

# Retaking Ability Tests in a Selection Setting: Implications for Practice Effects, Training Performance, and Turnover

John P. Hausknecht  
Pennsylvania State University

Charlie O. Trevor  
University of Wisconsin—Madison

James L. Farr  
Pennsylvania State University

This field study investigated the effect of retaking identical selection tests on subsequent test scores of 4,726 candidates for law enforcement positions. For both cognitive ability and oral communication ability selection tests, candidates produced significant score increases between the 1st and 2nd and the 2nd and 3rd test administrations. Furthermore, the repeat testing relationships with posthire training performance and turnover were examined in a sample of 1,515 candidates eventually selected into the organization. As predicted from persistence and continuance commitment rationales, the number of tests necessary to gain entry into the organization was positively associated with training performance and negatively associated with turnover probability.

The area of employment-related test preparation is undergoing considerable growth. Moreover, many employers, particularly those in the public sector, allow job applicants to retake employee selection tests. Thus, it is important to gauge, in the employment realm, the effectiveness and implications of what is commonly referred to as test coaching and practice. Although some research has addressed whether coaching and practice generally yield improvements in ability test scores (Kulik, Bangert-Drowns, & Kulik, 1984; Kulik, Kulik, & Bangert, 1984) and although more recent studies have examined test preparation courses in such contexts as college admission decisions (e.g., Briggs, 2001), no research exists on ability test practice in an employee selection context. Consequently, the important issue of whether retaking such selection tests might change test scores and subsequent hiring decisions has been, as of yet, unexplored. In addition, retaking these tests may have implications that extend beyond employee selection. In this study, we examined cognitive ability and oral communication ability selection tests to understand how scores change with repeat administrations. Furthermore, we investigated whether retaking such tests is related to posthire training performance and employee turnover.

Thus, the purpose of this article is threefold. First, we examined the relationship between test practice and test scores. Drawing

from research in educational psychology and other fields, we attempted to extend to employment settings the prior findings concerning practice effects on cognitive ability test scores. We also investigated a potential practice effect on oral communication ability test scores for the same sample. Second, we investigated the relationship between the number of attempts to gain entry into an organization and posthire training performance. On the basis of the applicants' persistence in the face of rejection, we drew on a motivation-based rationale for our hypothesis. Third, we examined the previously unstudied relationship between test repetition and employee turnover. In doing so, we offer an escalation-of-commitment rationale for hypothesizing that the persistence associated with repeated attempts to gain entry into the organization is associated with turnover probability.

## Background and Hypotheses

### *Practice and Testing*

The effect of coaching and practice (a subset of the broader coaching construct) on test scores has been investigated primarily in the educational literature. With an increased number of commercial coaching programs available for standardized tests, researchers have been interested in the impact of such programs on test scores. Kulik, Bangert-Drowns, and Kulik (1984) conducted a meta-analysis of coaching programs for aptitude tests. They found that coaching raised scores by an average of 0.15 *SD* and 0.43 *SD* in 14 studies on the Scholastic Assessment Test and 24 studies on other aptitude and intelligence tests, respectively. Powers and Rock (1999) more recently found that coaching raised scores on the revised version of the Scholastic Assessment Test by less than 0.10 *SD* and 0.20 *SD* on the verbal and math portions, respectively. In a recent large-scale study by Briggs (2001) of students' scores on the Scholastic Assessment Test and the Amer-

---

John P. Hausknecht and James L. Farr, Department of Psychology, Pennsylvania State University; Charlie O. Trevor, Department of Management and Human Resources, University of Wisconsin—Madison.

This article is based on John P. Hausknecht's minor area project completed at Pennsylvania State University under the supervision of Charlie O. Trevor.

Correspondence concerning this article should be addressed to John P. Hausknecht, Department of Psychology, Pennsylvania State University, 429 Bruce V. Moore Building, University Park, Pennsylvania 16802. E-mail: jhaus@psu.edu

ican College Test, similar findings emerged, demonstrating small but significant increases in subsequent test scores following participation in test preparation activities. As test preparation programs move into employment settings, published research on the effectiveness of such coaching interventions is beginning to emerge. For example, Ryan, Ployhart, Greguras, and Schmit (1998) examined both the effectiveness of test preparation programs in a selection setting and the factors related to self-selection into such programs. They found that voluntary attendees of test preparation programs scored no better on an ability test than did nonattendees.

Although the coaching studies are informative, test practice alone is the issue of interest in the present study. Kulik, Kulik, and Bangert (1984) summarized early research on practice effects using meta-analysis. The authors drew almost exclusively on studies with student populations to examine practice effects on aptitude and achievement test scores. They reported that test score increases in the second administration were larger when identical tests were used (0.42 *SD*) than when parallel forms of the tests were used (0.23 *SD*). The authors also found a significant positive relationship between test takers' ability and size of the practice effect, as effect sizes over two identical tests were 0.80 *SD*, 0.40 *SD*, and 0.17 *SD* for subjects of high, middle, and low ability, respectively. Finally, multiple test repetitions resulted in larger practice effects, with a 0.42-*SD* mean increase from the first to the second administration of an identical test (19 studies), a 0.70-*SD* improvement from the first to the third administration (6 studies), and a 0.96-*SD* increase from the first to the fourth administration (5 studies). In the most recent research on practice effects, psychologists have examined intelligence testing from a clinical perspective. Studies of the Wechsler Adult Intelligence Scale—Revised and numerous other neuropsychological measures indicate that improved scores tend to occur with repeat administrations of most measures (Rapport, Axelrod, et al., 1997; Rapport, Brines, Axelrod, & Theisen, 1997; Watson, Pasteur, Healy, & Hughes, 1994).

Recognizing the potential importance of practice effects for employee selection, Sackett, Burris, and Ryan (1989) reviewed a small number of test manuals, technical reports, and published studies that used ability and achievement tests designed for employment settings. These authors also found support for the existence of practice effects and replicated the earlier meta-analytic findings of larger practice effects for identical tests than for parallel tests (Kulik, Kulik, & Bangert, 1984). Sackett et al. (1989) summarized the extant literature at the time by stating, "Practice effects are found fairly consistently for a variety of test types and for a wide range of retest intervals, thus indicating that the practice issue should not be taken lightly by those organizations using ability tests" (p. 159). It is important to emphasize, however, that none of the studies reviewed by Sackett et al., and no published studies that we are aware of since then, were able to test for practice effects on ability tests in an authentic selection context. This point is critical from an ecological validity standpoint. For example, outside of the selection setting, test takers' motivation to improve on their initial test scores may well not be representative of the motivation of an actual job applicant attempting to improve scores enough to gain entry into the organization (Sackett et al., 1989).

Although no studies have examined practice effects on ability tests in an actual selection setting, Kelley, Jacobs, and Farr (1994)

investigated practice effects in a selection context with multiple screenings of the Minnesota Multiphasic Personality Inventory (MMPI). The MMPI produces personality profiles that have been used to screen out psychologically unfit applicants. Hence, its study in the context of practice effects is quite distinct from our emphasis on ability test score improvement, although the study's implications for test practice in selection contexts are quite relevant. Kelley et al. found an increasing normality of MMPI score profiles across administrations at a nuclear power plant. The authors concluded that the instrument's effectiveness at identifying, and thus screening out, psychologically unfit applicants and employees diminishes across repeat administrations. From an ability testing perspective, this study highlights the potentially important implications of practice effects, in that systematic score changes across repeated administrations ultimately can result in a qualitative change in the makeup of the workforce.

In sum, existing literature from varied settings suggests that test score increases occur with repeat administrations of the same or similar ability tests. Although no research has examined this dynamic under actual selection conditions, we expected to find a similar result when using cognitive ability tests and oral communication ability tests as employee selection devices.

*Hypothesis 1a:* Test scores of job candidates who retake the same cognitive ability test will increase across administrations.

*Hypothesis 1b:* Test scores of job candidates who retake the same oral communication ability test will increase across administrations.

### *Practice and Posthire Training Performance*

As in the preceding hypotheses, most research on practice effects focuses on ability test scores as the outcome. But because these tests are used for employee selection and because test scores presumably predict subsequent performance, another issue of considerable importance is whether practice on a selection hurdle might be related to posthire performance. Thus, we examined whether the number of attempts needed to pass a selection test is related to subsequent performance in posthire training.

The act of retaking a selection test after an applicant has failed to score well enough to gain entry into the organization on the preceding attempt exemplifies persistent behavior on the part of the applicant. To the extent that such persistence is indicative of motivation, it is possible that, having endured multiple tests to finally clear the hurdle and secure a position, repeat testers will be more motivated to perform well once they are hired.

Many studies have investigated ability and motivation as two basic determinants of performance. Results from several experiments by Kanfer and Ackerman (1989) support the notion that "motivation and cognitive abilities represent two basic determinants of learning and work performance" (p. 657). When considering motivational factors, the authors distinguished between distal and proximal motivational processes. Distal motivational processes are those involving the choice to engage some amount of resources toward the attainment of a goal. Proximal motivational processes, in contrast, are those that determine the level of effort expended during task engagement, including self-regulatory processes such as self-monitoring and self-evaluation. The studies by Kanfer and Ackerman primarily investigated proximal motivational processes. In this study, we investigated aspects of distal

motivational processes that also may indicate subsequent proximal motivation. That is, individuals had to determine to expend effort in the form of repeat testing well in advance of proximal task-engagement processes.

Four recent studies provide additional evidence as to the importance of ability and motivation in predicting job performance. Wright, Kacmar, McMahan, and Deleeuw (1995) provided support for the notion that cognitive ability and personality (said to reflect motivation) jointly determine performance. With personality tests used to assess achievement need, 12% of the variance in performance was explained by cognitive ability, achievement need, and the interaction between the two variables. In a study examining the effects of cognitive ability and motivation on career success, O'Reilly and Chatman (1994) found a significant interaction between Graduate Management Admission Test scores and conscientiousness variables when predicting early managerial career success. However, they did not find significant main effects for either predictor. The authors suggested that high levels of both cognitive ability and motivation are important for early managerial success. In contrast to these two studies, Mount, Barrick, and Strauss (1999) found no evidence of conscientiousness, as a proxy for trait motivation, interacting with cognitive ability in the prediction of job performance. However, the authors did find main effects for each variable across three independent samples. Finally, McCloy, Campbell, and Cudeck (1994) proposed that performance is determined by declarative knowledge (i.e., knowing facts, rules, and principles), procedural knowledge and skills (i.e., knowing how to perform a task), and motivation (i.e., choice to initiate, expend, and persist in expenditure of effort). Structural equation analyses of data from military personnel provided strong support for the three-factor model. The authors suggested that, to perform effectively, a person must have the requisite knowledge, master the required skills, and choose to exert some level of effort on job tasks for some time.

Taking multiple selection tests in repeated attempts to gain organizational entry appears to represent persistent, motivated behavior. Given the conceptual and empirical grounds for believing that motivation influences performance, we hypothesized the following:

*Hypothesis 2a:* With ability controlled for, the number of cognitive ability tests taken will be positively related to posthire training performance.

*Hypothesis 2b:* With ability controlled for, the number of oral communication ability tests taken will be positively related to posthire training performance.

### *Practice and Turnover*

Our third application of test practice to an employment setting involves turnover. A central feature of most employee turnover models is that job dissatisfaction and lack of commitment stimulate the turnover process. Extant literature provides evidence for a job investments–commitment–turnover relationship (Griffeth & Hom, 1995). That is, employees often make job investments that bolster their commitment to the organization, which makes them less likely to leave.

Becker (1960) proposed that commitment increases as employees make side bets, or investments that they would lose if they left

the organization. Building on this seminal work, commitment researchers have conceived of calculative commitment as the material investments that employees have in an organization that bind them to stay with the firm. In an organizational commitment meta-analysis, Mathieu and Zajac (1990) found moderate relationships between continuance commitment and turnover intentions as well as between continuance commitment and turnover. In a similar manner, Tett and Meyer (1993) found meta-analytic support for the relationship between organizational commitment and turnover. Griffeth and Hom (1995) contended that job investments underpin the commitment–turnover relationship. In other words, job investments such as those associated with pension plans or firm-specific training deter individuals from leaving the organization for fear of losing the investments.

The time, effort, and opportunity cost associated with taking multiple selection tests is one way in which individuals may make what they perceive to be investments in the organization. Although these elements may in fact be sunk costs rather than investments, the critical issue is whether individuals respond to them as if they were investments. Indeed, there is considerable evidence for the sunk-cost effect (e.g., Arkes & Blumer, 1985; Staw & Hoang, 1995), whereby people exhibit an arguably irrational commitment to courses of action as a result of irreversible expenditures. Thus, regardless of whether the “investments” associated with repeated test taking are in fact sunk costs, it seems reasonable that they should enhance continuance commitment, which should ultimately make employees less inclined to leave.

*Hypothesis 3a:* The number of cognitive ability tests taken will be negatively related to turnover.

*Hypothesis 3b:* The number of oral communication ability tests taken will be negatively related to turnover.

## **Method**

### *Setting, Procedure, and Participants*

This field study involved participants who were job candidates for an entry-level position in a law enforcement agency located in the eastern United States. Figure 1 illustrates the selection and applicant flow process. After a general screening of employment applications, candidates were invited to participate in the first hurdle of the selection process, which was a written cognitive ability examination. We obtained test scores for the four cognitive ability test administrations offered from 1990 to 1993. Candidates who failed to reach a cutoff score (approximately 70%) on the cognitive ability test were excluded from consideration for that year. However, the excluded candidates could reenter the selection process in subsequent years by retaking the cognitive ability test (test administrations were approximately 1 year apart). Those candidates reaching the cutoff score, regardless of prior test performance, were invited to participate in the next step in the selection process, which was an oral communication ability test. Consequently, we obtained scores on the yearly oral communication test administrations as well. Candidates were selected in a top-down manner on the basis of the combined cognitive and oral communication ability test scores, which were evenly weighted. Selected candidates were then given a physical examination and an extensive background check before beginning work in the training academy. The actual number of selected candidates each year was a function of organizational needs. Those not selected at this point were allowed to begin the process anew in subsequent years. Candidates who were selected and performed satisfactorily in the law enforcement training academy began working in the field after graduation. The organization provides paid training, which lasts

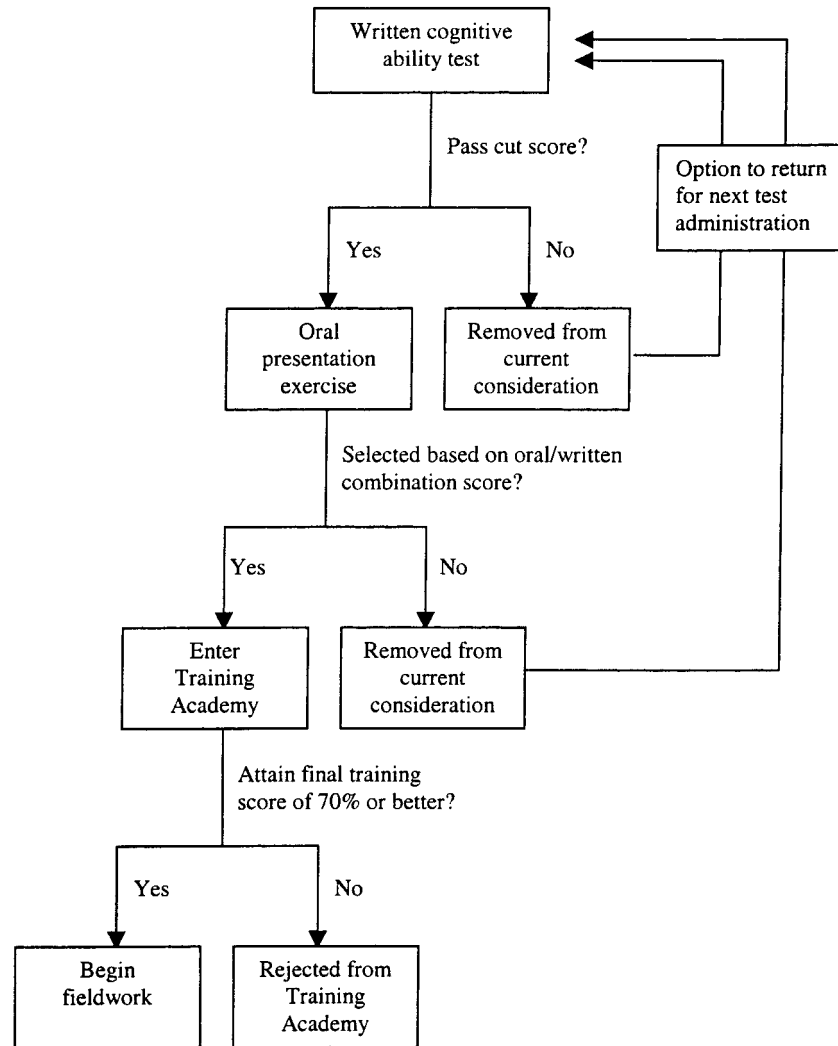


Figure 1. General model of applicant flow process.

approximately 6 months, and approximately 98% of candidates graduate successfully and begin fieldwork.

To begin to assess Hypothesis 1, which proposed test score increases across administrations, we first gathered test data on those individuals who took more than one cognitive ability test over the course of the four test administrations. Thus, we acquired data on the 4,726 candidates who failed to gain organizational entry in 1 year and returned at least once to attempt to gain entry in a subsequent year. We also gathered oral communication test data on the 375 candidates who took more than one such test, allowing us to also examine potential practice effects on this type of test. Substantially fewer candidates retook the oral communication test than retook the cognitive ability test as a result of the multiple-hurdle approach used.

Hypotheses 2 and 3 explored test practice associations with posthire training performance and employee turnover. To test these hypotheses, we used the number of cognitive ability tests taken and the number of oral communication ability tests taken as indicators of persistence, and thus as motivation proxies. Data were available for the 1,515 participants who had gained entry into the training academy on written tests administered in 1990, 1991, 1992, or 1993.

Several reasons led us to select the study window described above. First, both the cognitive and oral communication ability tests were identical

across years, which allowed us to test for practice effects over this time span (i.e., within-person test score change is best assessed with identical tests because they form a valid basis for comparison). We also had written test data for 1988, 1989, and 1994, but these tests differed from the 1990–1993 tests. The former tests assessed different dimensions and were created by two different consulting firms, whereas the latter tests were identical across years and were developed by the same firm. Standardizing scores within each year to compare within-person change across different tests was not an option because our applicant pools were not homogeneous across years and because the two tests were not parallel. The second factor that led us to choose this time frame was that examining whether scores improved from Test 1 to Test 2 necessitated being able to accurately identify Test 1. By beginning in 1990, when we had data for 1988 and 1989, we were able to better verify which testers, by virtue of being absent from the testing rolls in 1988 and 1989, were in fact taking tests for the first time. The third reason for our study window choice involved the testing of Hypotheses 2 and 3. Controlling for ability when investigating relationships between motivation proxies and outcomes that are correlated with ability (e.g., performance, turnover) is necessary for proper model specification. An acceptable ability measure requires, of course, that the construct be measured in the same way for all observations. By using these



study windows, we constrained ourselves to the ability scores from identical tests.

### Measures

**Cognitive ability.** As we stated previously, the initial hurdle in the selection process was a written cognitive ability test. The exact same test, which was developed for the organization by an independent consulting firm, was used across the four test administrations. Two multiple-choice verbal ability components served as the measure of cognitive ability. The first component was a vocabulary measure that asked candidates to identify the correct meaning of a number of words. The second component was a verbal comprehension measure that assessed participants' ability to identify important elements of written passages. Scores on the vocabulary and verbal comprehension components were evenly weighted and combined to provide the measure of cognitive ability for each participant. Test-retest correlations for consecutive years on the cognitive ability tests were .77 (1990–1991), .73 (1991–1992), and .74 (1992–1993), suggesting acceptable test reliability.

We initially used each subject's first cognitive ability test score as the cognitive ability control variable. Because initial test scores for repeat testers are free of potential practice effects, this ensured comparability on this measure regardless of the number of tests taken. Moreover, given that cognitive ability is generally characterized as a stable characteristic (e.g., Gottfredson, 1986), we assumed that any change in cognitive ability scores would be attributable to factors other than actual change in cognitive ability. However, because the impetus for practice effects could not be known with any degree of certainty in our study, we also present results using the cognitive ability test score used to gain entry into the organization as the control. Should practice effects in fact result from actual cognitive ability changes, or from reduced test anxiety, the entry-gaining score is arguably the better cognitive ability control.

**Oral communication ability.** After passing the written cognitive ability test, candidates proceeded to the next step of the selection process, which was an oral communication ability examination. Candidates took part in a role-playing exercise that involved delivering a job-related oral presentation to an audience of three raters. Candidates were asked to prepare answers to several questions that were provided in advance and also were asked to answer several situational questions without preparation. Candidates were allotted 45 min to complete the oral communication ability exercise. Raters provided scores on several dimensions (e.g., decision making), which were averaged across raters and across dimensions to arrive at a final score for each candidate.

Test-retest correlations for consecutive years on the oral communication ability tests were .23 (1990–1991), .41 (1991–1992), and .36 (1992–1993). Given identical tests, two major sources of constrained test-retest correlations are subjectivity of scoring and temporal instability in the attribute that the test measures (Nunnally & Bernstein, 1994). It is likely that a substantial portion of the score variation between years resulted from the organization's frequent use of different raters across test administrations, a typical scenario for such scoring subjectivity (Nunnally & Bernstein, 1994). This probable unreliability notwithstanding, we also suspected that temporal instability, which yields low test-retest correlations that clearly are not driven by unreliability (Guion, 1998), contributed to the low oral communication ability year-to-year correlations. Furthermore, because practice effects on the oral communication ability examination might involve actual improvement in oral presentation skills, the entry-gaining oral communication ability test scores may be the more valid indicator of oral communication ability. However, as with cognitive ability scores, we could not offer substantial evidence as to the source of practice effects on the oral communication ability test and thus also conducted analyses with initial oral communication scores as the oral communication ability control.

**Number of cognitive ability tests taken.** We examined the record of each participant to determine the number of times the candidate took a cognitive ability examination. The mean number of cognitive ability tests

taken by candidates was 1.65, with 42% of those accepted into training requiring more than one examination. Because those candidates who failed to gain entry into the training academy on initial or subsequent testing were free to take the annual cognitive ability examination again, this measure represents persistence in attempts to enter the organization.

**Number of oral communication ability tests taken.** We examined the record of each participant to determine the number of times the candidate took the oral communication ability test. The mean number of oral communication ability tests taken by candidates was 1.19, with 16% of those accepted into training requiring more than one examination. As with cognitive ability tests, the number of oral communication ability tests taken represents persistence in attempts to enter the organization.

**Training performance.** Once selected into the training academy, candidates spent approximately 6 months in paid training. Performance was based on three components: knowledge of criminal law, knowledge of traffic law, and performance on a final examination. To assess mastery of criminal and traffic law, written examinations were administered quarterly throughout the training period to arrive at a criminal law and traffic law average. Although the criminal and traffic law components assessed conceptually distinct material, averages between these two components were highly correlated ( $r = .71$ ). A comprehensive final examination, covering material from both domains, was administered at the end of training. Final examination scores were highly correlated with a composite of the average criminal and traffic law scores produced during training ( $r = .60$ ). According to organizational procedures, these three scores (criminal, traffic, and final exam) were weighted 40%, 40%, and 20%, respectively, to arrive at the final training average. This final average was the criterion chosen to represent training performance ( $M = 81.69$ ,  $SD = 7.23$ ).

**Turnover.** The law enforcement agency maintained separation dates for those who left the organization. The turnover variable was coded to identify those who had separated from the organization from the date of entry until the end of May 1999. Approximately 5.7% of the sample left the organization during the period under study. This figure largely represented voluntary turnover, because individuals were rarely (and not easily) terminated from the organization.

**Additional control variables.** In addition to the cognitive and oral communication ability controls previously described, several other control variables were used in this study to rule out alternative explanations for variation in training performance and turnover. We included gender, ethnic background, and age when examining the effects of repeat testing on performance and turnover. For the performance analyses, we also included dummy variables for the year in which one began the applicant process by taking his or her first cognitive ability test. In contrast, for the prediction of turnover, we included dummy variables for year of entry into the training academy as controls for general external job market differences. In addition, tenure and training performance were controlled when the effects of repeat testing on turnover were examined.

## Results

### Practice Effects

Table 1 provides the intercorrelations among the variables included in this study. Hypothesis 1 predicted that participants would demonstrate score increases with repeated cognitive ability and oral communication ability tests. For various breakdowns of candidates' test-taking experience, Table 2 presents the mean raw scores for consecutive tests taken by candidates. Table 2 includes consecutive test scores for those individuals who took between two and four cognitive ability tests and between two and three oral communication ability tests (only 1 candidate took four oral communication ability tests). Each row of Table 2 provides data on potential practice effects, as paired-sample *t* tests were conducted on each of the listed pairs of consecutive test administrations.

Table 1  
*Intercorrelations Among Study Variables*

Variable	<i>M</i>	<i>SD</i>	1	2	3	4	5	6	7	8	9	10	11	12
1. Age	32.07	2.44	—											
2. Gender <sup>a</sup>	1.97	0.16	.02	—										
3. Ethnic background <sup>b</sup>	1.91	0.28	.01	.01	—									
4. Tenure	6.68	1.62	.20	.14	.17	—								
5. Cognitive ability <sup>c</sup>	82.75	7.40	-.17	.02	.22	-.02	—							
6. Cognitive ability <sup>d</sup>	83.96	7.04	-.17	.04	.24	-.02	.85	—						
7. Oral communication <sup>c</sup>	69.21	19.76	-.02	-.01	.13	.04	.26	.26	—					
8. Oral communication <sup>d</sup>	70.06	19.87	-.03	-.02	.14	.01	.25	.27	.95	—				
9. No. of written tests	1.65	0.91	.20	.00	.04	.01	-.24	-.03	.04	.09	—			
10. No. of oral tests	1.19	0.47	.19	-.02	.09	.06	-.07	.03	.02	.11	.64	—		
11. Training performance	81.69	7.23	-.02	.17	.19	.24	.27	.31	.16	.15	.02	.06	—	
12. Turnover <sup>e</sup>	.06	.23	.00	-.15	-.13	-.73	-.13	-.15	-.06	-.05	-.05	-.01	-.29	—

Note.  $N = 1,519$ – $1,525$ . All correlations  $>.08$  are significant at  $p < .01$ .

<sup>a</sup> Coded as 1 = female, 2 = male. <sup>b</sup> Coded as 1 = minority status, 2 = nonminority status. <sup>c</sup> Initial test scores. <sup>d</sup> Entry-gaining test scores. <sup>e</sup> Coded as 0 = did not turnover, 1 = turnover.

Consistent support was found for Hypothesis 1 when we examined score change from the first to the second test administration, because both cognitive and oral communication ability scores revealed statistically significant increases, regardless of the number of tests ultimately taken. Candidates similarly improved from the second to the third test administration, regardless of both number of tests eventually taken and type of test. We note that for oral communication ability scores on the second and third administrations, we chose to reject the null hypothesis of no difference at a significance level of  $p < .10$  because the small number of cases ( $n = 23$  and  $n = 24$ ) seriously limited statistical power, although this decision does, of course, increase the likelihood of a Type I error. For the 137 candidates who took four cognitive ability tests,

we found no significant difference between scores on Test 3 and Test 4, suggesting the possibility that consecutive test practice effects may erode or disappear with repeated administrations. Although Kulik, Kulik, and Bangert (1984) reported that cumulative effect sizes (i.e., between the first test and the final test) increased as a function of the number of tests, they did not look at consecutive test score differences beyond Tests 1 and 2. In sum, for consecutive test administrations, we found clear evidence of practice effects between the first and second tests and between the second and third tests, but we did not find support between Tests 3 and 4.

In Table 3, we present results for practice effects across non-consecutive tests (i.e., more than one “practice”). Here, we found

Table 2  
*Mean Score Differences and Effect Sizes Over Consecutive Tests*

No. of tests applicants took	<i>n</i>	Applicant's test administration	<i>M</i> ( <i>SD</i> )	Applicant's test administration	<i>M</i> ( <i>SD</i> )	<i>d</i>
Cognitive ability						
2 or more	4,726	1st	74.69 (7.75)	2nd	77.32 (7.47)**	.34
Exactly 2	3,789	1st	75.04 (7.75)	2nd	77.39 (7.54)**	.30
Exactly 3	800	1st	73.21 (7.76)	2nd	77.24 (7.19)**	.52
Exactly 4	137	1st	73.59 (6.81)	2nd	75.79 (7.23)**	.32
3 or more	937	2nd	77.03 (7.21)	3rd	78.72 (7.07)**	.24
Exactly 3	800	2nd	77.24 (7.19)	3rd	78.56 (7.08)**	.18
Exactly 4	137	2nd	75.79 (7.23)	3rd	79.66 (6.94)**	.54
Exactly 4	137	3rd	79.66 (6.94)	4th	79.89 (6.62)	.03
Oral communication ability						
2 or more	375	1st	61.12 (18.87)	2nd	63.85 (20.47)*	.14
Exactly 2	351	1st	61.67 (18.71)	2nd	63.99 (20.32)*	.12
Exactly 3	23	1st	53.47 (20.26)	2nd	62.20 (23.37)*	.43
3 or more	24	2nd	61.81 (22.94)	3rd	70.14 (20.94)†	.36
Exactly 3	23	2nd	62.20 (23.37)	3rd	71.26 (20.66)†	.39

Note. Significance tests are between the two test administrations listed within each row. The *d* values are effect sizes generated by  $(M_{\text{Later}} - M_{\text{Earlier}})/SD_{\text{Earlier}}$ , where *Later* and *Earlier* subscripts represent test administration chronology.

†  $p < .10$ . \*  $p < .05$ . \*\*  $p < .01$ .

Table 3  
Mean Score Differences and Effect Sizes Over Nonconsecutive Tests

No. of tests applicants took	<i>n</i>	Applicant's test administration	<i>M</i> ( <i>SD</i> )	Applicant's test administration	<i>M</i> ( <i>SD</i> )	<i>d</i>
Cognitive ability						
3 or more	937	1st	73.27 (7.62)	3rd	78.72 (7.07)**	.76
Exactly 3	800	1st	73.21 (7.76)	3rd	78.56 (7.08)**	.69
Exactly 4	137	1st	73.59 (6.81)	3rd	79.66 (6.94)**	.89
Exactly 4	137	1st	73.59 (6.81)	4th	79.89 (6.62)**	.93
Exactly 4	137	2nd	75.79 (7.23)	4th	79.89 (6.62)**	.57
Oral communication ability						
3 or more	24	1st	53.21 (19.85)	3rd	70.14 (20.94)**	.85
Exactly 3	23	1st	53.47 (20.26)	3rd	71.26 (20.66)**	.88

Note.  $N = 1,515$ . Significance tests are between the two test administrations listed within each row. The  $d$  values are effect sizes generated by  $(M_{\text{Later}} - M_{\text{Earlier}})/SD_{\text{Earlier}}$ , where *Later* and *Earlier* subscripts represent test administration chronology.

\*\*  $p < .01$ .

clear support for Hypothesis 1, because regardless of test type and eventual test-taking experience, test scores always showed statistically significant increases over two-test and three-test intervals.

The effect sizes from Tables 2 and 3 were calculated as  $d$ , which is the difference in test score means between administrations, divided by the standard deviation of the score distribution from the earlier of the two administrations. Hence,  $d$  provides mean differences in terms of standard deviation units. Taking a second cognitive ability test, for example, was associated with an average gain of 0.34  $SD$ , or 2.64 points on the 100-point cognitive ability test. Taking a second oral communication ability test corresponded to an average gain of 0.14  $SD$ , or 2.77 points on the 100-point oral communication ability test. Consistent with Kulik, Kulik, and Bangert's (1984) meta-analysis, the effect size with multiple practice trials (i.e., nonconsecutive tests) was greater than with single practices. For cognitive ability tests, the average score increase between Tests 1 and 3 (i.e., with two practices) was 0.76  $SD$ , which translates to 5.79 points. For oral communication ability tests, the average score increase between Tests 1 and 3 was 0.85  $SD$ , yielding an increase of 16.87 points.

The 0.34- and 0.76- $SD$  effect sizes for cognitive ability test score change following single and dual practices were similar to the 0.42  $SD$  and 0.70  $SD$ , respectively, reported by Kulik, Kulik, and Bangert (1984) for identical tests. This similarity emerged despite the fact that of the 19 studies with identical tests across administrations that were used by Kulik et al., 18 were with student samples, and 10 of these were conducted at Grade 6 or below. Also, an unreported number of studies in their meta-analysis used test-retest intervals of less than 2 months, in contrast to our 1-year intervals.

### Test Repetitions and Training Performance

We used multiple regression analysis to assess the relationship between repeat testing and posthire performance in the training academy. Six different models testing the relationship are presented in Table 4. Again, because it seemed reasonable to use either the first or the entry-gaining cognitive and oral communication test scores as the appropriate ability control, we examined

six possible combinations. Each model included either the number of cognitive ability tests or the number of oral communication ability tests as the motivation proxy. Given the sequential and contingent nature of the testing, including both proxies in a single equation would mislead by partialing out the indirect effect when we were interested in the total (i.e., direct plus indirect) effect. The analyses indicated that, regardless of whether number of cognitive ability tests or number of oral communication ability tests was modeled as the motivation proxy, the number of tests taken was positively related to training performance. Moreover, this relationship emerged under various approaches to controlling for ability. These findings provide robust support for Hypothesis 2, which predicted that test repetitions would be associated with higher performance in the training academy. The inference we drew from these findings is that persistence through taking additional tests is indicative of motivation to perform after entry.

In terms of interpreting effect sizes from the regression results, we use, as an example, coefficients for number of tests from Models 1 and 2 in Table 4. When using number of cognitive ability tests as the motivation proxy, the .86 coefficient suggests that taking an additional test, when cognitive and oral communication ability were controlled for, corresponded to a 0.86-point increase in training performance, which was on a 100-point scale. Thus, each additional cognitive ability test that one took was associated with 12% of a standard deviation increase in training performance. For those candidates requiring three or four cognitive ability test administrations rather than one to merit organizational entry, we observed 1.72-point (i.e.,  $2 \times 0.86$ ) and 2.58-point (i.e.,  $3 \times 0.86$ ) increases in training performance, respectively. In contrast, when number of oral communication ability tests was used as the motivation proxy, the 1.49 coefficient suggests that taking a second test corresponded to a 1.49-point increase in training performance score, which is 21% of a standard deviation. For those requiring three oral communication ability test administrations rather than one, we predicted a 2.98-point increase in training performance. In sum, effect sizes were relatively small, with additional oral communication ability tests providing larger effects than additional cognitive ability tests.

Table 4

*Multiple Regression of Training Performance on Number of Tests Taken for Organizational Entry*

Variable	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Age	0.07 (0.08)	0.07 (0.08)	0.07 (0.08)	0.07 (0.08)	0.07 (0.08)	0.07 (0.08)
Gender	6.13** (1.08)	6.36** (1.11)	7.16** (1.11)	7.16** (1.11)	7.16** (1.11)	7.16** (1.11)
Ethnic background	5.13** (0.63)	5.13** (0.63)	5.11** (0.63)	5.11** (0.63)	5.11** (0.63)	5.11** (0.63)
Cognitive ability						
First score	0.21** (0.03)	0.20** (0.03)	0.21** (0.03)	0.20** (0.03)		
Entry-gaining score					0.25** (0.03)	0.24** (0.03)
Oral communication						
First score	0.03** (0.01)	0.03** (0.01)				
Entry-gaining score			0.03** (0.01)	0.03** (0.01)	0.03** (0.01)	0.02** (0.01)
No. of tests taken						
Cognitive ability tests	0.86** (0.22)		0.76** (0.22)		0.41* (0.22)	
Oral communication tests		1.49** (0.39)		1.30** (0.40)		1.03** (0.40)
R <sup>2</sup>	.14	.14	.14	.14	.15	.15
F	80.40**	79.35**	75.86**	74.86**	83.89**	87.15**

Note.  $N = 1,515$ . Standard errors are in parentheses. All models also include dummy variables for year in which the first test was taken.

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

### Test Repetition and Employee Turnover

Drawing on a commitment rationale, we predicted that the number of cognitive and oral communication ability tests would be negatively associated with employee turnover. Table 5 presents the results of the logistic regression analyses used to test these hypotheses, with controls for tenure, age, ethnic background, gender, cognitive ability, oral communication ability, year of entry, and training performance. Under several approaches to modeling the ability controls (similar to Table 4), the number of cognitive ability tests taken by the candidate consistently exhibited a negative, statistically significant relationship with turnover. The same finding emerged for oral communication ability tests. Thus, the more tests participants took, the less likely they were to leave the organization, thereby providing support for Hypotheses 3a and 3b

and for the rationale that test-taking persistence is associated with continuance commitment and subsequent retention.

To extend interpretation of the logistic regression coefficients beyond their sign and statistical significance, one must transform the raw coefficients that are presented in Table 5. As they are, these raw coefficients provide the effect of a 1-unit change in the predictor on the log odds of the outcome. A more intuitive interpretation results from calculating the relationship between the number of tests and the actual probability of turnover, which has the advantage of allowing the estimation of turnover probability at meaningful values of the predictors (for a detailed discussion of logistic regression interpretation, see Liao, 1994; Long, 1997; and Rethorford & Choe, 1993). Thus, for example, to use Model 1 for White men gaining entry in 1991 and with mean values on all other

Table 5

*Logistic Regression of Employee Turnover on Number of Tests Taken for Organizational Entry*

Variable	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Age	-0.45 (0.27)	-0.56 (0.29)	-0.51 (0.29)	-0.56 (0.29)	-0.51 (0.29)	-0.57* (0.29)
Gender	-0.90 (2.31)	0.42 (2.82)	-0.69 (2.45)	0.44 (2.81)	-0.69 (2.46)	0.45 (2.82)
Ethnic background	-2.24 (1.44)	-1.52 (1.88)	-2.01 (1.72)	-1.48 (1.90)	-2.01 (1.72)	-1.48 (1.90)
Tenure	-6.19** (1.58)	-5.59** (1.16)	-6.45** (1.68)	-5.58** (1.16)	-6.43** (1.67)	-5.60** (1.16)
Training performance	-0.02 (0.06)	-0.02 (0.05)	-0.02 (0.06)	-0.02 (0.05)	-0.02 (0.06)	-0.02 (0.05)
Cognitive ability						
First score	0.09 (0.08)	0.05 (0.07)	0.09 (0.08)	0.05 (0.07)		
Entry-gaining score					0.08 (0.08)	0.05 (0.07)
Oral communication						
First score	0.05 (0.03)	0.03 (0.03)				
Entry-gaining score			0.06 (0.03)	0.03 (0.03)	0.06 (0.03)	0.03 (0.03)
No. of tests taken						
Cognitive ability tests	-2.99** (1.27)		-2.98* (1.30)		-2.97* (1.32)	
Oral communication tests		-8.36* (3.87)		-7.93* (4.20)		-7.95* (4.21)
Pseudo R <sup>2</sup>	.34	.34	.34	.34	.34	.34
$\chi^2$	624.14**	621.42**	631.28**	627.24**	631.19**	627.24**

Note.  $N = 1,515$ . Raw coefficients with standard errors in parentheses. All models also include dummy variables for year of organizational entry.

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



variables, increasing the number of cognitive ability tests needed to gain entry from one to two was associated with a decline in turnover probability from .07 to less than .01. For the same individuals, increasing the number of oral communication ability tests needed to gain entry from one to two was associated with a decline in turnover probability from .15 to .01. Simply put, although turnover, at 5.7%, was quite low for the sample as a whole, needing additional tests to gain entry into the training academy was associated with substantially lower turnover probability.

## Discussion

### *Practice Effects*

Consistent with earlier research on practice effects that was conducted in nonemployment settings, this study demonstrated that ability test scores increased with each of the first two repeat administrations. Understanding the practice effects found in this study requires that we speculate as to what exactly drove the score improvements across test administrations. Researchers have previously suggested that test familiarity, decreases in variables such as anxiety or stress, and actual ability increases might lead to practice or coaching effects (Maurer, Solamon, & Troxtel, 1998; Sackett et al., 1989). In terms of enhanced scores due to familiarity with the test, examinees may develop a better understanding of the format of the exam and the types of questions involved and may develop “tricks” to aid in responding to test questions. A second potential explanation for practice effects is that test repetitions may lead to less anxiety and stress when examinees are taking tests. In other words, variance attributable to factors other than true scores is reduced or eliminated, which may improve predictive validity. The third possibility is that examinees might be showing actual improvement in skills or abilities. Sackett et al. suggested that practice effects might be especially likely if feedback has been delivered between tests. Because individuals in our sample received feedback concerning their performance on the test, those candidates who were persistent in taking multiple exams also may have been motivated to develop the requisite skills and abilities to perform well on them (consistent with Sackett et al., 1989, we adopted a broad approach to practice, in that experience with the test could work indirectly on future test scores through feedback and test takers’ subsequent development). Given the time lag of approximately 1 year between tests, candidates had ample time for such development.

It is unfortunate that the design of this study did not allow direct testing of the three potential practice effect explanations. However, examination of predictive validity coefficients (i.e., the correlations between test scores and training performance) may provide some guidance. The validity coefficient was significantly larger for those hired after taking only one test (initial testers) than for those hired after multiple tests (repeat testers) for cognitive ability tests (.36 vs. .24,  $p < .01$ ). A similar pattern emerged for oral communication ability tests (.16 vs. .07,  $p = .10$ ), although the likelihood of a Type I error was greater. One explanation for this finding could be that, in response to the rejection feedback, applicants succeeded at improving their scores by developing the ability that was being measured by the test without actually developing the underlying construct presumed to predict performance. Given that cognitive ability is a stable construct (Gottfredson, 1986), it seems

likely that rather than developing cognitive ability, candidates may have instead simply improved vocabulary skills that were distinct from cognitive ability. These skills would likely be differentially improved across subjects and may not translate to training performance, thus resulting in lower validity. In a similar manner, on the oral communication ability test, the nonhiring feedback may have led some candidates to develop alternative oral presentation strategies that allowed them to score well but that had little to do with posthire performance. One could alternatively argue that test familiarity explains the practice effects in our study. Some examinees may have gained a better understanding of the format of the exam and the types of questions involved and may have developed tricks to aid in responding to test questions. Both of these explanations involve the addition of construct-irrelevant variance that would be consistent with the decrement in validity for repeat testers. Because construct-irrelevant variance would be diminished with a reduction in test anxiety, we would not expect the lower validity for repeat testers if anxiety reduction were a determinant of the practice effects.

We note that range restriction did not appear to contribute to the differential validities. Entry-gaining test score variance was quite similar for repeat and initial testers. It is also worth noting that the lower validities for repeat testers emerged both when repeat testers scored statistically lower on the first test than did initial testers (the case with cognitive ability) and when repeat and initial testers did not differ on the first test (the oral communication ability case). These validity findings are particularly interesting given that “the question of whether practice increases or decreases the predictive validity of the test has largely been ignored” (Sackett et al., 1989, p. 159).

In addition to practice effect validity, cause, and definition, the size of the practice effects is noteworthy. For example, the 0.76-*SD* and 0.85-*SD* increases from Test 1 to Test 3 for the cognitive ability and oral communication ability tests, respectively, yield substantial selection implications. With relatively normal distributions, these cumulative practice effects suggest that someone in the 50th percentile on Test 1 would then move up to approximately the 75th or 80th percentile on Test 3 (assuming similar test score means across administrations). Such a percentile change would clearly have considerable implications for exactly who is hired if the organization allows repeat testing.

### *Retaking Selection Tests and Posthire Performance*

The second major finding in this study is that the persistence evident in retaking selection tests was positively associated with posthire training performance, after we controlled for ability. Although the size of the effect was small, the relationship emerged consistently across several methodological decisions. This previously unstudied relationship suggests that those who had to persevere to gain entry into training (repeat testers) were, on average, more motivated to perform than were those who gained entry after a single test (initial testers). However, given the design of the study, this inference is tenuous. That is, there is no way of knowing whether initial testers would have responded with persistence (i.e., retaking selection tests) if they had not been selected into the organization after the first test. Yet, assuming variance in motivation levels across initial testers and assuming test repetition is a reasonable proxy for motivation, it is probably unlikely that all

initial testers, had they not been selected, would have persevered by repeat testing. Thus, repeat testers, on average, may well have been more motivated than initial testers. In other words, we suggest that we were working with a sample of highly motivated repeat testers and a sample of initial testers with a relatively normal distribution of motivation. After ability differences were partialled out, the motivation advantage of repeat testers presumably manifested itself in slightly greater training performance.

On a conceptual level, two issues regarding test takers' motivation warrant discussion. As we noted earlier, distal motivation processes involve choosing to allocate resources toward the attainment of a goal, whereas proximal motivation processes involve the level of effort expended during task engagement (Kanfer & Ackerman, 1989). Our results suggest, and our interpretation assumes, that the distal test-taking persistence was associated with the proximal effort required for posthire training performance. Although this interpretation seems reasonable, distal motivation was proxied rather than directly measured, and proximal motivation was inferred from performance rather than measured. Hence, more research is necessary to directly examine the distal-proximal motivation relationship. A second issue involves individual personality differences in test takers. Given that personality is sometimes characterized as trait motivation, there may be individual differences that partially explain just who it is that persists in test taking to gain entry and subsequently performs well in posthire training. Conscientiousness, as one of the Big Five personality dimensions (e.g., McCrae & Costa, 1987), would seem to be a likely candidate for this role. Conscientiousness represents, among other characteristics, persistence, a will to achieve, and the energy and discipline required to maintain effort directed at performance (McCrae & Costa, 1987; O'Reilly & Chatman, 1994). These factors clearly should be related to our distal test-taking persistence and the more proximal motivation assumed to be at work in training performance. Moreover, conscientiousness has been linked to job performance (e.g., Barrick & Mount, 1991), further suggesting that the stable personality dimension may play a significant role in accounting for our results.

### *Retaking Selection Tests and Turnover*

Finally, we predicted that test repetitions would likely lead to retention in the organization, given the substantial effort and persistence invested to gain entry through testing. The results support this notion, suggesting a job investment-commitment-turnover link that is consistent with integrated models of employee turnover (Griffeth & Hom, 1995). Employees have developed preentry sunk costs in the form of repeat testing episodes. We suggest that these sunk costs led to increased commitment to the organization and subsequent retention. Alternatively, it is possible that applicants were committed to the idea of a career in law enforcement and this goal drove both the repeated attempts to gain entry and the lower turnover. Given that the overall low turnover rate suggests a strong level of commitment across the entire sample, the occupational aspirations explanation seems plausible. Future research is necessary to determine exactly what underlying processes precipitated the negative relationship between test repetitions and turnover.

### *Implications for Practice*

The implications from this study provide a mixed message as to whether practice effects and allowing applicants to retake selection tests bode well for the organization. The discussion on validity suggests that test familiarity or test-relevant ability increases might lead to the practice effects that we found. Practice effects through test familiarity are likely to be undesirable for the organization, because score improvements would not reflect actual increases in relevant knowledge, skills, or abilities. Such increases, however, may be either positive or negative for the organization, depending on what exact characteristic is being improved. Should the characteristic developed be a valid predictor of posthire performance, allowing multiple tests may be an effective hiring strategy. In our study, however, it appeared that construct-irrelevant improvements were being made (e.g., vocabulary instead of cognitive ability), as evidenced by lower predictive validity for repeat testers. Consequently, with both test familiarity and construct-irrelevant ability improvements, the practice effects with the cognitive and oral communication ability tests in our sample may have resulted in less qualified applicants (in terms of ability) being hired, suggesting that organizations limit selection test repetition.

In contrast, we found considerable evidence for a small, positive relationship between test repetitions and training performance. Thus, organizations may want to allow individuals who do not pass selection hurdles to make additional attempts to do so. The apparent motivation-ability trade-off, however, would likely mean different things for different occupations, as the relative contributions of motivation and ability to job performance change. This point becomes increasingly important as one considers the cumulative nature of the practice effects, which suggests that applicants well below the ability cutoff may eventually score well enough for admittance. In some occupations, the presumed high motivation of such applicants might well compensate for the ability deficit, whereas in other occupations, it would not. Of course, using a measure to assess motivation might be a less expensive and more accurate alternative to multiple test administrations.

The preceding logic assumes that practice effects involve either familiarity or construct-irrelevant ability improvement, as was the apparent case in this study. Some selection tests, however, are designed to assess factors that can be improved between administrations. For example, a lifeguard candidate who passes a lifesaving test on the fourth attempt would likely possess the requisite skill as well as the desired motivation and commitment to excel. Thus, high malleability in what the test is designed to assess may suggest allowing multiple entry attempts, particularly if the measure remains construct-valid on repeated administrations.

Yet, our findings suggesting test familiarity or construct-irrelevant improvement serve to remind us that employers must be careful when allowing for repeated entry attempts. At times, job applicants appear to be able to use multiple tests to score in a way that will get them hired. Although outside of the ability testing domain, Kelley et al.'s (1994) study of the use of the MMPI to screen out psychologically unfit applicants from being hired at a nuclear power plant addressed this issue. The authors found an increasing normality of score profiles across administrations, making it difficult to screen out potentially unstable workers. The authors concluded that the instrument's effectiveness at identifying, and thus screening out, unfit applicants and employees dimin-

ished across repeat administrations. The relevance for ability testing is that when a threshold ability level becomes absolutely critical for performance or safety, practice effects that do not reflect construct-relevant improvements may be problematic. Thus, the source of practice effects should be an important concern for employers allowing multiple entry attempts.

Finally, the finding that repeat testers remained with the organization longer than those who gained entry on a single test seems to support organizations' use of multiple entry attempts. This is particularly true considering that repeat testers may offer organizations the valuable combination of increased performance and lower turnover. The retention effect may be an amplification factor of sorts, in that if practice effects result in adequate or high-performing hires, as in our sample, retention further increases the utility of allowing multiple tests. In contrast, should practice effects bring in employees whose high motivation cannot compensate for ability deficits, the relationship between repeat testing and retention might exacerbate the problem, in that poor performers hired after multiple tests may be even less likely to leave voluntarily than those hired after a single test.

### Limitations

Several limitations of the present study must be noted. First, because this field study used a sample of actual job candidates who were vying for real jobs in a law enforcement organization, participants were not randomly assigned to the number of tests required for entry. Thus, disentangling potentially confounding variables from the repeat-testing variable is a concern. For example, it is unknown whether the employees who gained entry on the first attempt would have persisted had they failed, thereby limiting the confidence with which we can say that number of tests taken, our proxy for motivation, accurately represents this construct. However, we did attempt to control for several possible confounds when we were testing our hypotheses. Furthermore, the use of data from participants in an actual organization improves the ecological validity of the study. A second limitation is that we were unable to comment on actual job performance of the individuals in this sample. Though the posthire training performance finding is suggestive, how repeat testing relates to posttraining performance on the job remains unknown. Third, although we made inferences about and from the explanations for the practice effects, data were unavailable to directly test whether anxiety reduction, test familiarity, or ability improvement was in fact driving the effects. Fourth, it is unknown whether in fact all employee turnover was voluntary, even though our conceptual model assumed that this was the case. Discussions with the law enforcement agency, however, indicated that this likely was the case. Finally, this study examined a sample of participants from a large law enforcement agency. Although these data are valuable, concern must be taken in generalizing the findings until they are replicated with other samples and settings.

### Conclusion

This study is an initial attempt at investigating practice effects on ability tests under actual selection conditions. In addition, this is the first study to examine the relationship between repeat testing and two important outcomes—posthire training performance and

employee turnover. Results showed that candidates improved on the first and second tests by posting higher scores on the second and third tests, respectively. Also, repeat testing was positively associated with training performance, suggesting that individuals who persisted in gaining entry may have been more motivated in subsequent performance domains. Finally, repeat testers were less likely to leave the organization than were those who took a single test to gain entry, suggesting a continuance commitment explanation. Given the paucity of research in this area, more research is needed to understand the relationship between repeat testing and important organizational phenomena. Our study suggests that, within certain constraints, those individuals who are willing to undergo the costs of repeat testing, and eventually succeed in gaining entry to the organization, may be motivated and committed employees who are well worth hiring.

### References

- Arkes, H. R., & Blumer, C. (1985). The psychology of sunk costs. *Organizational Behavior and Human Decision Processes*, 35, 124–140.
- Barrick, M. R., & Mount, M. K. (1991). The Big Five personality dimensions and job performance: A meta-analysis. *Personnel Psychology*, 44, 1–26.
- Becker, H. S. (1960). Notes on the concept of commitment. *American Journal of Sociology*, 66, 32–42.
- Briggs, D. C. (2001). The effect of admissions test preparation: Evidence from NELS: 88. *Chance*, 14, 10–18.
- Gottfredson, L. S. (1986). Societal consequences of the *g* factor in employment. *Journal of Vocational Behavior*, 29, 379–410.
- Griffeth, R. W., & Hom, P. W. (1995). The employee turnover process. *Research in Personnel and Human Resources Management*, 13, 245–293.
- Guion, R. M. (1998). *Assessment, measurement, and prediction for personnel decisions*. Mahwah, NJ: Erlbaum.
- Kanfer, R., & Ackerman, P. L. (1989). Motivation and cognitive abilities: An integrative/aptitude–treatment interaction approach to skill acquisition. *Journal of Applied Psychology*, 74, 657–690.
- 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.
- Kulik, J. A., Bangert-Drowns, R. L., & Kulik, C. C. (1984). Effectiveness of coaching for aptitude tests. *Psychological Bulletin*, 95, 179–188.
- Kulik, J. A., Kulik, C. C., & Bangert, R. L. (1984). Effects of practice on aptitude and achievement test scores. *American Educational Research Journal*, 21, 435–447.
- Liao, T. F. (1994). *Interpreting probability models: Logit, probit, and other generalized linear models*. Thousand Oaks, CA: Sage.
- Long, J. S. (1997). *Regression models for categorical variables and limited dependent variables*. Thousand Oaks, CA: Sage.
- Mathieu, J. E., & Zajac, D. M. (1990). A review and meta-analysis of the antecedents, correlates, and consequences of organizational commitment. *Psychological Bulletin*, 108, 171–194.
- Maurer, T., Solamon, J., & Troxter, D. (1998). Relationship of coaching with performance in situational employment interviews. *Journal of Applied Psychology*, 83, 128–136.
- McCloy, R. A., Campbell, J. P., & Cudeck, R. (1994). A confirmatory test of a model of performance determinants. *Journal of Applied Psychology*, 79, 493–505.
- McCrae, R. R., & Costa, P. T. (1987). Validation of the five-factor model of personality across instruments and observers. *Journal of Personality and Social Psychology*, 57, 17–40.
- Mount, M. K., Barrick, M. R., & Strauss, J. P. (1999). The joint relation-

- ship of conscientiousness and ability with performance: Test of the interaction hypothesis. *Journal of Management*, 25, 707–721.
- Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). New York: McGraw-Hill.
- O'Reilly, C. A., III, & Chatman, J. A. (1994). Working smarter and harder: A longitudinal study of managerial success. *Administrative Science Quarterly*, 39, 603–627.
- Powers, D. E., & Rock, D. A. (1999). Effects of coaching on SAT I: Reasoning test scores. *Journal of Educational Measurement*, 36, 93–118.
- Rapport, L. J., Axelrod, B. N., Theisen, M. E., Brines, B. D., Kalechstein, A. D., & Ricker, J. H. (1997). Relationship of IQ to verbal learning and memory: Test and retest. *Journal of Clinical and Experimental Neuropsychology*, 19, 655–666.
- Rapport, L. J., Brines, B. D., Axelrod, B. N., & Theisen, M. E. (1997). Full scale IQ as a mediator of practice effects: The rich get richer. *Clinical Neuropsychologist*, 11, 375–380.
- Rethorford, R. D., & Choe, M. K. (1993). *Statistical models for causal analysis*. New York: Wiley.
- Ryan, A. M., Ployhart, R. E., Greguras, G. J., & Schmit, M. J. (1998). Test preparation programs in selection contexts: Self-selection and program effectiveness. *Personnel Psychology*, 51, 599–621.
- Sackett, P. R., Burris, L. R., & Ryan, A. M. (1989). Coaching and practice effects in personnel selection. In C. Cooper & I. Robertson (Eds.), *International review of industrial and organizational psychology* (pp. 145–183). New York: Wiley.
- Staw, B. M., & Hoang, H. (1995). Sunk costs in the NBA: Why draft order affects playing time and survival in professional basketball. *Administrative Science Quarterly*, 40, 474–494.
- Tett, R. P., & Meyer, J. P. (1993). Job satisfaction, organizational commitment, turnover intention, and turnover: Path analyses based on meta-analytic findings. *Personnel Psychology*, 46, 259–293.
- Watson, F. L., Pasteur, M. L., Healy, D. T., & Hughes, E. A. (1994). Nine parallel versions of four memory tests: An assessment of form equivalence and the effects of practice on performance. *Human Psychopharmacology*, 9, 51–61.
- Wright, P. M., Kacmar, K. M., McMahan, G. C., & Deleeuw, K. (1995).  $P = f(M \times A)$ : Cognitive ability as a moderator of the relationship between personality and job performance. *Journal of Management*, 21, 1129–1139.

Received August 22, 2000

Revision received April 25, 2001

Accepted May 29, 2001 ■