

## Research Article

## THE SUSCEPTIBILITY OF OVERT AND COVERT INTEGRITY TESTS TO COACHING AND FAKING

George M. Alliger, Scott O. Lilienfeld, and Krystin E. Mitchell

*The University at Albany, State University of New York*

**Abstract**—Although previous research has indicated that faking can affect integrity test scores, the effects of coaching on integrity test scores have never been examined. We conducted a between-subjects experiment to assess the effects of coaching and faking instructions on an overt and a covert integrity test. Coaching provided simple rules to follow when answering test items and instructions on how to avoid elevated validity scale scores. There were five instruction conditions: "just take," "fake good," "coach overt," "coach covert," and "coach both." All subjects completed both overt and covert tests and a measure of intelligence. Results provided strong evidence for the coachability of the overt integrity test, over and above the much smaller elevation in the faking condition. The covert test apparently could be neither coached nor faked successfully. Scores on both integrity tests tended to be positively correlated with intelligence in the coaching and faking conditions. We discuss the generalizability of these results to other samples and other integrity tests, and the relevance of the coachability of integrity tests to the ongoing debate concerning the prediction of counterproductive behavior.

Paper-and-pencil integrity tests are used primarily to detect prospective employees who are thought to be at high risk for committing theft and engaging in other related counterproductive behaviors in the workplace. These tests are administered to millions of individuals each year (Sackett & Harris, 1984). Integrity tests can be placed into two categories. Overt tests have a purpose that is clear to the test taker, and covert tests have a disguised purpose (Murphy, 1993; Sackett, Burns, & Callahan, 1989). Overt tests consist of questions concerning respondents' history of, and attitudes toward, counterproductive workplace activities such as theft. Covert tests consist of questions concerning personality traits, such as conscientiousness and impulse control, that are presumably related to integrity. Despite compelling meta-analytic evidence of their validity for a variety of work-related criteria (Ones, Viswesvaran, & Schmidt, 1993), integrity tests remain controversial (Camara & Schneider, 1994; Goldberg, Grenier, Guion, Sechrest, & Wing, 1991; Lilienfeld, 1993; Sackett, 1994; U.S. Congressional Office of Technology Assessment, 1990).

One important question regarding integrity tests is their susceptibility to impression management. A few published studies have investigated the ability to fake honesty on integrity tests. Ryan and Sackett (1987) found that, for an overt integrity test closely modeled after existing overt tests, scores in a fake-good condition were significantly higher (more "honest") than those

in a respond-honestly condition. Ryan and Sackett also reported that subjects in the fake-good condition received markedly elevated scores on a social desirability scale embedded within the integrity test, suggesting that subjects who attempt to create a positive impression on overt integrity tests may be detectable by validity indices. Gough (1972) found that subjects could raise their scores slightly on the Personnel Reaction Blank, a commonly used covert test, when instructed first to take the test honestly and then to "fake the responses of an ideal applicant for employment" (p. 5). LoBello and Sims (1993) reported that male prisoners asked to "fake good" produced higher scores on a commercial overt integrity test than did both prisoners given no instructions and those asked to respond candidly.

In addition to the issue of fakability, the question of whether integrity tests are susceptible to coaching is important, for at least two reasons. First, most integrity tests assess a circumscribed number of domains and are generally quite similar to one another in content (Murphy, 1993), raising the possibility that they could be passed by means of a relatively simple set of coaching instructions. In fact, at least one informal but surprisingly sophisticated tutorial for passing integrity tests is already popularly available (Source, 1993), although the effectiveness of such tutorials has not been evaluated empirically. Second, coachability historically has been seen as an important psychometric issue concerning tests that have selection consequences (e.g., Johnson & Wallace, 1989; Jones, 1986). The vulnerability of tests to coaching is a critical scientific question that has important implications for practice. One major concern regarding use of the polygraph, for example, is its vulnerability to easily coached countermeasures (Honts, Raskin, & Kircher, 1994; Lykken, 1981). Researchers seem to be becoming aware that integrity test scores may be elevated if subjects possess information about the content of these tests. Cunningham, Wong, and Barbee (1994) explained briefly to subjects some of the rationale underlying a commercial overt test, and suggested that subjects might want to keep this rationale in mind while they took the test. Cunningham et al. concluded that the information provided had no practical effect on test scores (p. 650), but the information manipulation fell far short of the coaching test takers might encounter (e.g., Source, 1993).

The purpose of the present study was fourfold. First, we examined the extent to which integrity tests are susceptible to coaching. We hypothesized that integrity tests are susceptible to coaching, and that coaching instructions would demonstrate effects over and above faking instructions. Second, we compared an overt and a covert integrity test in terms of their susceptibility to coaching and faking. We expected the overt test to be more susceptible to both sets of instructions because of the greater transparency of most of its items. Third, we examined the relations between scores on an intelligence test and

Scott O. Lilienfeld is now at Emory University.

Address correspondence to George M. Alliger, Department of Psychology, The University at Albany, State University of New York, Albany, NY 12222.

performance on integrity tests under normal, faking, and coaching conditions. Although researchers have generally reported weak or essentially nonexistent relations between scores on integrity tests and measures of intelligence (e.g., Werner, Jones, & Steffy, 1989), we predicted that subjects' ability to engage in response distortion following faking or coaching instructions would partly depend on intelligence. Fourth, we examined the extent to which coaching and faking instructions affected scores on a validity scale designed to detect impression management. We predicted that faking instructions would increase scores on this scale (see Ryan & Sackett, 1987), but that the coaching instructions, which included strategies to avoid detection on such validity scales, would not.

## METHOD

### Subjects

The subjects in this experiment were 195 undergraduates. Forty-six percent of the subjects were male, the mean age was 19.3. Subjects were enrolled in either introductory psychology or industrial-organizational psychology at a large Northeastern university.

### Design

The design of this study was a five-condition (instructions) between-subjects factorial. Subjects in each group received one of five sets of instructions indicating how to answer the questions on either the overt test, the covert test, or both. Subjects in all five conditions completed both the overt and the covert tests, along with lie scales embedded within each test. The order of presentation was counterbalanced within each condition.

Development of the instructions occurred as follows. Based on a review of the literature on integrity tests, we developed a list of topic areas assessed by a wide range of overt and covert integrity tests. To ensure that the coaching instructions were general and widely applicable, none of the coaching instructions provided sample items from existing integrity tests. Moreover, the coaching instructions were developed prior to any examination of the integrity tests used in this study. Thus, these instructions were not tailored to the specific integrity tests used.

### Procedure

Subjects completed the tests in groups of 2 to 8 individuals, each group being randomly assigned to a condition. Subjects first completed a brief demographics sheet. The experimenter then read some background information regarding the uses and purposes of integrity tests. Subjects then completed the Shipley Institute of Living Scale (Shipley, 1940), a 20-min test of intelligence.

Subjects received one of five sets of audiotaped instructions. While the appropriate tape was played, the instructions were simultaneously displayed on an overhead screen. Subjects were not allowed to write down any of the instructions, or to ask the

experimenter how to answer the integrity test questions once the instructions were given.

After receiving instructions, subjects in all conditions completed the overt and covert tests at their own pace. They were not told which test was overt or covert, or even that there is such a distinction. They then completed a multiple-choice test with 11 items, covering the information in the mnemonics taught in the coaching instructions. The multiple-choice test was included as a manipulation check for all conditions. In addition to the multiple-choice test, subjects in the coaching conditions completed a mnemonic recall test. Finally, all subjects were given a questionnaire measuring work history.

### Instructions

After completing the intelligence test, all subjects were asked to consider that they had just graduated and were applying for a job in which they were very interested. The instructions they were given are presented here in summarized form only. There have been recent discussions concerning whether detailed accounts of the instructions used to coach individuals to dissimulate their responses on personality tests and validity scales should be provided in published articles (e.g., Ben-Porath, 1994; Berry, Lamb, Wetter, Baer, & Widiger, 1994). It seems generally agreed that although the degree to which a test is susceptible to coaching is an important research issue, there is a conflict between maintaining test confidentiality and providing details in published research articles. Accordingly, only very general instructions are presented here.

#### *Condition 1 Just take*

Subjects were given no coaching on how to answer the questions on either the overt or the covert integrity test, and were told to answer the questions as candidly as possible. Subjects were not alerted to the existence of or provided with instructions concerning how to answer the items on the validity scales embedded within the integrity tests.

#### *Condition 2 Fake good*

Subjects were asked to try to "beat the tests" by appearing as honest as possible. As in Condition 1, subjects were not given coaching on how they should answer any questions assessing honesty or the items on the validity scales.

#### *Condition 3 Overt coaching*

Subjects received coaching on how to answer items on the overt integrity test only. Subjects in this condition were told that each letter in the mnemonic "PLACE" stood for a hint for answering questions on the overt test. *P* stood for "protect." Subjects were told that they should not protect dishonest co-workers from punishment. *L* stood for "lenient." Subjects were told they should not be lenient in punishing co-workers who are caught stealing and should endorse the strongest possible punishment for infractions. *A* stood for "admit." Subjects were told not to admit to any illicit activities, or even to being tempted to perform such activities. *C* stood for "common." Subjects were told to endorse responses indicating that dishonest behaviors are not common. *E* stood for "excuses." Subjects

## Coaching Integrity Tests

were told to state that no excuse is sufficient to justify illicit activities

*Condition 4 Covert coaching*

Subjects received coaching for items on the covert integrity test only, via the mnemonic "SCARED." S stood for "safety conscious." Subjects were told to endorse responses indicating that they would never take risks of any kind or entertain thoughts of dangerous activities. C stood for "conformist." Subjects were told to indicate that they were traditional and conventional. A stood for "agreeable." Subjects were told to endorse responses implying that they get along well with other people. R stood for "restrained." Subjects were instructed to indicate they had good impulse control and to state that they carefully consider the consequences of their activities. E stood for "even tempered." Subjects were instructed to indicate that they are emotionally stable and relaxed. D stood for "dependable." Subjects were instructed to indicate that they are conscientious, reliable, and hardworking.

*Condition 5 Overt and covert coaching*

Subjects received both the overt coaching and the covert coaching instructions, in the form just described.

*Instructions for lie items*

Subjects in the three coaching conditions also received guidance on how to identify and answer lie items. They were told that lie items assess whether respondents are trying to make themselves appear too virtuous or perfect, and were given two examples of such items. Subjects were told first to ascertain whether each item was a lie item, and were instructed to answer any apparent lie item so as not to appear perfect.

**Measures***Overt integrity test*

The overt integrity test was the Employee Integrity Index (EII), developed by Ryan and Sackett (1987). This test contains 62 items modeled closely after items used on published overt integrity tests. Ryan and Sackett (1987) reported acceptable internal consistency for the EII ( $\alpha = .77-.93$  for attitude items and  $.56-.74$  for admissions items, under various instructional conditions). The EII was used in lieu of an overt integrity test owned by a test-publishing company, despite numerous attempts, we were unable to obtain permission to use a published overt integrity test for this study. A comparison of items on the EII and published integrity tests, however, showed that the EII was remarkably similar to the published tests.

*Covert integrity test*

The Personnel Reaction Blank (PRB, Gough, 1972) was used as the covert measure of integrity. The PRB is a commonly used 70-item personality-oriented measure designed to assess the construct of *wayward impulse*. Wayward impulse subsumes dependability, conscientiousness, and conformity (Gough, 1971). Gough (1972) reported a split-half reliability of  $r = .73$  for college males and  $r = .65$  for college females, and a 5-year interval

test-retest reliability of  $r = .56$ . Gough (1972) also reported that the PRB distinguished known delinquents from nondelinquents ( $r_{pb} = .58$  for males and  $r_{pb} = .57$  for females). Frost and Rafilison (1989) reported that the PRB correlated significantly with counterproductivity.

*Lie scales*

We embedded in the integrity tests 14 items taken from the Unlikely Virtues scale of Tellegen's (1982) Multidimensional Personality Questionnaire. Seven items were embedded in each integrity test. Tellegen's items assess the degree to which an individual endorses extremely rare but virtuous qualities or behaviors. The Unlikely Virtues scale is similar to the Lie scale of the Minnesota Multiphasic Personality Inventory (Hathaway & McKinley, 1940), except that it contains an equal number of items keyed true and false. A typical Unlikely Virtues item is "I have at times eaten too much."

*Shipley Institute of Living Scale*

This is a 60-item test of intelligence containing 40 items measuring verbal ability and 20 items measuring problem solving. Deaton (1992) reported that this test has good test-retest reliability and correlates approximately .85 with the Wechsler Adult Intelligence Scale-Revised.

*Multiple-choice test*

As noted earlier, we developed an 11-item multiple-choice test on the material presented in the coaching instructions. Each question corresponded to material covered by one of the letters in one of the mnemonics, PLACE and SCARED. Thus, the test consisted of a 5-item subtest covering the overt coaching content and a 6-item subtest covering the covert coaching content. Subjects in all five conditions received this test to determine the extent to which they attended to the coaching instructions.

*Mnemonic recall test*

Subjects in the coaching conditions were asked to recall the mnemonic (or mnemonics, in Condition 5) provided in their instructions, and to fill in a blank indicating the word associated with each letter in the mnemonic (or mnemonics).

*Inventory of work experience*

This inventory, adapted from Ryan and Sackett (1987), was designed to assess if subjects had ever taken an integrity test or had worked in industries where such tests are common.

**RESULTS****Manipulation Check**

Table 1 presents means and standard deviations both for the overt coaching ( $\alpha = .76$ ) and the covert coaching ( $\alpha = .69$ ) sections of the multiple-choice test and for the mnemonic recall test. On the overt multiple-choice subtest, subjects in the overt coaching conditions scored significantly better than subjects in the other conditions. For scores on the covert multiple-choice subtest, the coach-both and coach-covert conditions were not

**Table 1** Means and standard deviations for heuristic multiple-choice and mnemonic recall tests by coaching condition

Coaching condition	Multiple-choice test								
	N	Overt coaching subtest			Covert coaching subtest		Mnemonic recall test		
		Mean	SD	Mean	SD	N	Mean	SD	
Just take	41	2.51 <sup>b</sup>	1.47	3.71 <sup>b</sup>	1.68	—	—	—	
Fake good	36	2.64 <sup>b</sup>	1.79	3.64 <sup>b</sup>	1.78	—	—	—	
Coach overt	38	4.26 <sup>a</sup>	1.41	3.66 <sup>b</sup>	1.66	37	72 <sup>a</sup>	26	
Coach covert	36	2.31 <sup>b</sup>	1.58	4.25	1.59	37	41 <sup>b</sup>	30	
Coach both	42	4.24 <sup>a</sup>	1.10	4.81 <sup>a</sup>	1.49	41	37 <sup>b</sup>	29	
Analysis of variance		$F(4, 118) = 16.96^{***}$			$F(4, 118) = 3.94^{**}$		$F(2, 112) = 17.00^{***}$		

Note: The multiple-choice test was scored for number correct, the mnemonic recall test was scored for proportion correct. Superscripts represent within-condition means significantly different by least significant difference test,  $p < .05$  (i.e., a > b).

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

significantly different from each other, and subjects in the coach-both condition scored higher than those in the conditions with no coaching for the covert test.

Table 1 also shows means for the mnemonic recall test. Performance on this test was scored to indicate the proportion of words recalled from either one or both mnemonics. Subjects in the coach-overt condition performed significantly better than subjects in the other two coached conditions.

### Mean Test Score Differences Across Conditions

Performance for each condition on the two integrity tests is shown in Table 2. The analysis of variance (ANOVA)  $F$  between conditions for the EII was statistically significant,  $F(4, 190) = 25.97, p < .001, \eta^2 = .35$ . Post hoc pair-wise tests indicated that the EII means in the overt coaching conditions were significantly higher than in the other conditions. In addition,

the average score for the just-take condition was significantly lower than average scores for the other conditions. The analogous ANOVA  $F$  for the PRB was not statistically significant,  $F(4, 190) = 0.54, p = .70, \eta^2 = .01$ . The mean for the PRB in the just-take condition was typical of that found in employee samples (e.g., Frost & Rafilston, 1989). Because the mean for the EII in the coach-covert condition is significantly higher than that for the just-take condition, there is an indication of transfer of covert coaching effects on the overt test. There is little or no indication of coaching transfer in the other direction. Also, no Coaching Condition  $\times$  Test Order effect was found.

We also examined the degree to which coaching or faking instructions permitted subjects to attain the highest possible score on the integrity tests. Because no significant mean differences between conditions were found for the PRB, we report ceiling analyses only for the EII. Because the EII has a maximum score of 310, the average score in the just-take condition

**Table 2** Means and standard deviations for integrity tests by coaching condition

Coaching condition	N	Overt test (EII)		Covert test (PRB)	
		Mean	SD	Mean	SD
Just take	41	210.27 <sup>b,d</sup>	23.33	31.37	4.05
Fake good	36	235.36 <sup>b,c</sup>	50.34	31.94	5.60
Coach overt	38	278.11 <sup>a</sup>	29.17	32.24	4.18
Coach covert	38	229.00 <sup>b,c</sup>	34.40	32.53	4.76
Coach both	42	267.43 <sup>a</sup>	32.44	32.67	3.86
Analysis of variance		$F(4, 190) = 25.97^{***}$		$F(4, 190) = 0.54$	

Note: EII items were rated 1 through 5, PRB items were rated 0 or 1. Superscripts represent within-condition means significantly different by least significant difference test,  $p < .05$  (i.e., a > b, c > d). EII = Employee Integrity Index, PRB = Personnel Reaction Blank.

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

## Coaching Integrity Tests

is about 68% of the possible score. In the other conditions, this same conversion yields 76% of maximum for the fake-good condition, 90% for the coach-overt condition, 74% for the coach-covert condition, and 86% for the coach-both condition. Another way to look at this question would be to ask what percentage of subjects attained, for instance, 90% of the maximum possible score on the EII. Such an analysis showed that no students in the just-take condition attained a score this high. In contrast, 19% of the subjects in the fake-good condition, 58% in the coach-overt condition, 5% in the coach-covert condition, and 45% in the coach-both condition did so.

Table 3 displays the means for the lie scales embedded within the EII and the PRB. For both tests, the lie scale scores in the fake-good condition were higher than those of the other conditions (this difference was significant over each of the other conditions for the EII lie scale, and significant over three of the other four conditions for the PRB lie scale). For both tests, the lie scale means for the coaching conditions were no higher than for the just-take condition.

## Correlations

Correlations among scores on both tests and the lie scales are presented in Table 4. The correlation between the EII and the PRB is .25 in the just-take condition (see Frost & Rafelson, 1989, for a similar finding). This correlation is much higher in the fake-good condition and in the coach-covert and coach-both conditions.

In the just-take condition, the correlation between the EII and its lie scale was .23. In the faking condition, the EII-lie scale correlation was .52 ( $p < .001$ ), whereas the correlation was nonsignificant in the coaching conditions. This pattern effectively repeats for the correlations between the PRB and its lie scale. These results, together with the means shown in Table 3, indicate that the lie scales functioned appropriately in the noncoaching conditions. In the coaching conditions, however, the lie scales showed no elevation and essentially no correlation with their respective tests.

Correlations between the Shipley measure of intelligence and each integrity test tended to be highest in the faking and relevant coaching conditions, suggesting that more intelligent subjects responded to the faking and coaching instructions somewhat more effectively. As can be seen in Table 5, correlations between scores on the integrity tests and on the multiple-choice tests were statistically significant in nearly every case. Thus, even in the noncoaching conditions, individuals who could intuit the heuristics for the integrity tests obtained higher scores. The correlation between mnemonic recall and integrity score was positive for the overt test in the overt-coaching conditions.

## DISCUSSION

In this study, we showed that coaching of an overt integrity test can be remarkably successful, and that the effects of coaching significantly surpass those of faking. The overt coaching consisted of a few simple heuristics based on the content of items found in overt tests in general. Coaching may surpass the effects of faking in part because some subjects who attempt to fake overt integrity tests may adopt misguided strategies. Some respondents, for example, may believe that greater openness regarding their flaws and transgressions will cause them to appear honest (Cunningham, 1989). Coaching counteracts these strategies by eliminating misconceptions concerning the scoring of integrity tests.

There are several reasons to believe that coaching instructions would achieve the same success for overt integrity tests in the real world as found in this study. First, the subjects were drawn from the same pool as many applicants for jobs for which integrity testing is common. Seventy-eight percent of the subjects indicated that they had worked or were working in a retail store, restaurant, or bank (where integrity tests are often administered), and 17% reported having taken a test similar to those they encountered in this study. Generalizability of the study to the relevant job applicant population can therefore be

Table 3 Means and standard deviations for lie scales on each integrity test by coaching condition

Coaching condition	N	Lie scale (EII)		Lie scale (PRB)	
		Mean	SD	Mean	SD
Just take	41	17.54 <sup>b,c</sup>	3.73	1.44 <sup>b</sup>	1.34
Fake good	36	20.47 <sup>a</sup>	6.00	2.31 <sup>a</sup>	1.94
Coach overt	38	16.55 <sup>b</sup>	4.60	1.37 <sup>b</sup>	1.48
Coach covert	38	15.18 <sup>b,d</sup>	4.19	1.26 <sup>b</sup>	1.25
Coach both	42	17.54 <sup>b,c</sup>	5.29	1.79	1.54
Analysis of variance		$F(4, 190) = 6.00^{***}$		$F(4, 190) = 2.88^*$	

Note. Lie scale items had the same rating format as the relevant test (i.e., EII lie scale items were rated 1-5, and PRB lie scale items were rated 0 or 1). Superscripts represent within-condition means significantly different by least significant difference test,  $p < .05$  (i.e., a > b, c > d). EII = Employee Integrity Index, PRB = Personnel Reaction Blank.  
\* $p < .05$  \*\*\* $p < .001$

**Table 4** Correlations among integrity tests, lie scales, and Shipley test of intelligence, by coaching condition

	PRB	EII lie scale	PRB lie scale	Shipley
		Just take		
EII	25	23	17	20
PRB		-.06	23	17
EII lie scale			33*	.00
PRB lie scale				-.15
		Fake good		
EII	52***	52***	44**	36*
PRB		20	61***	22
EII lie scale			52***	-.04
PRB lie scale				13
		Coach overt		
EII	24	-.11	27	35*
PRB		-.09	22	23
EII lie scale			34*	-.33*
PRB lie scale				-.02
		Coach covert		
EII	58***	.02	37*	16
PRB		.04	21	35*
EII lie scale			49***	-.41*
PRB lie scale				-.08
		Coach both		
EII	52***	.07	14	28
PRB		.00	29	15
EII lie scale			43**	-.23
PRB lie scale				-.14

Note: *N*s for the correlations range from 36 to 42. EII = Employee Integrity Index. PRB = Personnel Reaction Blank.

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

assumed to be high. Second, the items on the EII are extremely similar to those of published overt tests. Generalizability of the results to other overt integrity tests can therefore also be assumed to be high.

In contrast to the effects of coaching on the overt test, coaching for the covert test failed completely. Why should this bifurcation occur? One possibility resides in the fact that the coaching instructions were devised prior to examination of the content of either test. It could be argued that we did not do as good a job in developing the coaching instructions for the covert test as we did for the overt test. After reviewing the covert coaching instructions, however, we concluded that they do correctly address the rationale of personality-based integrity test theory (see also Murphy, 1993). Another possible explanation of the failure to observe covert-coaching effects is that subjects were not able to remember the covert mnemonic well. This explanation was suggested by a reviewer and is supported by the mnemonic recall means shown in Table 1. To address this possibility, we reran the three coaching conditions (40 undergraduates per condition). However, in this follow-up study, we

handed out copies of the relevant mnemonic (or mnemonics, in the coach-both condition) and its associated words to all subjects for reference while they took the tests. A main effect for coaching was again found for the overt test among the coach-overt, coach-covert, and coach-both conditions ( $M = 250.40, 233.20, \text{ and } 258.43$ , respectively,  $F[2, 117] = 4.98, p < .01, \eta^2 = .08$ ), pair-wise post hoc tests showed the means for the coach-overt and coach-both conditions to be significantly higher than that for the coach-covert condition,  $p < .05$ . The means for the PRB again showed no significant differences ( $M = 31.68, 32.18, \text{ and } 32.98$ , respectively,  $F[2, 117] = 0.97, p = .38, \eta^2 = .02$ ). This finding indicates that the ineffectiveness of coaching on the covert test cannot be attributed to poor memory of the mnemonics.

Instead, it may be that covert integrity tests are in general more resistant to coaching than overt tests. Item subtlety may be an explanation, item subtlety on the empirically keyed PRB, for example, seems substantial. Although greater item subtlety is associated with lower item validity (Hough, Eaton, Dunnette, Kamp, & McCloy, 1990), it should be noted that covert and

## Coaching Integrity Tests

**Table 5** Correlations between integrity tests and the multiple-choice and mnemonic recall tests, by coaching condition

Integrity test	Multiple-choice test		Mnemonic recall test
	Overt coaching subtest	Covert coaching subtest	
	Just take		
EII	38*	27	---
PRB	39*	40**	---
	Fake good		
EII	56***	73***	---
PRB	24	60***	---
	Coach overt		
EII	49***	59***	30
PRB	-07	37*	-21
	Coach covert		
EII	66***	37*	09
PRB	39*	48**	06
	Coach both		
EII	53***	36*	40**
PRB	07	30	09

Note: *N*s for the correlations range from 36 to 42. EII = Employee Integrity Index, PRB = Personnel Reaction Blank. \**p* < .05. \*\**p* < .01. \*\*\**p* < .001.

overt integrity tests have approximately equal validity (Ones et al., 1993). In any case, given that other covert integrity tests appear to be modeled largely after the PRB (Frost & Rafilison, 1989), the resistance of the PRB to coaching may extend to other covert tests.

The present study suggests several avenues for further research. First, constructive replication (Lykken, 1968) would seem desirable. To what extent will coaching effects be found with other overt tests? We hypothesize, for reasons listed earlier, that such replications will be successful. Second, examination of the duration of coaching effects is necessary. Third, the degree of resistance of other covert tests to coaching instructions similar to those outlined here, or to alternative strategies, is important. Fourth, although previous studies (e.g., Hough et al., 1990) did not find that response distortion on employment tests moderated their validity, subjects in those studies were not provided with strategies for how to obtain higher scores. Future research should examine the extent to which coaching attenuates the validity of integrity tests.

In sum, coaching of the overt test seems to have been completely successful. Test score means were higher with overt coaching than in the just-take or fake-good conditions, but lie scale means were not higher than in the just-take condition. Although it appears that more intelligent individuals may be somewhat more likely to benefit from coaching, an increase in scores due to coaching occurred across the board. Indeed, many coached subjects will obtain very high scores on such

tests, and yet not be identifiable by validity scales as engaging in impression management. This result is analogous to findings on counterdetection methods for the polygraph (e.g., Honts et al., 1994) and suggests that, like the polygraph, overt integrity tests can be passed by means of a simple set of instructions. The finding that overt integrity tests are susceptible to coaching is thus one important piece of information that needs to be considered in the ongoing debate concerning the use of integrity tests.

**Acknowledgments**—A previous version of this manuscript was presented at the May 1995 meeting of the Society for Industrial and Organizational Psychology in Orlando, Florida. The authors would like to express gratitude to Margot Lutvak-Ganster for help in contacting test publishers, Brian Andrews for assistance in developing coaching instructions, Judy Cannata, Lauren McKillop, Stephanie Penner, David Rapp, Gretchen Scherer, and Melissa Feigelson for assistance in running subjects, and Glenn Sanders, the reviewers, and the editor for comments on an earlier version of this manuscript.

## REFERENCES

- Ben Porath, Y. S. (1994). The ethical dilemma of coached malingerer research. *Psychological Assessment*, 6, 14-15.
- Berry, D. T. R., Lamb, D. G., Wetter, M. W., Baer, R. A., & Widiger, T. A. (1994). Ethical considerations in research on coached malingerer. *Psychological Assessment*, 6, 16-17.
- Camara, W. J., & Schneider, D. L. (1994). Integrity tests: Facts and unresolved issues. *American Psychologist*, 49, 117-119.
- Cunningham, M. R. (1989). Test-taking motivations and outcomes on a standardized measure of on-the-job integrity. *Journal of Business and Psychology*, 3, 54-66.
- Cunningham, M. R., Wong, D. T., & Barbee, A. P. (1994). Self-presentation dynamics on overt integrity tests: Experimental studies of the Reid Report. *Journal of Applied Psychology*, 79, 643-658.
- Deaton, W. L. (1992). The Shipley Hartford. In J. J. Kramer & J. Cloze (Eds.), *The eleventh mental measurements yearbook* (pp. 822-824). Lincoln: University of Nebraska Press.
- Frost, A. G., & Rafilison, F. M. (1989). Overt integrity tests versus personality-based measures of delinquency: An empirical comparison. *Journal of Business and Psychology*, 3, 269-277.
- Goldberg, L. R., Gremier, J. R., Gunn, R. M., Sechrest, L. B., & Wing, H. (1991). Questionnaires used in the prediction of trustworthiness in pre-employment selection decisions: An APA task force report. Washington, DC: American Psychological Association.
- Gough, H. G. (1971). The assessment of wayward impulse by means of the Personnel Reaction Blank. *Personnel Psychology*, 24, 669-677.
- Gough, H. G. (1972). *Manual for the Personnel Reaction Blank*. Palo Alto, CA: Consulting Psychologists Press.
- Hathaway, S. R., & McKinley, J. C. (1940). A multiphasic personality schedule (Minnesota): I. Construction of the schedule. *Journal of Psychology*, 10, 249-254.
- Honts, C. R., Raskin, D. C., & Kircher, J. C. (1994). Mental and physical countermeasures reduce the accuracy of polygraph tests. *Journal of Applied Psychology*, 79, 252-259.
- Hough, L. M., Eaton, N. K., Dunnette, M. D., Kamp, J. D., & McCloy, R. A. (1990). Criterion-related validities of personality constructs and the effect of response distortion of those validities [Monograph]. *Journal of Applied Psychology*, 75, 581-595.
- Johnson, S. T., & Wallace, M. B. (1989). Characteristics of SAT quantitative items showing improvement after coaching among Black students from low-income families: An exploratory study. *Journal of Educational Measurement*, 26, 133-145.
- Jones, R. F. (1986). The effect of commercial coaching courses on performance on the MCAT. *Journal of Medical Education*, 61, 273-284.
- Lilienfeld, S. O. (1993). Do "honesty tests really measure honesty?" *Skeptical Inquirer*, 18, 32-41.
- LoBello, S. G., & Sims, B. N. (1993). Feasibility of a commercially produced pre-employment integrity test. *Journal of Business and Psychology*, 8, 265-273.

- Lykken D.T. (1968) Statistical significance in psychological research. *Psychological Bulletin* 70, 151-159
- Lykken D.T. (1981) *A tremor in the blood: Uses and abuses of the lie detector*. New York: McGraw-Hill
- Murphy, K.R. (1993) *Honesty in the workplace*. Pacific Grove, CA: Brooks/Cole
- Ones D.S., Viswesvaran C., & Schmidt, F.L. (1993) Meta analysis of integrity test validities: Findings and implications for personnel selection and theories of job performance [Monograph]. *Journal of Applied Psychology* 78, 679-703
- Ryan, A.M. & Sackett P.R. (1987) Pre-employment honesty testing: Falsability reactions of test takers and company image. *Journal of Business and Psychology* 1, 248-256
- Sackett P.R. (1994) Integrity testing for personnel selection. *Current Directions in Psychological Science* 3, 73-76
- Sackett P.R., Burris, I.R. & Callahan C. (1989) Integrity testing for personnel selection: An update. *Personnel Psychology* 42, 491-529
- Sackett P.R. & Harris M.M. (1984) Honesty testing for personnel selection: A review and critique. *Personnel Psychology* 37, 221-245
- Shuply W. (1940) A self-administering scale for measuring intellectual impairment and deterioration. *Journal of Psychology* 26, 371-377
- Source U.R. (1993) How to hack honesty. 2600, 10, 20-23
- Tellegen A. (1982) *Manual for the Multidimensional Personality Questionnaire*. Unpublished manuscript, University of Minnesota, Minneapolis
- U.S. Congressional Office of Technology Assessment (1990) *The use of integrity tests for pre-employment screening* (Report No. OTA-SET-442). Washington, DC: U.S. Government Printing Office
- Werner S.H., Jones J.W., & Steffy B.D. (1989) The relationship between intelligence, honesty, and theft admissions. *Educational and Psychological Measurement* 49, 921-927

(RECEIVED 9/5/94, REVISION ACCEPTED 4/28/95)

## Forthcoming Papers

### Feature Article

The Verdict on Freud  
*Frederick Crews*

### General Article

The Anatomy of Impact: What Makes an Article Influential?  
*Robert J. Sternberg and Tamara Gordeeva*

### Research Articles

Cognitive Suppression During Saccadic Eye Movements  
*David E. Irwin and Laura A. Carlson-Radvansky*

Cerebral Asymmetry and Cognitive Development in Children:  
A Magnetic Resonance Imaging Study  
*Christiana M. Leonard, Linda J. Lombardino, Laurie R. Mercado,  
Samuel R. Brown, Joshua I. Breier, and O. Frank Agee*

The Role of Task and Context in Preference Measurement  
*Barbara A. Mellers and Alan D. J. Cooke*

### Research Reports

Effects of Evidence on Attitudes: Is Polarization the Norm?  
*Deanna Kuhn and Joseph Lao*

Word Frequency and Word Difficulty: A Comparison of Counts in Four Corpora  
*Hunter M. Breland*

Infant Attachment Security and Affective-Cognitive Information Processing at Age 3  
*Jay Belsky, Becky Spritz, and Keith Crnic*

Averageness, Exaggeration, and Facial Attractiveness  
*Gillian Rhodes and Tanya Tremewan*

Repetition Deficits Interact With Syntactic Grouping in Rapid Serial Visual  
Presentation Sentences  
*Lise Abrams, Jennifer R. Dyer, and Donald G. MacKay*



This document is a scanned copy of a printed document. No warranty is given about the accuracy of the copy. Users should refer to the original published version of the material.