hogan et al., 2007; Landers et al., 2009), and from elements of retesting frameworks found elsewhere in the literature (Knowles, Coker, Scott, Cook, & Neville, 1993; Lievens et al., 2005), two basic arguments are put forth to explain why personality test practice effects are likely.

The *true change* argument suggests that personality test scores change over time because of actual change in adult personality. Although some researchers argue that small changes do indeed occur, particularly over large intervals (10-20 years) and for individuals in early and very late adulthood (Ardelt, 2000), others contend that personality is fairly immutable (Costa & McCrae, 1988). Ellingson and colleagues (2007) provide empirical support for the latter argument. The authors examined personality test practice effects across intervals ranging from 12 days to 7 years and concluded that although there was some evidence of practice effects, the amount of time elapsed between assessments was unrelated to score changes. Following this line of evidence and the logic of the true change argument, candidates who retest on personality measures in selection contexts should exhibit stable scores over time because the retest interval is too short for actual personality change to occur. In other words, “scores simply should not drift when nothing happens” (Knowles et al., 1996, p. 352).

Conversely, the *response distortion* argument may explain personality test practice effects in selection contexts. Response distortion has been defined as “deliberate tailoring of answers to create a positive impression” (Rosse, Stecher, Miller, & Levin, 1998, p. 636). The response distortion argument suggests less personality measure stability (i.e., more within-person change) because candidates realize the contingencies between their responses and the likelihood of receiving a job offer (Birkeland, Manson, Kisamore, Brannick, & Smith, 2006; Ellingson et al., 2007; Leary & Kowalski, 1990). Such a scenario represents a rational response to the initial selection decision. Candidates who generate profiles that are incongruent with the employer’s desired characteristics would be unwise to respond similarly upon retesting since the same (undesirable) outcome would result.



Only a handful of studies examine response distortion in operational selection contexts using within-person designs<sup>2</sup>. Two recent studies reported fairly small average practice effects across personality dimensions (Ellingson et al., 2007; Hogan et al., 2007). However, important differences between these two studies and the present investigation suggest practice effects are context-dependent and potentially larger than previously believed. For example, Hogan and colleagues studied practice effects among external applicants to a customer service job, and found practice effects ( $d$ ) on the Hogan Personality Inventory (HPI) that never exceeded .10. However, internal candidates may generate larger practice effects than the external applicants studied by Hogan et al. (2007). Internal candidates can benefit from test-specific feedback and can consult with co-workers for strategies and tips to help them develop alternative response strategies (i.e., “grapevine coaching”; Sackett et al., 1989). Further, internal candidates are often well-aware of the extent to which selection decisions are tied to test performance. Thus, when candidates compete for internal promotions or managerial positions, as is the case here, larger score changes may be observed because of selection process transparencies. They have more invested in the outcome because of its career-advancing potential and have greater access to test feedback and tips from internal colleagues.

Ellingson et al. (2007) studied practice effects on the California Psychological Inventory (CPI) among 713 individuals who represented a range of jobs and organizations. Relative to other combinations of test purpose and sequence, candidates who completed a personality inventory first for development purposes and second for selection purposes produced the largest

---

<sup>2</sup> Outside of operational selection contexts, there is a large literature on personality test response distortion that has been conducted in laboratory settings and/or using between-person designs (see for reviews Alliger & Dwight, 2000; Birkeland et al., 2006; Dilchert, Ones, Viswesvaran, & Deller, 2006; Viswesvaran & Ones, 1999). Following arguments provided by Ellingson et al. (2007, p. 386-388), only studies that are germane to the issue of within-person retesting and practice effects in operational selection settings are reviewed here.

practice effects (average  $d = .26$ ). This group also generated larger effect sizes for 17 of 20 personality dimensions when compared to those who completed the personality inventory for development purposes on both occasions. More broadly, these findings illustrate that different factors influence practice effects depending on the testing context. Practice effects may be larger in high-stakes selection settings where the outcome is desirable and tied to test performance.

More recently, Landers et al. (2009) studied 470 managers who completed a personality measure twice as part of a promotional test battery for a large retailer and found practice effects ( $d$ ) that ranged from .72 to 1.17 across dimensions. The authors identified several possible explanations for the observed practice effects. One possibility is that after an unsuccessful initial attempt, candidates revise initially-honest (but unsuccessful) responses toward what they believe is an ideal personality profile for the job. Alternatively, candidates may revise initially-distorted profiles toward extremely honest responses. Regardless of their strategy, the conceptual rationale outlined above and recent studies suggest that score changes are likely when applicants retest in an actual selection setting.

*Hypothesis 3: Candidates who retest on a personality measure after initial failure will exhibit non-zero change on personality subscales.*

#### *Personality Test Practice Effects for Passing and Failing Candidates*

Candidates who retest are often, but not always, those who failed their first attempt. Individuals who initially earn a passing score may retest for several reasons. For instance, numerous organizations maintain “eligibility lists” where scores are rendered outdated after a certain period (AERA, APA, & NCME, 1999). Scarce job openings would require passing candidates to retest when they are not selected or promoted during the eligibility period. A change in the organization’s testing policies and procedures can also compel passing candidates

to retest. For example, one study reports that a large group of managers was allowed to retest after the organization clarified that test scores would affect salary and promotion decisions following a major restructuring effort (Tuzinski, Laczko, & Sackett, 2005).

Regardless of the means by which passing candidates eventually choose to retest, they should produce very similar personality profiles across testing occasions because their initial response pattern matched the desired profile for the job (Sackett et al., 1989). Classic motivation models such as the *law of effect* assert that people tend to repeat behaviors that generate positive outcomes and learn not to repeat those that lead to aversive outcomes (Baum, 1973; Herrnstein, 1970). In retesting contexts, the behavior-reward contingencies associated with receiving a passing score after the initial attempt would prompt passing candidates to repeat the same response pattern upon retesting. On the other hand, initially unsuccessful candidates should learn from the first attempt that the same response pattern will produce the same undesirable failing outcome. This natural grouping provides a novel test of response distortion arguments, as each group of candidates has a distinct motive upon retesting; passing candidates are motivated to remain the same, failing candidates are motivated to change. Thus, assuming that passing the test battery is desirable to candidates who retest, the group of failing candidates would be more likely than passing candidates to change their responses upon retesting.

*Hypothesis 4:* Personality test practice effects will be larger for failing versus passing candidates.

#### *Initial Scores and Personality Test Practice Effects*

Finally, among failing candidates, the degree of score change on any given scale may be contingent on the initial score that is received, particularly when candidates need not always improve on every dimension of the personality inventory. When the personality measure

involves a forced-choice scoring format, as is the case here, candidates cannot produce extreme score increases on all dimensions. Faced with choices, and consistent with the motivation-based arguments presented above, failing candidates would be expected to target score increases toward dimensions upon which they performed relatively poorly on the initial attempt, thereby exhibiting larger practice effects on those dimensions where initial scores were relatively low.

Landers and colleagues (2009) correlated initial scores with the magnitude of personality test practice effects and found that low initial scores were associated with the highest gains. The authors concluded that these relationships could not be readily explained as a ceiling effect phenomenon. That is, negative correlations might be expected when the full range of scorers retests because high scorers have little opportunity to raise scores; however, given that high scorers were generally absent from the retest sample, they argued that ceiling effects would not fundamentally constrain the range of practice effects, and that the relationship between initial scores and practice effects provides substantive evidence that low scorers improve most. Thus, given the rationale cited above and recent empirical work, the largest score gains are expected in instances where the initial score on any given scale is relatively low.

*Hypothesis 5:* Initial scores on any given personality subscale will be negatively related to the magnitude of practice effects.

## Method

### *Sample*

The total sample included 15,338 applicants for supervisory positions at various hotel and casino properties located throughout the United States between July 2003 and March 2007. The total sample was primarily male (58%) and White (65%). A majority of applicants were internal

candidates seeking promotion to a supervisory role (55%), while the remaining candidates were external (45%).

### *Procedure*

*Candidate flow.* Figure 1 illustrates the progression of the 15,338 applicants through the selection and retesting process. After the first attempt, 3,161 candidates failed the test battery, and 12,177 passed. In total, 357 candidates retested. Of these, 301 had failed the first attempt and 56 had passed. Passing candidates retested because of time lag, location, and/or position-related issues. When the test battery was first adopted, passing test scores were considered valid for 12 months, but this policy was later revised to allow such scores to remain valid indefinitely. Because of the former policy, some candidates who initially passed the test battery (but were never promoted) retested because more than 12 months had passed since the first attempt. In other instances, candidates passed the test battery initially, but then re-applied (and retested) at a different property and/or for different positions over time.

Predictions involving the propensity to retest (Hypotheses 1 and 2) were tested using data from 2,535 candidates who failed their first attempt and who had complete data for all model variables. The practice effects prediction (Hypothesis 3) was examined with the subsample of 301 failing candidates. The hypothesis that practice effects would be larger for failing candidates versus passing candidates (Hypothesis 4) was tested with the subsamples of 56 passing and 301 failing candidates. The relationship between initial scores and practice effects (Hypothesis 5) was tested using data from 265 of the 301 failing candidates who had complete information on all model variables.

*Selection process information.* The organization maintained a six-month waiting period before candidates were permitted to repeat the selection process. Test records show that this

policy was followed in 98% of cases where retesting occurred (six candidates retested at shorter intervals, ranging from two to five months). The median time lapse between the first and second test administration was 10.0 months ( $Mean = 13.2$ ,  $SD = 8.2$ ,  $Min = 2.0$ ,  $Max = 40.0$ ). Candidates retested on their own time rather than company time. Pass/fail status was determined by an overall composite test score that weighted the personality (65%) and cognitive ability (35%) test components based on a concurrent validation study that was conducted prior to the exam's operational use (specific weights: abstract reasoning=15%, quantitative reasoning=20%, ascendancy=15%, personal relations=20%, original thinking=10%, vigor=20%). The organization provided pass/fail feedback to external candidates, and gave internal candidates pass/fail feedback along with a three-page narrative summary of their performance for development purposes. Fifteen candidates completed the test battery a third time, but given the small number, no analysis of this group is presented.

### *Measures*

*Decision to retest.* Candidates who did not retest (i.e., those who completed only one test) were coded "0", and candidates who retested at any point during the study window were coded "1."

*Personality scales.* Applicants completed the Gordon Personal Profile and Inventory (GPPI), which is a partially-ipsative measure that contains 152 items assessing aspects of personality that are important to the daily functioning of a normal person (Gordon, 1993). Items are organized into 38 four-item "tetrads", and applicants are asked to indicate which of the four descriptive phrases is "most like you" and which is "least like you." Responses are then scored and classified into one of the following eight primary scales: *ascendancy* (18 items); *responsibility* (18 items); *emotional stability* (18 items); *sociability* (18 items); *cautiousness* (20

items); *original thinking* (20 items); *personal relations* (20 items); and *vigor* (20 items). The test manual (Gordon, 1993) reports that six of the GPPI factors correspond to dimensions of the five-factor personality model (ascendancy and sociability reflect extroversion, responsibility reflects conscientiousness, emotional stability reflects emotional stability, original thinking reflects openness to experience, and personal relations reflects agreeableness). Applicants completed identical forms of the GPPI on both occasions. The GPPI has been used in selection settings for many decades (e.g., Bass, 1957; Gordon & Stapleton, 1956; Schwab & Packard, 1973). Unlike fully-ipsative measures, partially-ipsative measures such as the GPPI yield scores that are highly correlated with measures that are based on normative response formats (Heggstad, Morrison, Reeve, & McCloy, 2006).

*Cognitive ability*<sup>3</sup>. The test battery included two measures designed by the organization to assess two facets of cognitive ability. First, a 25-item *quantitative reasoning* test was included to measure understanding of job-related arithmetic skills (addition, subtraction, multiplication, and division) using multiple-choice questions. The second component was a 30-item *abstract reasoning* test. Each item contained four problem figures that illustrated a particular sequence or series, and candidates had to identify out of five possible options the correct figure that logically would appear next in the series. Identical forms were used across administrations. Based on the sample of 357 retest candidates, the test-retest correlation was .65 for each measure.

*Pass/fail status*. Candidates who passed the test battery on their first attempt were coded “0” and failing candidates were coded “1.”

*External/internal status*. Candidates who were external to the organization were coded “0” and internal candidates were coded “1.”

---

<sup>3</sup> Although personality test practice effects represent the focus of this study, descriptive analyses of cognitive ability practice effects are presented in Table 2.

*Control variables.* Three variables were controlled in the regression models. Test records were consulted for information regarding candidate *race* and *sex*. Whites and males were coded “1” , and non-Whites and females were coded “0”. The *time lag* between the first and second test administrations (in months) was controlled to account for potential time-based effects.

## Results

### *Propensity to Retest*

The decision to retest was treated as a binary outcome and modeled using multiple logistic regression to test whether high scorers on responsibility (Hypothesis 1) and internal candidates (Hypothesis 2) had greater odds of retesting. These analyses involve the subset of failing participants from the total sample who had complete predictor data ( $N = 2,535$ ). The decision to retest was regressed upon the responsibility subscale score (obtained upon initial testing), external/internal status, and the race and sex controls. Results are shown in Table 1. Coefficients are presented in two forms, typical regression coefficients ( $B$ ) and odds ratios ( $\exp(B)$ ). The statistical significance of each predictor is calculated in part based on the magnitude of the regression coefficient, but these are less interpretable because they represent linear increments in the *logit* given a one-unit increment in the predictor (Cohen, Cohen, West, & Aiken, 2003). Thus, it is more useful to interpret the odds ratios, which estimate the odds that a candidate would retest given his/her standing on the predictor variables. Odds ratios of 1.00 indicate no relationship between the predictor and the outcome, whereas values greater than 1.00 indicate a positive relationship (higher odds of retesting) and values less than 1.00 indicate a negative relationship (lower odds of retesting). As shown in Table 1, controlling for other variables in the model, the logit-based coefficient for responsibility ( $B = .00, p > .05$ ) was not statistically significant, indicating a lack of support for Hypothesis 1. The coefficient for



external/internal status ( $B = 1.52, p < .001$ ) was statistically significant, and interpreting the associated odds ratio reveals that internal candidates were 4.55 times more likely to retest than those who were external to the organization. Hypothesis 2 was supported<sup>4</sup>.

*Practice Effects: Descriptive Statistics*

Table 2 contains descriptive statistics for personality and cognitive ability dimensions at Time 1 and Time 2 for the retest groups as well as comparison data for the other subgroups that did not retest (cf. Lievens et al., 2005). The table also contains an effect size estimate, Cohen's  $d$ , to index the magnitude of observed score changes from Time 1 to Time 2 for the retest samples. Positive values of  $d$  indicate that scores increased from Time 1 to Time 2. Cohen (1992) suggested that values of  $d$  that are near .20 could be considered "small", those near .50 "medium", and those near .80 "large."

*Practice Effects: Hypothesis Tests*

Hypothesis 3 predicted that failing candidates would exhibit change on the personality dimension subscales across test administrations, and Hypothesis 4 predicted that practice effects would be larger for failing versus passing candidates. These hypotheses were tested by conducting a repeated measures mixed design multivariate analysis of variance (MANOVA) on the eight personality dimensions. The within-subjects factor (time) consisted of two levels to indicate the two testing occasions (Time 1 and Time 2), and the between-subjects factor (pass/fail status) consisted of two levels to indicate whether the candidate passed or failed the initial test. Main effects of time  $F(8, 348) = 2.37, p < .05, \eta^2 = .05$  and pass/fail status  $F(8, 348) = 7.71, p < .001, \eta^2 = .15$  were statistically significant, as was the interaction between time and

---

<sup>4</sup> Given that retesting requires less effort for internal candidates, the perseverance-based rationale for Hypothesis 1 may be restricted to external candidates. However, empirical tests were unresponsive, as the interaction term involving responsibility and external/internal status was not statistically significant ( $B = .00, p > .05$ ).

pass/fail status,  $F(8, 348) = 3.21, p < .01, \eta^2 = .07$ . Figure 2 illustrates the average personality test scores across testing occasions for failing and passing candidates. Paired sample t-tests were conducted to determine the patterns of dimension-level change for each group. In support of Hypothesis 3, failing candidates exhibited significant change on six of eight personality subscales, including cautiousness ( $t = 4.25, p < .001$ ), ascendancy ( $t = -8.87, p < .001$ ), sociability ( $t = -7.26, p < .001$ ), original thinking ( $t = -6.76, p < .001$ ), personal relations ( $t = -4.50, p < .01$ ), and vigor ( $t = -5.89, p < .001$ ). No change was observed for responsibility ( $t = .74, p > .05$ ) or emotional stability subscales ( $t = 1.61, p > .05$ ). Paired sample t-tests conducted with passing applicants revealed no change on any of the eight personality test subscales: ascendancy,  $t = .45$ ; responsibility,  $t = -.25$ ; emotional stability,  $t = .62$ ; sociability,  $t = .41$ ; cautiousness,  $t = -.83$ ; original thinking,  $t = .58$ ; personal relations,  $t = .04$ ; vigor,  $t = .14$ , all  $p > .05$ . The average effect size estimate for failing candidates (average  $d = .23$ ) was larger than that for passing candidates (average  $d = -.02$ ; see Table 2). For some of the personality dimensions (ascendancy, sociability, original thinking), effect size estimates were .50 to .60 higher for failing applicants when compared to passing applicants. Finally, as shown in Table 3, failing candidates produced markedly lower test-retest correlations. The average of the test-retest correlations ( $r_{12}$ ) across the eight dimensions was .35 for failing candidates ( $Min = .21, Max = .47$ ), and was .50 for passing candidates ( $Min = .36, Max = .64$ ). This result is particularly striking given that the average retest interval for passing candidates (20.6 months) was actually *longer* than that for failing candidates (11.8 months;  $t = 8.01, p < .01$ ). Recall that although previous research has indicated that personality stability declines with time, results of this study show that passing candidates actually had more stable profiles despite their longer average retest interval. In summary, when compared with passing candidates, failing candidates produced statistically significant score

changes for more personality dimensions, yielded larger practice effects, and exhibited less stability in their personality profiles over time. Overall, Hypothesis 4 was supported.

Hypothesis 5 predicted that among failing candidates, initial scores on any given personality subscale would be negatively related to the magnitude of practice effects. The regressed change approach advocated by Cohen et al. (2003) was used to test this hypothesis. Time 2 personality test scores were regressed upon the set of control variables (time lag, race, sex, external/internal status) and the Time 1 personality test score for the 265 failing candidates with complete data. As shown in Table 4, the coefficient for the Time 1 scores was positive and statistically significant for all eight personality dimensions, indicating that initial scores and retest scores were positively related (all else equal). To interpret the effects in terms of score *change* on any given scale, Time 1 scores that were one standard deviation above and below the mean were calculated for each personality subscale and entered into the regression equation to determine predicted Time 2 scores. Time 1 scores were then subtracted from the predicted Time 2 scores to estimate the magnitude of change for each dimension. Figure 3 illustrates the estimated score changes across the eight personality dimensions for high and low Time 1 scores. As shown in the figure, in all cases, score gains at Time 2 were larger when the Time 1 score was low rather than high. In particular, the predicted average change across dimensions for high Time 1 scores was -1.74 points (*Min* = -.85, *Max* = -2.80), whereas for low Time 1 scores it was +3.91 points (*Min* = 3.31, *Max* = 4.80). In standard deviation (*SD*) terms, when the Time 1 score was high, the average change was -.42 *SD* (*Min* = -.22, *Max* = -.61). When it was low, the average change was .98 *SD* (*Min* = .72, *Max* = 1.41). Overall, these findings indicate support for Hypothesis 5, as failing candidates showed larger increases on the subscales for which they performed relatively poorly on the initial attempt.

## Discussion

The purpose of this study was to contribute to the selection literature by examining predictors of retesting behavior and personality test practice effects. Analysis of nearly four years of selection data yields several insights into the profile of retest candidates as well as the nature and magnitude of practice effects in operational selection settings. The implications of these findings for research and practice are described below.

### *Research Implications*

Although previous researchers have suggested that persistence in selection settings may represent legitimate motivation among candidates, responsibility did not predict retest propensity, as persevering and determined candidates were not any more likely to retest. One possible explanation is that highly responsible candidates were selected by the organization after initial testing, leaving mostly low scorers as potential retest candidates. Although the responsibility dimension was not used in the weighting of candidates' overall composite scores, responsibility was positively correlated with three dimensions that did affect the candidate's pass/fail status, which raises potential indirect range restriction concerns that could lead to a Type II error. It should be noted, however, that the amount of truncation on the responsibility scale was minimal (e.g., more than half of the sample scored 30 or higher out of 36), and unstandardized regression coefficients are relatively robust to range restriction (Raju, Steinhaus, Edwards, & DeLessio, 1991; Sackett & Yang, 2000), suggesting that any effects could be detected if they existed.

Although the persistence-based prediction was not supported, results showed that internal candidates were nearly five times more likely to repeat the selection process than external candidates. These findings support conceptual arguments grounded in spatial separation and

feedback availability suggesting that internal candidates have lower search costs and perhaps greater self-efficacy based on feedback availability. Assuming this pattern is generalizable, organizations that adopt both internal and external candidate sourcing strategies may find a larger portion of unsuccessful internal applicants in their applicant pools over time (all else equal). Additional work is needed to tease apart which underlying processes are responsible for the observed effects. It is unclear whether internal applicants repeated the process because they were encouraged by the developmental feedback they received or simply did so out of proximity and convenience.

Drawing from previous theory and research, true change and response distortion arguments were developed to explain why personality test practice effects occur in selection settings. Although true change in personality is possible, the restricted range of the retest interval in this study (and in nearly all studies conducted in applied selection settings) suggests that substantial change in such a limited period seems an unlikely explanation for the score gains observed here. The median retest interval was less than one year, which is considerably smaller than the decade-long intervals reported in previous research as being necessary to detect true personality change. Thus, although intensive feedback and developmental opportunities may account for a portion of the variance in practice effects, previous theory and research suggests that such effects would be minimal given the time interval studied.

Alternatively, personality test practice effects may reflect response distortion. Recognizing the contingencies between test performance and job offers, candidates may alter their response strategies on any given testing occasion in an attempt to better fit the desired profile. One of the novel features of this study is that both failing and passing candidates retested, providing two distinct performance motives upon retesting. Results confirmed that

failing candidates pursued alternative response strategies and generated practice effects ( $d$ ) as high as .40 to .60 for some dimensions, whereas passing candidates generally replicated their response profile (average  $d = -.02$ ). Very low test-retest correlations for failing applicants (.21 to .47) provide further evidence of response pattern instability that true change cannot readily explain. Yet despite the general applicability of response distortion arguments, it is unclear exactly what strategies candidates adopted on any given attempt. Future research should examine how, when, and why applicants select alternative response strategies when completing personality tests.

To place this study's findings in context, recall that three recent studies have examined personality test practice effects using a within-person design involving applicant retesting. Two studies found limited evidence of response distortion (Ellingson et al., 2007; Hogan et al., 2007), whereas the third reported more substantial change (Landers et al., 2009). Comparisons between these studies and the present investigation reveal important differences in sample and design characteristics that support the context-dependence of practice effects. For example, all four papers involve different personality inventories, leaving open the question of whether practice effects may be test-specific. Three studies hold the job and organization constant, and one (Ellingson et al., 2007) examined individuals across a range of occupations, organizations, and testing contexts. Taken together, direct comparisons of effect size estimates across studies must be undertaken with care, as different forces likely account for variability in the magnitude of score change in any given setting.

Finally, consistent with recent evidence (Landers et al., 2009), score changes were larger on those dimensions where failing candidates' initial scores were relatively low. Low scores were associated with practice effects that exceeded a full standard deviation for some

dimensions. Conversely, on those scales where initial scores were relatively high, predicted change was smaller and in the opposite direction. A portion of these effects is likely attributable to regression to the mean, as extreme scores at one time tend to be less so at another (Campbell & Kenny, 1999), but the asymmetrical change observed for low scorers supports the more substantive conclusion that candidates appear to target those dimensions upon which they initially obtained low scores.

### *Practical Implications*

The results of this study may be useful to staffing specialists as they design selection systems and draft retesting policies. Small to moderate personality test practice effects were found despite the general presumption that personality is a stable construct (Costa & McCrae, 1988). The practice effects found here present a challenge to organizations that use personality assessment as part of the staffing process. Coupled with fairly low test-retest correlations, these findings call into question the practice of using such measures at all (cf. Morgeson et al., 2007). For organizations that see value, the question then becomes how to manage the sizable pool of failing candidates who seek opportunities to retest.

Disallowing retesting entirely may have potential social, moral, and legal implications that argue against such a policy. Particularly when organizations base promotion decisions partly on personality test scores, a “no retesting” policy may raise substantial fairness concerns and would force unsuccessful candidates to choose between an internal career plateau and an external job search. Such a policy ignores the individual’s capacity to develop important job skills over time through targeted learning and development. A final argument against limited retesting concerns those organizations that struggle to fill open positions. Given that retest candidates can

make up a sizable portion of the overall applicant pool, prohibitive retesting policies would work against staffing goals.

In contrast, permissive retesting policies alleviate fairness concerns, recognize that qualifications may develop over time, and help broaden the applicant pool. However, for personality testing, these benefits are offset by practice effects and lower reliability, suggesting that retesting may decrease the usefulness of personality data over time (Kelley, Jacobs, & Farr, 1994). Organizations may address some of these competing concerns by using different forms of assessment (e.g., structured interviews, assessment centers). Several studies provide guidance for developing alternative methods of assessing personality characteristics (e.g., Mount, Barrick, & Strauss, 1994), although more research is needed given recent findings that structured interviews developed to assess personality constructs share little overlap with self-report measures of those same dimensions (Van Iddekinge, Raymark, Eidson, & Attenweiler, 2004).

Another practical matter concerns the use of forced-choice personality measures in selection contexts. Depending on their design, forced-choice measures can be termed either *fully-ipsative* or *partially-ipsative*. Fully-ipsative measures contain response interdependencies that force scale scores to sum to the same total. As such, fully-ipsative measures place predictable constraints on practice effects, as score increases on any given dimension would necessarily cause score decreases on another. Moreover, fully-ipsative measures are mostly appropriate for within-person comparisons, and thus have been discouraged in selection settings (Meade, 2004). In contrast, partially-ipsative measures (such as the GPPI, studied here) contain interdependent items, but scales are not forced to sum to the same total. Scale scores derived from partially-ipsative measures are highly related to normative measures (Heggstad et al., 2006), permit between-person comparisons, and allow candidates to improve on all dimensions. Overall,



staffing specialists must recognize how these alternative response formats influence the meaning of test scores and the magnitude of practice effects.

### *Limitations*

One study limitation is that data were collected from only one company and a single personality measure. Thus, particularly when comparing these results to past research, it is impossible to tell whether the practice effects found here are attributable to occupational differences, test differences, or other contextual factors that differ between studies. Another limitation is that although several mechanisms were identified that could explain retest propensity (e.g., spatial separation, feedback availability), the design did not allow direct tests of these possible explanations. Finally, because criterion data were unavailable, the effects of retesting on personality test validity could not be examined. Although it is generally argued that response distortion does not compromise personality test validity (Hough, Eaton, Dunnette, Kamp, & McCloy, 1990), whether validity coefficients remain unchanged in retesting contexts remains an open question.

### *Conclusion*

Practice effects research has proliferated in recent years, yet little work has addressed the propensity to retest or examined personality test practice effects in actual selection contexts. Overall, these findings reveal some of the characteristics of retest candidates, suggesting that internal job market characteristics and feedback delivery mechanisms enhance the likelihood of candidate persistence. This study also clarifies the conditions under which personality test practice effects will be found, illustrating larger practice effects for failing candidates and for those subscales where initial scores were relatively low. As research continues in this area,

organizations should pay close attention to the patterns and processes that unfold when applicants persist.

## References

- Alliger, G.M., & Dwight, S.A. (2000). A meta-analytic investigation of the susceptibility of integrity tests to faking and coaching. *Educational and Psychological Measurement, 60*, 59-72.
- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.
- American Management Association. (2001). *2001 AMA survey on workplace testing: Basic skills, job skills, psychological measurement*. New York: Author.
- Ardelt, M. (2000). Still stable after all these years? Personality stability theory revisited. *Social Psychology Quarterly, 63*, 392-405.
- Bass, B.M. (1957). Faking by sales applicants of a forced choice personality inventory. *Journal of Applied Psychology, 41*, 403-404.
- Baum, W.M. (1973). The correlation-based law of effect. *Journal of the Experimental Analysis of Behavior, 20*, 137-153.
- 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*, 317-335.
- Bourdeau, N.R. (2008, April). Applicant retesting policy: Key considerations and best practices. Panel discussion presented at the meeting of the Society for Industrial and Organizational Psychology, San Francisco, CA.
- Campbell, D.T., & Kenny, D.A. (1999). *A primer on regression artifacts*. New York: Guilford.
- Cohen, J. (1992). A power primer. *Psychological Bulletin, 112*, 155-159.

- Cohen, J., Cohen, P., West, S.G., & Aiken, L.S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences* (3rd ed.). Mahwah, NJ: Erlbaum.
- Costa, P.T., Jr., & McCrae, R.R. (1988). Personality in adulthood: A six-year longitudinal study of self-reports and spouse ratings on the NEO Personality Inventory. *Journal of Personality and Social Psychology, 54*, 853-863.
- Costa, P.T., Jr., McCrae, R.R., & Dye, D.A. (1991). Facet scales for agreeableness and conscientiousness: A revision of the NEO Personality Inventory. *Personality and Individual Differences, 12*, 887-898.
- Dilchert, S., Ones, D.S., Viswesvaran, C., & Deller, J. (2006). Response distortion in personality measurement: Born to deceive, yet capable of providing valid self-assessments? *Psychology Science, 48*, 209-225.
- 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*, 386-395.
- Gobillon, L., Selod, H., & Zenou, Y. (2007). The mechanisms of spatial mismatch. *Urban Studies, 44*, 2401-2427.
- Gordon, L.V. (1993). *Gordon Personal Profile-Inventory Manual*. San Antonio, TX: Harcourt Brace.
- Gordon, L.V., & Stapleton, E.S. (1956). Fakability of a forced-choice personality test under realistic high school employment conditions. *Journal of Applied Psychology, 40*, 258-262.

- Hausknecht, J.P., Halpert, J.A., Di Paolo, N.T., & Moriarty Gerrard, M.O. (2007). Retesting in selection: A meta-analysis of practice effects for tests of cognitive ability. *Journal of Applied Psychology, 92*, 373-385.
- Hausknecht, J.P., Trevor, C.O., & Farr, J.L. (2002). Retaking ability tests in a selection setting: Implications for practice effects, training performance, and turnover. *Journal of Applied Psychology, 87*, 243-254.
- Heggestad, E.D., Morrison, M., Reeve, C.L., & McCloy, R.A. (2006). Forced-choice assessments of personality for selection: Evaluating issues of normative assessment and faking resistance. *Journal of Applied Psychology, 91*, 9-24.
- Herrnstein, R.J. (1970). On the law of effect. *Journal of the Experimental Analysis of Behavior, 13*, 243-266.
- Hogan, J., Barrett, P., & Hogan, R. (2007). Personality measurement, faking, and employment selection. *Journal of Applied Psychology, 92*, 1270-1285.
- Holzer, H.J. (1991). The spatial mismatch hypothesis: What has the evidence shown? *Urban Studies, 28*, 105-122.
- Hough, L.M., Eaton, N.K., Dunnette, M.D., Kamp, J.D., & McCloy, R.A. (1990). Criterion-related validities of personality constructs and the effects of response distortion on those validities. *Journal of Applied Psychology, 75*, 581-595.
- Ihlanfeldt, K. (1997). Information on the spatial separation of job opportunities within metropolitan areas. *Journal of Urban Economics, 41*, 218-242.
- Kain, J. (1968). Housing segregation, negro employment, and metropolitan decentralization. *Quarterly Journal of Economics, 82*, 175-197.

- Kelley, P. L., Jacobs, R. R., & Farr, J. L. (1994). Effects of multiple administrations of the MMPI for employee screening. *Personnel Psychology, 47*, 575-591.
- Knowles, E.S., Coker, M.C., Scott, R.A., Cook, D.A., & Neville, J.W. (1996). Measurement-induced improvement in anxiety: Mean shifts with repeated assessment. *Journal of Personality and Social Psychology, 71*, 352-363.
- Landers, R. N., Sackett, P. R., & Tuzinski, K. A. (2009). Retesting after failure, coaching rumors, and warnings against faking in using personality measures for selection. Manuscript submitted for publication.
- Leary, M.R., & Kowalski, R.M. (1990). Impression management: A literature review and two-component model. *Psychological Bulletin, 107*, 34-47.
- Lievens, F., Buyse, T., & Sackett, P. R. (2005). Retesting effects in operational selection settings: Development and test of a framework. *Personnel Psychology, 58*, 981-1007.
- Lievens, F., Reeve, C.L., & Heggestad, E.D. (2007). An examination of psychometric bias due to retesting on cognitive ability tests in selection settings. *Journal of Applied Psychology, 92*, 1672-1682.
- McFarland, L.A., & Ryan, A.M. (2000). Variance in faking across noncognitive measures. *Journal of Applied Psychology, 85*, 812-821.
- Meade, A.W. (2004). Psychometric problems and issues involved with creating and using ipsative measures for selection. *Journal of Occupational and Organizational Psychology, 77*, 531-552.
- 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*, 683-729.

- Mount, M.K., Barrick, M.R., & Strauss, J.P. (1994). Validity of observer ratings of the Big Five personality factors. *Journal of Applied Psychology, 79*, 272-280.
- Raju, N.S., Steinhaus, S.D., Edwards, J.E., & DeLessio, J. (1991). A logistic regression model for personnel selection. *Applied Psychological Measurement, 15*, 139-152.
- Raymond, M. R., Neustel, S., & Anderson, D. (2007). Retest effects on identical and parallel forms in certification and licensure testing. *Personnel Psychology, 60*, 367-396.
- Rosse, J.G., Stecher, M.D., Miller, J.L., & Levin, R.A. (1998). The impact of response distortion on preemployment personality testing and hiring decisions. *Journal of Applied Psychology, 83*, 634-644.
- Sackett, P. R., Burris, L. R., & Ryan, A. M. (1989). Coaching and practice effects in personnel selection. In C. L. Cooper & I. T. Robertson (Eds.), *International review of industrial and organizational psychology* (pp. 145-183). New York: Wiley.
- Sackett, P. R., & Yang, H. (2000). Correction for range restriction: An expanded typology. *Journal of Applied Psychology, 85*, 112-118.
- Schwab, D.P., & Packard, G.L. (1973). Response distortion on the Gordon Personal Inventory and the Gordon Personal Profile in a selection context: Some implications for predicting employee tenure. *Journal of Applied Psychology, 58*, 372-374.
- Society for Industrial and Organizational Psychology. (2003). *Principles for the validation and use of personnel selection procedures* (4th ed.). College Park, MD: Author.
- Tuzinski, K.A., Laczko, R.M., & Sackett, P.R. (2005, April). *Impact of response distortion on retaking of cognitive and personality tests*. Paper presented at the meeting of the Society for Industrial and Organizational Psychology, Los Angeles, CA.

Van Iddekinge, C.H., Raymark, P.H., Eidson, C.E., Jr., & Attenweiler, W.J. (2004). What do structured interviews really measure? The construct validity of behavior description interviews. *Human Performance, 17*, 71-93.

Viswesvaran, C., & Ones, D.S. (1999). Meta-analyses of fakability estimates: Implications for personality measurement. *Educational and Psychological Measurement, 59*, 197-210.

Wang, F.H. (2003). Job proximity and accessibility for workers of various wage groups. *Urban Geography, 24*, 253-271.



Author Note

John P. Hausknecht, Department of Human Resource Studies, School of Industrial and Labor Relations, Cornell University.

An earlier version of this study was presented at the Society for Industrial and Organizational Psychology Annual Conference in San Francisco, April 2008. I would like to thank Bradford S. Bell, Michael J. Howard, Paul R. Sackett, and Charlie O. Trevor for helpful suggestions and comments related to this work.

Correspondence concerning this article should be directed to John P. Hausknecht, Department of Human Resource Studies, 365 Ives Hall, School of Industrial and Labor Relations, Cornell University, Ithaca, NY 14853. E-mail: [jph42@cornell.edu](mailto:jph42@cornell.edu)

Table 1

*Multiple Logistic Regression Predicting Decision to Retest*


---

Variable	<i>B</i>	<i>SE</i>	Wald	df	exp( <i>B</i> )
Race	.15	.14	1.09	1	1.16
Sex	.23	.13	2.88	1	1.26
Responsibility	.00	.02	0.06	1	1.00
External/internal status	1.52***	.18	73.01	1	4.55

---

*Note.*  $N = 2,535$ . Analysis includes candidates who failed their first attempt. Whites, males, and internal candidates coded higher.

\*\*\*  $p < .001$

Table 2

*Personality and Cognitive Ability Test Scores for Different Applicant Groups*

	Total Sample (N = 15,338)														
	<u>Applicants Who Do Not Retest</u>				<u>Failing Applicants Who Retest</u>					<u>Passing Applicants Who Retest</u>					
	Passed		Did Not Pass		1 <sup>st</sup> Attempt		2 <sup>nd</sup> Attempt			1 <sup>st</sup> Attempt		2 <sup>nd</sup> Attempt			
	(N = 12,121)		(N = 2,860)		(N = 301)					(N = 56)					
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>d</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>d</i>	
<u>Personality Dimension</u>															
Ascendancy	25.15	2.85	22.58	3.34	22.37	3.46	24.64	3.62	.64	25.34	3.13	25.18	3.20	-.05	
Responsibility	30.55	2.95	28.83	4.41	28.92	4.70	28.72	4.05	-.04	29.45	3.67	29.57	3.78	.03	
Emotional Stability	27.07	3.26	26.18	3.88	26.27	4.07	25.86	3.71	-.10	27.07	3.07	26.75	3.81	-.09	
Sociability	23.26	3.48	22.00	3.52	21.68	3.57	23.54	4.12	.48	23.71	3.94	23.50	4.30	-.05	
Cautiousness	25.81	3.50	26.52	4.31	26.40	4.67	25.20	4.15	-.27	25.14	3.37	25.57	3.68	.12	
Original Thinking	30.03	3.14	26.63	3.73	26.62	3.91	28.37	3.65	.46	29.32	3.19	29.05	4.28	-.07	
Personal Relations	29.69	3.48	27.12	4.29	26.82	4.50	28.11	4.38	.29	30.07	3.25	30.05	4.22	.00	
Vigor	31.00	3.33	28.40	3.97	28.32	3.89	29.73	3.72	.37	30.79	2.95	30.71	3.54	-.02	
<u>Cognitive Ability Dimension</u>															
Quantitative Reasoning	19.86	3.24	13.56	4.29	14.41	4.34	16.43	4.46	.46	19.21	3.04	19.68	3.03	.15	
Abstract Reasoning	19.07	5.88	10.09	5.57	11.42	5.63	14.12	6.40	.45	16.95	6.00	18.04	5.33	.19	

Table 3

*Intercorrelations of Study Variables*

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
<u>Personality</u>																								
1. Ascendancy 1		-.69	-.69	.57	-.25	.35	-.09	.15	.64	-.27	-.24	.28	-.18	.30	-.09	.21	.00	-.06	-.03	-.07	-.38	.15	.06	.27
2. Responsibility 1	-.18		.59	-.66	.15	-.08	-.03	.19	-.44	.51	.44	-.42	.11	-.02	.06	.03	.10	.17	.06	.31	.40	-.19	-.06	-.18
3. Emotional Stability 1	-.07	.61		-.75	.22	-.13	-.01	.06	-.56	.28	.37	-.45	.07	-.07	-.09	-.12	.16	-.02	.07	.10	.18	-.02	.01	-.06
4. Sociability 1	.59	-.07	-.12		.01	.11	.30	-.12	.40	-.39	-.35	.55	-.05	.01	.17	-.03	-.20	-.06	-.09	-.19	-.21	.08	-.04	-.06
5. Cautiousness 1	-.22	.44	.38	-.14		-.19	.12	-.11	-.24	.30	.19	-.03	.41	.00	.14	-.20	-.03	-.23	-.04	-.19	-.06	-.17	-.09	-.11
6. Original Thinking 1	.24	.37	.19	.26	.06		-.21	.00	.27	.07	.06	.01	-.22	.61	-.16	.19	-.08	.27	.03	.29	-.10	.07	-.04	.07
7. Personal Relations 1	-.06	.41	.43	.14	.32	.24		-.33	-.02	-.06	.00	.36	.13	-.18	.52	-.12	-.27	-.18	-.08	-.15	.17	-.01	-.13	-.19
8. Vigor 1	.19	.25	.05	.16	.00	.25	-.10		.10	.02	.29	-.25	.06	.09	-.30	.36	.14	-.21	-.05	-.10	.07	-.05	-.07	.17
9. Ascendancy 2	.21	.16	.09	.15	.04	.15	-.02	.10		-.20	-.06	.35	-.20	.43	.20	.26	-.11	.08	-.09	-.01	-.19	.04	-.01	.35
10. Responsibility 2	-.08	.47	.35	-.16	.24	.13	.22	.10	-.23		.64	-.35	.41	.35	.25	.21	.01	.12	.04	.07	.22	-.21	-.03	-.10
11. Emotional Stability 2	-.14	.23	.37	-.21	.23	.03	.21	-.03	-.18	.56		-.55	.30	.39	.24	.11	-.08	-.15	-.11	-.14	.03	-.10	-.18	-.05
12. Sociability 2	.21	.07	-.07	.34	-.07	.15	-.02	.09	.57	-.25	-.46		-.14	.06	.25	.01	-.04	.02	-.18	-.04	.04	.09	-.16	.10
13. Cautiousness 2	-.15	.21	.20	-.08	.39	.04	.14	.02	-.20	.39	.42	-.22		-.17	.30	-.27	-.30	-.34	-.17	-.43	-.04	-.14	.05	-.05
14. Original Thinking 2	.06	.30	.17	-.07	.03	.30	.09	.02	.44	.20	.10	.23	-.18		.06	.31	.08	.19	-.14	.20	-.07	-.10	-.38	.13
15. Personal Relations 2	-.16	.24	.19	-.03	.22	.13	.37	-.04	.16	.22	.16	.22	.28	.26		-.33	-.32	-.06	-.20	-.13	.05	-.09	-.16	.01
16. Vigor 2	.09	.22	.07	.02	-.06	.14	-.06	.40	.33	.30	.16	.16	.00	.33	.06		.09	.12	.01	.12	.09	.00	-.16	.01
<u>Cognitive Ability</u>																								
17. Quantitative Reasoning 1	-.20	.11	.04	-.18	.25	-.15	-.08	-.21	.17	.12	.13	.07	.13	.15	.14	.03		.34	.53	.36	.20	-.11	.00	.11
18. Abstract Reasoning 1	-.26	.12	.09	-.22	.08	-.13	-.11	-.21	.06	.21	.16	.08	.11	.19	.17	.04	.29		.32	.72	.34	-.04	.09	-.01
19. Quantitative Reasoning 2	-.13	.15	.09	-.15	.15	-.05	-.02	-.13	.22	.16	.13	.10	.10	.26	.17	.12	.62	.35		.32	.19	.11	.19	-.11
20. Abstract Reasoning 2	-.11	.11	.14	-.13	.05	-.04	-.06	-.12	.14	.18	.22	.05	.05	.19	.11	.04	.33	.62	.43		.30	.07	.05	-.09
<u>Other</u>																								
21. Race	-.18	.28	.18	-.03	.14	-.08	.14	-.03	.08	.21	.09	.07	.04	.16	.24	.12	.28	.26	.25	.15		-.04	.13	-.14
22. Sex	.04	-.07	.01	-.01	-.09	.07	.00	-.05	.04	-.11	.05	-.05	.06	.03	-.02	-.09	-.04	-.05	-.05	.00	-.01		.03	-.12
23. External/internal status	.05	-.04	-.04	.03	.01	-.02	.00	-.06	.07	-.09	-.13	.02	-.19	.05	-.04	-.03	.08	.05	.10	.09	.04	-.08		.01
24. Time lag	.10	-.11	-.06	-.03	-.11	-.01	-.02	-.01	-.10	-.10	-.02	-.01	-.05	-.07	-.04	-.06	-.18	-.06	-.21	-.11	-.06	.02	-.17	

*Note.* Failing candidates listed below the diagonal ( $N = 301$ ), where correlations with an absolute value greater than .11 are statistically significant at  $p < .05$ .  
 Passing candidates listed above the diagonal ( $N = 56$ ), where correlations with an absolute value greater than .25 are statistically significant at  $p < .05$ .

Table 4

*Multiple Regressions Predicting Failing Candidates' Retest Scores*

Variable	Personality Dimension (Time 2)							
	Ascendancy	Responsibility	Emotional Stability	Sociability	Cautiousness	Original Thinking	Personal Relations	Vigor
Constant	20.64*** (1.62)	20.45*** (1.71)	19.53*** (1.61)	14.71*** (1.65)	17.84*** (1.64)	20.79*** (1.68)	19.40*** (1.69)	19.50*** (1.75)
Time lag (months)	-.04 (.03)	-.03 (.03)	-.01 (.03)	.02 (.03)	-.01 (.03)	-.02 (.03)	.01 (.03)	-.02 (.03)
Race	.58 (.44)	.62 (.46)	.17 (.43)	.50 (.46)	-.08 (.47)	1.28** (.44)	1.67** (.49)	.98* (.43)
Sex	.30 (.44)	-.67 (.44)	.20 (.42)	-.23 (.46)	.65 (.47)	.08 (.44)	-.12 (.49)	-.59 (.43)
External/internal status	.48 (.62)	-1.04 (.63)	-1.16 (.60)	.17 (.65)	-2.10** (.66)	.47 (.62)	-.59 (.69)	-.17 (.60)
Personality (Time 1)	.17* (.07)	.33** (.05)	.28*** (.05)	.38*** (.07)	.34*** (.05)	.26*** (.06)	.31*** (.05)	.37*** (.05)

*Note.*  $N = 265$ . Personality (Time 1) reflects the same dimension as the dependent variable in each model. Standard errors listed in parentheses.

Internal candidates, Whites, and males coded higher. Model  $R^2$  values range from .04 to .20 ( $M = .14$ ).

\* $p < .05$ , \*\* $p < .01$ , \*\*\* $p < .001$ .

Figure 1

Model of Candidate Flow through the Selection Process

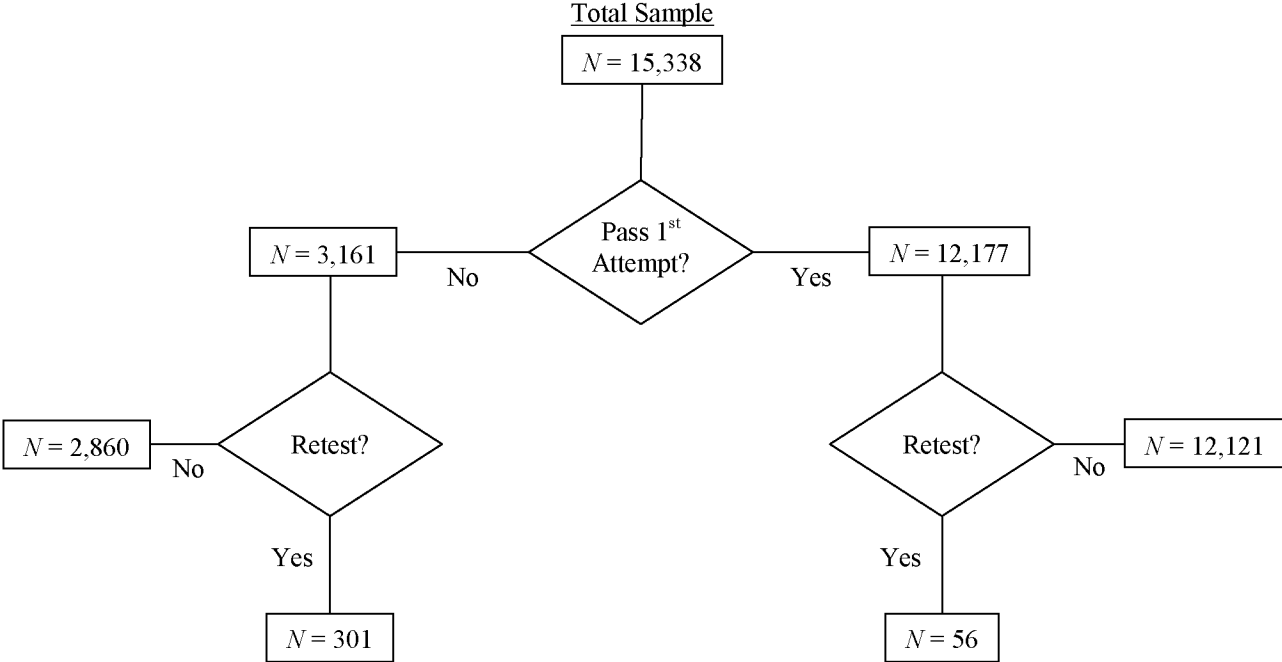


Figure 2

*Personality Test Practice Effects for Failing and Passing Candidates*

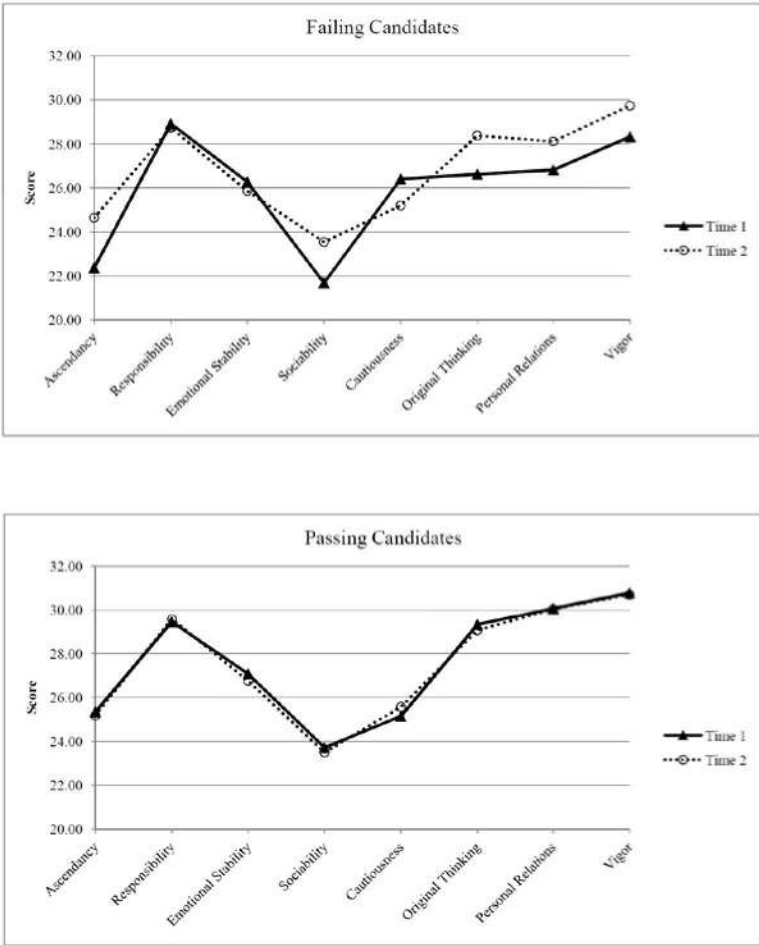
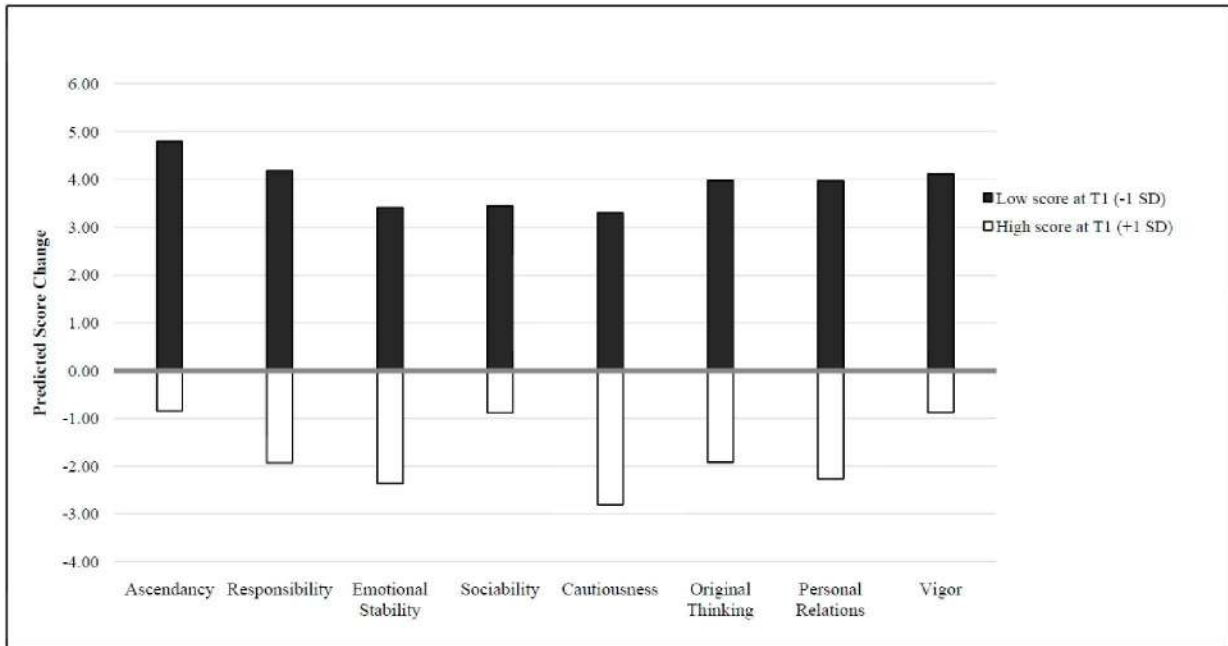


Figure 3

*Relationships Between Initial Scores and Personality Test Practice Effects*

*Note.* High and low scores defined as  $\pm 1$  SD from the mean, respectively, and are based on failing candidates' initial scores. Predicted score change reflects the regression-based estimate of the extent to which the retest score deviates from the initial score on each dimension.