

5-5-1995

A Comparative Study of Integrity Tests: the Effect of Situational and Individual Variables on Response Distortion

Nannette Shayne Searles
Portland State University

Follow this and additional works at: https://pdxscholar.library.pdx.edu/open_access_etds



Part of the [Psychology Commons](#)

Let us know how access to this document benefits you.

Recommended Citation

Searles, Nannette Shayne, "A Comparative Study of Integrity Tests: the Effect of Situational and Individual Variables on Response Distortion" (1995). *Dissertations and Theses*. Paper 5061.
<https://doi.org/10.15760/etd.6937>

This Thesis is brought to you for free and open access. It has been accepted for inclusion in Dissertations and Theses by an authorized administrator of PDXScholar. Please contact us if we can make this document more accessible: pdxscholar@pdx.edu.

THESIS APPROVAL

The abstract and thesis of Nannette Shayne Searles for the Master of Science in Psychology were presented May 5, 1995, and accepted by the thesis committee and the department.

COMMITTEE APPROVALS:

[Redacted Signature]

Dean E. Frost, Chair

[Redacted Signature]

Leslie B. Hammer

[Redacted Signature]

Peter Bullard

[Redacted Signature]

Ann Fullerton
Representative of the Office of
Graduate Studies

DEPARTMENT APPROVAL:

[Redacted Signature]

James A. Paulson, Chair
Department of Psychology

ACCEPTED FOR PORTLAND STATE UNIVERSITY BY THE LIBRARY

by

[Redacted Signature]

on

17 July 1995

Abstract

An abstract of the thesis of Nannette Shayne Searles for the Master of Science in Psychology presented May 5, 1995.

Title: A Comparative Study of Integrity Tests: The Effect of Situational and Individual Variables on Response Distortion

The paper-and-pencil integrity test, which is used in industry as an employee selection device, has been largely developed outside the mainstream of psychological testing. The result has been that some testing programs have inadequately standardized testing conditions and/or other safeguards to ensure valid test results. Studies have shown that response distortion, or faking, is a problem with all types of tests, integrity tests being no exception. A correlation between the construct underlying integrity testing, such as the personality trait of conscientiousness, has yet to be investigated.

The primary purpose of the present study was to determine how response distortion on integrity tests is affected by the instructions given by test administrators. Also, the connection between integrity tests and conscientiousness is examined. Finally, comparisons were made between currently published integrity tests/scales.

Two hundred and forty-nine college students participated in this study by completing the following three integrity tests and two personality test scales: The London House Personnel Selection Inventory (London House Press, Inc., 1980); The Tesco Survey (Bullard, 1992); The Reid Report (Reid, 1967); The Value Orthodoxy Scale from The Jackson

Personality Inventory (Jackson, 1976), and the Work Orientation Scale from the California Personality Inventory (Gough, 1985). The three integrity tests offer a variety of validation studies in support of their scales. Also, all three test publishers participated in a pre-publication review of the 1991 APA Task Force Report on integrity testing (Goldberg, Grenier, Guion, Sechrest, & Wing).

Results show that responses were affected by instructions given by the administrator. For example, the analysis shows that in a job application situation, an applicant who believes a prospective employer is using an integrity test to identify undesirable applicants will tend to distort his/her responses on a theft scale to appear more favorable. Also for the theft scales, instructions to deliberately give false responses in a socially desirable way did not differ from a standard job applicant instruction set. Instructions emphasizing the ability of the tests to identify high integrity employees also did not differ from the standard instruction set's results. Results also show that conscientiousness is correlated with integrity test scales. Finally, the integrity tests and personality scales studied here are significantly correlated. Implications of these findings are discussed.

**A COMPARATIVE STUDY OF INTEGRITY TESTS: THE EFFECT OF
SITUATIONAL AND INDIVIDUAL VARIABLES ON
RESPONSE DISTORTION**

by

NANNETTE SHAYNE SEARLES

A thesis submitted in partial fulfillment
of the requirements for the degree of

**MASTER OF SCIENCE
in
PSYCHOLOGY**

Portland State University

1995

Introduction

Problem Statement

Today's employer is concerned about hiring workers who will be reliable and not prone to theft. As shown by recent researchers, their concerns may be well-founded. In 1989, Slora and Boye (cited in Jones, Slora, & Boye, 1990) found that in the supermarket industry 43% of anonymously surveyed employees admitted to some type of theft of company cash or property, with an estimated average loss of \$44.72 per year per employee. Similar research conducted in 1989 by The Food Marketing Institute (cited in Jones, Slora, & Boye, 1990) found that 52.9% of all supermarket theft is attributable to supermarket employees. In 1986, Hefter reported that one in three employees steals from their employer. Finally, in 1987, Shephard and Duston (cited in Camara & Schneider, 1994) reported estimates that American businesses lose from \$15 billion to \$25 billion per year due to employee theft.

Up until 1988 when the Employee Polygraph Protection Act (House of Representatives, 1988) was passed, organizations widely used the polygraph test as a pre-employment screening device to identify applicants who might steal from their company. Since that is no longer an option for most areas of the private sector (i.e., the 1988 Act forbids most employers from using such tests), paper-and-pencil integrity tests have become widely used as selection devices by employers concerned about employee theft.

Of concern in the current research is the context within which these tests are being given and how that influences response distortion, also

referred to as faking, on these tests. Context here means the instructions (i.e. the rationale for administering the test) given the applicant when taking the integrity test. Another concern involving integrity tests is that different tests have not been compared using the same set of subjects. The final issue that this study addressed was whether or not the personality trait of conscientiousness correlates positively with current integrity tests and whether this trait can help detect individuals who are distorting their responses to these tests.

What is an Integrity Test?

Integrity tests have been defined as "psychological tests designed to predict job applicants' proneness for theft and other forms of counterproductivity" (Jones & Terris, 1991, p. 124). These tests, commonly used by organizations in which employees will be given access to cash and/or merchandise, question applicants about their attitudes toward theft, self-reports of any past thefts the applicant may have committed, and a variety of other questions concerning things such as safety attitudes, general work ethic, drug use habits, and overall integrity.

In 1989, O'Bannon, Goldinger, and Appleby (cited in Jones & Terris, 1991) found that over 40 integrity tests were in use. In a 1994 report, Camara and Schneider identified 46 separate publishers or developers of integrity tests. Most of these tests have proprietary scoring keys, making it difficult to make comparisons across tests. According to Sackett, Burris, and Callahan (1989) this practice is "not common to mainstream psychological tests used for employment purposes" (p. 494). The present research attempted to compare three currently available integrity tests and

two personality scales to determine if they are truly measuring the same constructs.

How Valid Are Integrity Tests?

There is evidence that integrity tests can help organizations eliminate certain problems related to employing dishonest workers. One example is a time series study conducted by Brown, Jones, Terris, and Steffy (1987). In this study, after a three year baseline period, the Personnel Selection Inventory (PSI) was introduced as part of the selection procedure at a chain of home improvement centers located in the western United States. For two years, only those candidates that passed the PSI, along with other selection criteria, were hired. At the end of those two years, the organization had a 50% reduction in employee terminations for theft, illegal drug use, and violence. In addition, the organization saved over two million dollars in inventory shrinkage losses over the two year period. These results extend the findings produced by Terris and Jones who used the PSI to study shrinkage reduction in convenience stores in 1982.

Another question is which integrity tests are valid tools for predicting which prospective employees will exhibit unwanted behaviors once hired. A handful of researchers have offered reviews of available integrity tests. In 1984, Sackett and Harris reviewed over 40 studies conducted using 10 available integrity tests and found that both "skeptics and advocates" of the tests could find fuel for their fire. The skeptic, they say, could complain about such things as faulty criteria and the reliance upon self-report measures, the latter of which are susceptible to the effects

of faking and social desirability (i.e., a response bias tending to create a more favorable impression). On the other hand, advocates of paper-and-pencil integrity testing can note the "consistency of positive findings across tests and across validation strategies" (Sackett & Harris, 1984, p. 241).

As recently as 1994, Camara and Schneider released a review of two independent reports done on available integrity tests. One report was from a 20-month study, ending in 1991, conducted by the American Psychological Association (APA). The other report was from a two-year study conducted by the U.S. Congress Office of Technology Assessment (OTA) and ending in 1990. The methods of review were very different as well as the results. The OTA, who limited their review to five predictive validity studies, all of which used detected theft or a "close proximity" as a criterion, concluded that: ". . .these reports were inconclusive in supporting or dismissing the assertion that integrity tests can reliably predict dishonest behavior in the workplace (p. 115)." The APA, on the other hand, reviewed close to 300 studies which covered a large variety of validity designs. Their conclusion was that: ". . .for those few tests for which validity information is available, the preponderance of the evidence is supportive of their predictive validity (cited in Camara & Schneider, 1994, p. 115)."

In 1993, Ones, Viswesvaran, and Schmidt conducted the first comprehensive meta-analysis of integrity test validities. The database for their study included 665 criterion-related validity coefficients which came from over 180 studies, technical reports, and personal communications.

Their results indicate that integrity tests are indeed substantial predictors of job performance and counterproductive behaviors on the job.

The Issue of Response Distortion

One of the problems most often mentioned in a discussion on paper-and-pencil integrity tests is that of the risk that the test taker will respond in a socially desirable manner or fake their answers on the tests to make themselves appear desirable to the employer. This issue is not new. It has been a concern with all personality and self-report tests.

Research on response distortion has reaped contradictory conclusions. In a now classic study on response bias, Dunnette, McCartney, Carlson, and Kirchner (1962) asked subjects to either respond honestly to a self-description checklist or to respond to it as a successful salesman would. They found that not only could subjects respond in a chosen direction, but that even a small amount of "faking" could distort the validity of this personnel test. More recently, Nicholson and Hogan (1990) have claimed that subjects responding in a socially desirable manner is not really a problem after all. Their argument is that there is considerable content overlap between personality scales and social desirability scales. Therefore, when social desirability is controlled for, validity coefficients are actually decreased. In 1990, Hough, Eaton, Dunnette, Kamp and McCloy asked groups of recent U.S. Army entrants to answer items on a temperament inventory in one of three ways: in a way that would ensure that the Army would select them (Fake Good); in a way that would ensure that the Army would not select them (Fake Bad); or in a way that describes how they really are (Honest). They too discovered that test takers "can

distort their self-descriptions when instructed to do so" (p. 593). But their study found little evidence that response distortion, or faking, significantly alters test validities.

In 1987, Ryan and Sackett conducted a study similar to those described above. These researchers asked college students to complete an "honesty test" (designed by them and patterned after existing integrity tests) under one of three instructional sets: "answer honestly; fake good (e.g., attempt to appear honest); and respond as you would if you were applying for a job" (p. 250). These researchers found that individuals can indeed respond to such a test in a desirable manner. Their contribution to the literature, however, was their discovery that those subjects instructed to respond as a job applicant "respond in a manner far more similar to subjects told to respond truthfully than to subjects told to fake good" (Ryan & Sackett, 1987, p. 255).

The present study was not conducted to assess criterion validities. In other words, no criterion data will be studied. Instead, response patterns to the various tests will be compared. This study's primary purpose was to determine if the instructions given to job applicants when taking an integrity test (i.e., context of test taking) would influence their tendency toward response distortion. At the same time this study compared response distortion rates on different integrity tests and/or scales using the same set of subjects.

The Test-Taking Context: Instructions and Rationale

The interest in integrity testing has developed outside the mainstream of Industrial/Organizational psychology, and most research on

the topic continues to be conducted by test publishers. In light of this, several reviewers (Camara & Schneider, 1994; Sackett et al., 1989; Sackett & Harris, 1984) have made pleas for I/O psychologists to take a more active role in the area "in order to provide sound advice to organizations concerned about theft and counterproductivity" (Sackett et al., 1989, p. 524). To this end, integrity test publishers formed the Association of Test Publishers (Association of Personnel Test Publishers, 1990), which now, among other things, helps educate test users about the legal and ethical responsibilities in the use of the tests.

A consequences of this early lack of careful professional scrutiny of test use, however, is that some publishers have had little regard for the standardization of testing conditions. In fact, according to a survey of integrity test publishers (with a 65% response rate by those identified as publishers) conducted over a 20-month period by the APA and ending in 1991 (Camara & Schneider, 1994), of those surveyed:

. . .64% reported that they do not use test user qualifications forms to screen potential purchasers, with 56% of these using no formal screening practices of any kind. . .[and] only 5% of the publishers required either a graduate degree or specialized licensure or certification for test users. (p. 114)

It appears that some publishers have made little or no effort to control for things such as demand characteristics, which are "aspects of the situation itself that *demand* that people behave in a particular way" (Myers & Hansen, 1993, p. 254), or experimenter biases, in which the person administering the test gives cues as to what is expected of the test taker, or

even behaves differently in subsequent testing situations, differentially influencing test takers.

Although research by Jones and Joy (cited in Jones & Terris, 1991), Ryan and Sackett (1989), and Stone, Stone, and Hyatt (cited in Camara & Schneider, 1994) has shown that job applicants do not tend to have negative attitudes about taking integrity tests as part of the selection process, values of the employer and the culture of the organization can be communicated by the act and manner of asking a potential employee to complete such a test. According to Chatman (1989), when there is "congruence between the norms and values of organizations and the values of persons" (p. 339), this is referred to as "person-organization fit." Chatman says that there are two methods by which organizations can ensure that this fit occurs. An employer can socialize new employees in a way that helps them incorporate the values and norms of the organization, or they can select employees who already have incorporated the same values and norms into their belief system.

Of concern in this study is that the instructions given a job applicant when they are asked to complete an integrity test may in fact influence them to respond in a manner to fit themselves to the organization, thus creating a false perception of person-organization fit. Therefore, organizations might unintentionally create testing environments which lead to more applicants distorting responses. A second unintentional result would be the possibility of new hires constructing an incorrect impression of the organization's values based on the stated reasons for administering the integrity test. Even more likely, however, is the situation

in which applicants take integrity tests without administrative explanation and then each applicant makes his or her own assumptions about the values of the given organization.

Personality and Integrity Testing: Conscientiousness

In their comprehensive meta-analysis described previously, Ones et al. (1993), argue that all integrity tests are in fact measuring the broad construct of conscientiousness, which they claim reflects such characteristics as dependability, carefulness, and responsibility. In their words:

. . .these findings raise the question of whether general conscientiousness is actually the motivation variable that has been so elusive in personnel psychology. . .that is, conscientiousness may be the most important trait motivation variable. (p. 696)

A recent addition to the California Personality Inventory, the Work Orientation Scale (Gough, 1985), which is described below, claims to measure the personality trait of conscientiousness. It was used in the current study to correlate this trait with three integrity tests and another personality scale. In addition, this study looked at the connection between conscientiousness and response distortion on integrity tests.

Research Goals

Based on the above discussion, the following were goals of the current research:

- 1) to complete a comparative study of three integrity tests and two personality scales, including response distortion rates,

- 2) to investigate a personality characteristic which may correlate with integrity tests as well as predict faking,
- 3) and, to examine how test administration instructions affect levels of response distortion on integrity tests.

Research Hypotheses and Questions

The current research was exploratory in nature, and there were three general questions that the researcher attempted to answer:

1. Comparison of Integrity Tests

- a) Are integrity test theft scales measuring the same or different constructs? For example, to what extent are the theft scales on different integrity tests correlated?
- b) Will faking vary between the different integrity tests and will the impact of the experimental conditions affect the results of the tests in similar or different ways?

2. Personality Characteristic: Conscientiousness

Is conscientiousness positively correlated with integrity test theft scales and can it be used to predict faking?

3. Test Administration Instructions

Will there be a consistent difference in test scores of subjects across the four experimental conditions? Response distortion will be most pronounced in the Fake Good condition and least pronounced in the Applicant condition. No directional hypothesis is made as to whether there will be more faking found in the Prevent Theft or the High Integrity conditions.

(see Procedures section for an explanation of the experimental conditions)

Method

Subjects

Subjects consisted of students taking psychology courses at Portland State University. They were offered extra course credit for participation. This is an appropriate sample for these research questions. As pointed out by Ryan and Sackett (1987), since college students tend to be a part of the population of "low level applicants in industries such as retailing and finance" (p. 255), they are representative of the population being studied.

Materials

For this experiment a group of current integrity tests were used to gather the necessary data. The tests used were: The London House Personnel Selection Inventory (London House Press, Inc., 1980); The Tesco Survey (Bullard, 1992); The Reid Report (Reid, 1967); and The Value Orthodoxy Scale from The Jackson Personality Inventory (Jackson, 1976). As has been described previously, the Work Orientation Scale from the California Personality Inventory (Gough, 1985) was administered to assess conscientiousness. In addition to these tests, demographic information was gathered for exploratory purposes.

London House PSI. The Personnel Selection Inventory was designed to identify job applicants with tendencies toward dishonesty and other forms of counterproductive behavior. The version PSI-35 was used for this

study. It includes the following scales: Honesty, Nonviolence, Drug Avoidance, Work Values, Validity, and Safety. The Honesty scale, which measures attitudes about theft and the probability that a person will not engage in theft behavior will be referred to here as the London House Theft Scale. The Validity scale, which measures to what extent a subject is answering in a socially desirable manner, will be referred to here as the London House Faking scale. The Theft and Faking scales have 47 and 22 items respectively, with the Theft scale scores ranging from 47 to 190 and the Faking scale scores ranging from 22 to 110.

Questions on the PSI come in three forms: rating scales, checklists, and open-ended questions. Only the objectively scored rating scales and checklists were included in the current research. Sample questions are as follows: "Would you say that you are too honest to steal?," which is rated on the following Likert-type scale: 1 = definitely yes, 2 = probably yes, 3 = uncertain, 4 = probably no, and 5 = definitely no. Checklist questions deal exclusively with theft admissions with questions such as: "Within the last 3 years, check the nearest total dollar value of all money you have taken without proper authorization from jobs. Include all money from all employers in the last 3 years." Eleven dollar amount options were offered for this question, ranging from "\$0" to "\$5,000 or more," as well as a category of "can't remember."

A great variety of validity studies have been conducted on the PSI. In 1979, Terris found a correlation of .56 between test scores and theft admissions obtained during a polygraph examination. Similarly, in 1982, Jones (cited in Sackett & Harris, 1984) found that the honesty scale of the

PSI correlated .56 with theft admissions. A Spearman-Brown estimate of .95 has been reported for the reliability of the PSI (Terris, 1979). Finally, a reviewer for the Mental Measurements Yearbook (Sauser, 1985) concluded, after reviewing 21 studies conducted on the PSI, that the test "appears to be a reliable and fair instrument" (p. 871).

Tescor Survey. The Tescor Survey is a general test of integrity. The test instrument is 96 items in length and consists of four scales used for scoring: Theft, Drug, Hostility, and Faking. Two scales, Work Traits and Safety, are also on the current form but are for research interests of the publisher. All questions are multiple choice with four possible answers. For example, a question from the Theft Scale is as follows: "Excluding minor office supplies, what would you guess is the value of things you have secretly taken from work in the past 2 years? 1) None; 2) \$5-\$25; 3) \$26-\$500; 4) \$501 or more." The Theft and Faking scales were included in this study. There are 13 items on the Theft scale and 18 items on the Faking scale with scale scores ranging, respectively, from 13 to 28, and 18 to 72.

The reported reliability of the Faking and Hostility scales are .81 and .86 respectively (Bullard, 1992). The publisher contends that biographical data and self-report items concerning Theft and Substance Abuse yield nominal rather than ordinal or interval scales, and that traditional reliability measures are thus inappropriate for such scales. A correlation coefficient of .54 has been reported between the Tescor Survey theft scale and the PSI theft scale (American Tescor, Inc., 1985). In addition, the publisher monitors EEOC compliance quarterly, and offers ample data that

the Tesco Survey does not discriminate on the basis of age, sex, or race (American Tesco, Inc., 1994).

The Reid Report. The Reid Report was designed to "assess the honesty and potential dishonesty of applicants for employment" (Brodsky, 1978, p. 1025). The version used in this research is referred to as "The Short Form" and consists of only 55 items (whereas the earlier forms consisted of up to 150 questions). For purposes of this research, only 48 items were used. Four open-ended questions dealing with topics such as work history, two questions about owning a driver's license, and one question asking for dates and descriptions of felony actions were not used. Scales on the Reid Report measure Theft, Faking, Hostility, Integrity, and Work Ethic. The response formats for the items are either yes-no items, rating scales, or checklists. An example of a yes-no item is as follows: "Were you ever tempted to take company merchandise without actually taking any?" An example of a rating scale item is as follows: "I rate my own honesty as follows: 1) far below average, 2) below average, 3) slightly below average, 4) average, 5) slightly above average, 6) above average, and 7) far above average." The Theft scale, which was included in this research, consists of 25 items with the scale score ranging from 25 to 50.

Large amounts of reliability and validity information have been gathered on the Reid Report. Brodsky (1978) reports that, from 11 samples, K-R 20 reliabilities for males ranged from .89 to .93, with a median of .92. For females, the range was from .71 to .86, with a median of .80. Brodsky also gives information on validity based on correspondence between polygraph results and results on the test. Results show an overall

validity coefficient of .62. The Reid Report has also been shown to be nondiscriminatory on both variables of race and gender (Brodsky 1978; Willis, 1985).

JPI Value Orthodoxy Scale. The Jackson Personality Inventory was developed to "assess a variety of personality characteristics in normally functioning individuals" (Dyer, 1985, p. 369) and is used for prediction in industry settings as well as an aid in counseling and in personality research. It is a test of 320 True/False items that constitute 16 distinct personality scales. The only scale of interest to this investigation was Value Orthodoxy (Vo), which contains 20 items dealing with "conscientiousness" issues similar to the Wo scale of the CPI. Exploratory comparisons will be made between the Vo scale and the Wo scale. The following is an example of a question from that scale: "Cheating and lying are always wrong, no matter what the situation." Scale scores for the Vo scale range from 20 to 40.

Studies have shown the JPI to be a reliable tool. Using the method of internal consistency, two studies showed median coefficients of .93 and .90 (in Dyer, 1985). Concerning validity of the test, the manual for the JPI (Jackson, 1976) reports average scale correlations between the test and two peer rating studies between .35 and .40.

CPI Work Orientation Scale. The Wo (or Work Orientation) scale is a recent adaptation of the California Personality Inventory. It is a "special purpose scale" intended for use in the occupational world and has as its function the assessment of "the sense of commitment and obligation to work that one finds in persons of exceptionally conscientious, dependable,

and self-disciplined temperament" (Gough, 1985, p. 505). The Wo scale consists of 40 True/False questions. Sample questions are: "Most people would tell a lie if they could gain by it," and "I do not mind taking orders and being told what to do." The Work Orientation Scale measured the personality construct conscientiousness in this study. Scale scores for the Wo range from 40 to 80.

The Wo scale has been shown to be reliable in several ways. Gough (1985) reports alpha coefficients of .75 for both male and female college students. Test-retest correlations of .70 and .62 were found for male and female high school students respectively. Validity information shows that the Wo scale correlates in the following manner on scales of similar content: .84 (males) and .85 (females) with the Well-Being scale of the CPI; .78 (males) and .79 (females) with the Emotional Stability scale of the Guilford-Zimmerman Temperament Survey; .53 (males only) with the Adjustment scale of the Hogan Personality Inventory; and .77 (males) and .75 (females) with the Personal Integration scale of the Omnibus Personality Inventory (reported in Gough, 1985).

Test Scoring. Scoring keys were available for Tesco Survey, the Wo scale, and the Vo scale. Since scoring keys were not available for the Reid Report and London House, each item was examined to determine what scale it belonged to (theft, faking, etc.). For all scales included in this experiment, scores were keyed so that lower scores predict less of a tendency to steal from the employer. In other words, low scores indicate that more desirable employee behaviors are likely.

Demographics. The following demographic information was requested from the subjects for exploratory purposes: age; sex; self-reported G.P.A.; hours worked per week (currently); years/months of life spent in full-time employment (i.e., at 35 hours a week or more). In addition, subjects were asked whether they have ever been asked to take a test similar to these when applying for a job, as well as asked the following open-ended questions: "If you can recall, what were the instructions given for completing these tests," and "What do you think is the employer's reason for using these tests?"

Procedure

Subjects were randomly assigned to one of four conditions: 1) Fake Good; 2) Applicant; 3) Prevent Theft; and 4) High Integrity. Participants in the "Fake Good" group were given the following instructions:

"Take the following tests in a way that will present yourself in the best possible light. That is, respond to each item in the most socially acceptable manner, in a way that creates the best possible impression of who you are."

Subjects in the "Applicant" condition received these instructions:

"Assume that you are applying for a job and your prospective employer has asked you to take the following tests as part of the application process. Answer these items just as you would if you were applying for a job you needed."

Subjects in the "Prevent Theft" condition were instructed to:

"Take the following tests as if you were applying for a job. Your prospective employer has informed you that their establishment has

had problems with employee theft in the past and they are having you take these tests so that they can select employees who will not be dishonest."

Finally, instructions for the "High Integrity" condition were the following:

"Take the following tests as if you were applying for a job. Your prospective employer has informed you that their organization has a work force that is honest and has high moral standards. They have asked you to take these tests because they would like to hire employees who will fit in well with their organizational culture."

In each session, which lasted no more than two hours, participants were assured complete anonymity and, in an effort to detect order-effects, the sequence of tests taken by each group was counter-balanced. Three different orders were randomly selected from among the large set of all possible orders. Among the subjects in each condition, equal numbers received the three test orders. While this design did not control for order effects, it does allow for partial testing of the presence of any order effects in the three sequences used.

Results

Two hundred and forty-nine subjects, 99 male and 150 female, participated in this study. Subjects ranged in age from 18 to 72 ($M = 24.65$; $sd = 7.96$). Only 65 of the 249 subjects said that they had been asked to take an integrity test in a job application situation prior to the testing. At the time of testing subjects were working anywhere from zero hours per week to 50 hours per week, with a mean of 16.17 hours ($SD =$

12.85). The range of years subjects had spent in full-time employment (35 hours a week or more) was zero to 33, with a mean of 4 years ($SD = 6$). Finally, participation by number in experimental conditions was as follows: Fake Good = 61; Applicant = 76; Prevent Theft = 56; and High Integrity = 56.

As a manipulation check, subjects were asked the following open-ended question: "If you can recall, what were the instructions given for completing these tests?" Results revealed that 57% of subjects were aware of and did remember the specific instructions they were given prior to taking the tests (for example, to complete the tests as if they were applying for a job). There were 22.5% of the subjects that recalled incorrectly, while 18.1% recalled alternative instructions (i.e., instructions to use a #2 pencils, etc.), and 2.4% left the question blank or answered unintelligibly. Because this item required free recall from the subjects, it would be expected that there would be fewer correct responses than if the item had been a multiple choice, or recognition, item. A MANOVA revealed no significant main effect on the theft scales for those who had or had not been asked to take an integrity test prior to this testing, which would show that prior exposure to integrity tests did not affect testing results.

Correlation coefficients were calculated to explore Research Question 1: "Do integrity tests containing scales with the same names measure the same or different constructs?" Correlations reveal that the scales are measuring similar constructs. For example, Table 1 shows correlations of .73 and .72 ($p < .001$) between the London House theft scale and the Reid Report and Tesco Survey theft scales respectively.

Despite the fact that all correlations among theft scales are significant, a test of significance of the difference between r_1 and r_2 shows that there is a significant difference between the correlation based on the Reid Report and Tesco Survey scores ($r = .54$) and the correlation based on the Reid Report and London House scores ($r = .73$), $Z = 3.39$, $p < .01$. Similarly, the correlation of the Reid Report and the Tesco Survey theft scales ($r = .54$) is significantly different from the correlation based on the London House and Tesco Survey scores for theft ($r = .72$), $Z = 3.17$, $p < .01$. There are moderately strong negative correlations between the Work Orientation Scale of the CPI and the three theft scales (see Table 1 below). In fact, the correlation between the Work Orientation Scale and the London House theft scale ($r = -.61$, $p < .001$) is higher than the correlation between the theft scales of the Tesco Survey and the Reid Report ($r = .54$, $p < .001$).

TABLE 1

Descriptive Statistics for Dependent Measures

	Mean	SD	TS	LH	RR	WO	VO
TS	15.95	3.43	1.00				
LH	81.01	17.38	.72**	1.00			
RR	37.76	5.45	.54**	.73**	1.00		
WO	29.59	5.42	-.47**	-.61**	-.42**	1.00	
VO	9.95	4.09	-.20*	-.26**	-.33**	.15	1.00

-TS = Tesco Theft Scale; LH = London House Theft Scale; RR = Reid Report Theft Scale; WO = Work Orientation Scale; VO = Value Orthodoxy Scale.

* $p < .01$

** $p < .001$

N = 219

Research Question 2, which asked whether conscientiousness is correlated with integrity tests, is answered by examining the correlations in Table 1 (above). Correlations between the Wo scale and the theft scales range from $-.42$ to $-.61$ ($p < .001$). The negative correlations indicate an inverse relationship between conscientiousness and theft behavior. The correlations also indicate that the personality trait of conscientiousness can indeed be considered as a potential predictor of theft behavior. A test of the difference between two correlation coefficients shows that the correlation between the Wo scale and the London House theft scale is not significantly different than the correlation between the Reid Report theft scale and the Tesco theft scale, $Z = 1.09$, ns. The correlation between the Wo scale and the Tesco theft scale is not significantly different than the correlation between the Reid Report theft scale and the Tesco theft scale, $Z = .98$, ns.

Conscientiousness can also be used to predict rates of faking since it is significantly correlated with the faking scales. There is a moderately high negative correlation between the Wo Scale and the faking scales on Tesco Survey and London House ($r = -.53$ & $-.51$, respectively, $p < .001$). In other words, the Wo Scale correlates with both the Theft and Faking scales of the London House test. Although they are significantly different from zero, the correlations between the Vo scale and the Wo scale and the three theft scales are considerably lower than the other correlations (the correlation coefficients range from $+.15$ to $-.33$). This indicates that the Vo scale may not be measuring the same construct as the other scales.

A multivariate analysis of variance (MANOVA) was used to examine Research Question 3 regarding the influence of instructional set on test scores. There is a significant multivariate main effect for the three theft scales with $F(3,108) = 2.14, p < .05$, but subsequent univariate tests are not significant. When dependent variables are highly intercorrelated this may happen. One way to interpret the results is to use 95% Confidence Intervals. Examination of Table 2 (below) reveals the following means on the Tesco theft scale for the four instructional sets: Prevent Theft, 15.38; Fake Good, 15.80; High Integrity, 16.11; and Applicant, 16.36. With the Applicant condition constituting the control or comparison group, comparisons made to it show that only the Prevent Theft condition strays outside the Applicant group's 95% Confidence Interval. This indicates that

TABLE 2

**Tesco Survey: Means, Standard Deviations, and
95% Confidence Intervals for Theft Scale**

Condition^a	Mean^b	SD	95% Conf. Interval
Fake Good	15.80	3.59	14.89 to 16.72
Applicant	16.36	3.64	15.52 to 17.19
Prevent Theft	15.38	2.88	14.60 to 16.16
High Integrity	16.11	3.46	15.17 to 17.05

^aFake Good=responses made in 'most socially acceptable manner'; Applicant=responses made as if applying for a job that was needed; Prevent Theft=responses made as if applying for a job in which the employer mentioned a past problem with employee theft; and High Integrity=responses made as if applying for a job in which the employer noted the company's highly moral work force.

^bN = 247

instructions given to test takers that mention the prospective employer wanting to hire employees who will not steal from their organization causes a significant bias in subjects' responses. Those subjects in the Prevent Theft condition appear to have responded in the most socially desirable fashion, appearing least likely to be undesirable employees in the future. Table 3 (below) shows that results were similar for the London House test, with means for instructional set as follows: Prevent Theft, 77.80; Fake Good, 80.35; High Integrity, 81.45; and Applicant, 83.50. Again, Prevent Theft was the only condition falling outside of the Applicant group's 95% Confidence Interval.

TABLE 3

London House: Means, Standard Deviations, and 95% Confidence Intervals for Theft Scale

Condition^a	Mean^b	SD	95% Conf. Interval
Fake Good	80.35	19.32	75.23 to 85.48
Applicant	83.50	17.82	79.31 to 87.69
Prevent Theft	77.80	14.54	73.71 to 81.89
High Integrity	81.45	16.98	76.68 to 86.23

^aFake Good=responses made in 'most socially acceptable manner'; Applicant=responses made as if applying for a job that was needed; Prevent Theft=responses made as if applying for a job in which the employer mentioned a past problem with employee theft; and High Integrity=responses made as if applying for a job in which the employer noted the company's highly moral work force.

^bN = 231

Table 4 (below) shows a different pattern of results for the Reid Report. In this instance there is considerable overlap in the intervals of the

Applicant, Prevent Theft, and High Integrity conditions. It is the Fake Good condition that is significantly different from the Applicant condition. This suggests that the Reid Report is more easily faked than the London House or Tesco Survey, but is not as sensitive to other types of instructional sets.

TABLE 4

**Reid Report: Means, Standard Deviations, and
95% Confidence Intervals for Theft Scale**

Condition^a	Mean^b	SD	95% Conf. Interval
Fake Good	36.57	5.46	35.18 to 37.97
Applicant	38.00	5.48	36.74 to 39.26
Prevent Theft	38.26	5.21	36.83 to 39.70
High Integrity	38.30	5.57	36.77 to 39.84

^aFake Good=responses made in 'most socially acceptable manner'; Applicant=responses made as if applying for a job that was needed; Prevent Theft=responses made as if applying for a job in which the employer mentioned a past problem with employee theft; and High Integrity=responses made as if applying for a job in which the employer noted the company's highly moral work force.

^bN = 242

To determine if subjects had pre-conceived explanations for why employers use integrity tests, the following open-ended question was asked: "What do you think is the employer's reason for using these tests?" Coding for the question was as follows: 1=to get high integrity employees; 2=to test the person's honesty; 3=to screen out undesirable applicants; 4=specific mention of drug use, theft, etc.; 5=a combination of above reasons; 6=all other reasons; 7=blank/don't know/unintelligible. Based on

their open-ended responses, subjects were placed into one of the seven categories, and differences between categories were tested with the three theft scales as dependent variables. A MANOVA revealed a significant main effect, with $F(3,106.5) = 1.99, p < .01$. As can be seen from the means in Table 5 (below), subjects who assumed that the employer was using the

TABLE 5

**Means and Percentages for Open-ended Item
Concerning Beliefs About Test Administration**

Category	Mean^a	SD	%
High integrity	15.44	2.88	3.6
Personal honesty	16.22	3.44	26.9
Screen out undesirables	14.65	2.21	6.8
Mention of drug use, theft, etc.	15.64	3.11	8.8
Combination of above reasons	15.16	2.93	14.9
All other reasons	16.32	3.80	36.9
Blank/don't know/unintelligible	18.20	4.09	2.0

^aN = 249

test to screen out individuals who would be "undesirable" had the lowest average scores, implying that they distorted their answers to appear more desirable to the prospective employer. This supports the previous finding contained above in Tables 2 and 3 in which those subjects, given instructions concerning the prospective employer's desire to not hire potential thieves, distorted their answers in a desirable direction. This held true for Tesco Survey and London House, but not the Reid Report. A one-way ANOVA showed that assumptions about an employer's use of integrity

tests were not affected by the instructions given at the beginning of the testing session, $F(3,243) = 3.25$, ns.

Regarding gender differences in test scores, a MANOVA run on the three theft scales found a significant difference by gender, with $F(1,109) = 2.86$, $p < .04$. As can be seen below in Table 6, female means were consistently lower than male means. Univariate tests showed those differences were significant only for the London House and Tesco Survey scales. This is a possible indication of a lower tendency toward theft for females since a MANOVA revealed no significant difference on the faking scales by gender, with $F(1,113) = .99$, ns.

TABLE 6

Sex Differences on Theft Scales

Theft Scale	Female Mean	Female SD	Male Mean	Male SD
Tesco Survey	15.49	3.01	16.64	3.90
London House	78.60	16.33	85.01	18.39
Reid Report	37.55	5.49	38.10	5.40

A MANOVA did reveal a significant order effect on the three theft scales with $F(2,104.5) = 7.69$, $p < .0001$. Table 7 (below) shows the three orders used in this experiment. Inspection of the means for each theft scale reveals that the London House theft scale scores are significantly higher when these items are not preceded by other scales (i.e., order 3). Apparently, when subjects are exposed to either the Tesco Survey or Reid Report items first, a learning effect occurs. This learning effect creates a

bias in responding to the London House items in such a way as to produce lower, more socially desirable scale scores.

TABLE 7

Means and Standard Deviations
for Theft Scales by Test Order

Order	Test	Mean	SD
1	Tescor Survey	16.57	3.32
	London House	79.64	16.37
	Reid Report	37.30	5.37
2	Tescor Survey	15.53	3.59
	Reid Report	37.54	5.86
	London House	75.58	18.02
3	London House	87.46	15.81
	Reid Report	38.48	5.07
	Tescor Survey	15.72	3.33

DISCUSSION

The intercorrelations between the integrity tests examined in this study are significant and are in the predicted directions, therefore Research Question 1 was supported. On the theft scales, for example, high positive correlations were found between the London House scale and both the Tescor Survey and the Reid Report. There are differences between the scales, however, and this is indicated by the more moderate correlation between the Tescor Survey and the Reid Report theft scales ($r = + .54$). More comparative studies need to be conducted in order to replicate this

difference and to determine if the other scales within each test are as highly intercorrelated.

Research Question 2 was also supported. Conscientiousness, as measured by the Work Orientation Scale of the CPI, was found here to be significantly correlated with three currently used theft scales. This supports the idea suggested by Ones et. al. (1993) that conscientiousness is a construct underlying all integrity tests. It seems reasonable, therefore, that the Wo scale could be used in place of integrity tests. While the Wo scale is empirically derived from the items on the CPI and the items contained in the Wo scale do not directly deal with employee theft, other integrity tests' items are often transparent and the questions very personal. The Wo scale could offer a more subtly phrased alternative.

In support of Research Question 3, this study helps further establish the fact that response distortion, or faking, is a problem in integrity testing. The problem was shown here to differ from test to test, from situation to situation, as well as from individual to individual. In regards to the tests, similar response distortion was seen on the Tesco Survey and the London House theft scales, and different distortion was seen on the Reid Report. Of the four instruction sets (Fake Good, Applicant, Prevent Theft, and High Integrity), distortions were most pronounced on the Tesco Survey and London House tests when test takers were given instructions to take the test as a job applicant and that the prospective employer wanted to stop a past problem with theft by not hiring potential thieves. This shows that the mention of theft and knowledge of it's undesirability to the prospective employer influenced the respondents to answer questions in a

more socially acceptable manner. This illuminates the differences by situation. Results differed on The Reid Report theft scale. Here, subjects responded with more socially desirable answers when told to do so, or to “Fake Good.” This indicates that The Reid Report is more easily faked, but also less susceptible to unintentional bias induced by instructions.

Individual differences in response distortion came in two forms. The first difference is evident in the rationale for testing attributed by test takers. Regardless of instruction set, the most desirable answers came from those subjects who believed that the test giver was interested in weeding out “undesirables.” The second difference came in the form of gender. Publishers may be correctly claiming that their tests do not present a threat of “adverse impact” by gender, but the current study did find a significant difference between the responses of the two genders on two of the three theft scales studied. Females tend to have scores indicating a lower tendency toward theft.

The results of this study have practical implications for the integrity testing industry. First of all, those entrusted with the administration of such tests must be made aware of the impact of situational factors on test takers. Greater steps need to be made toward controlling demand characteristics of the situation and experimenter bias. This includes standardizing testing instructions and environment in order to reduce the risk of response distortion. Test distributors should consider screening those purchasing the tests for their qualifications to be administering psychological tests. A second implication of this study has to do with the differences between genders. Even though tests have been shown not to

create “adverse impact” for one gender or the other, significant differences do exist between scale scores of males and females. It would be worth finding out why those differences exist.

Yet another practical implication of this study has to do with the personality characteristic of conscientiousness. It has been shown here that conscientiousness is indeed correlated with currently used theft scales. The Work Orientation Scale of the CPI, a short and unobtrusive test, appears to be an alternative to the numerous integrity tests on the market. The task now is to subject the Wo scale to scrutiny to determine if it has the desired effect when implemented in a business environment.

The integrity testing industry can also benefit from the theoretical implications of this study. For example, the data indicates that there are converging correlations between tests, suggesting that there is a psychological construct of “tendency toward theft.” And, this construct can be measured in both transparent and nontransparent ways. Also, the construct of conscientiousness has emerged as a useful predictor. Not only does it correlate with desirable behaviors, such as work quality, but it also is negatively correlated with the undesirable behavior of theft and can be used to predict it as well. Another implication relates to response distortion, or faking. The data shows that response distortion is influenced by pre-held beliefs about the reasons for test administration. This has not previously been taken into account. In addition, instructions that mention theft specifically tend to cue the test taker to pay special attention to items related to theft. Therefore, the theft items, for two of the three integrity tests, were answered in the most socially desirable way.

A limitation of the present study is that order effects were found to play a significant role. The problem cannot meaningfully be addressed here, however, since only three orders of the many possible were used. Only the London House test was significantly affected. It appears that prior inquiry about theft lowered the subsequent theft scale score on the London House test. Further studies will be needed to determine the significance of this finding.

The present research has laid the groundwork for much study that needs to be conducted in the area of integrity testing. In addition to investigations into the utility of the construct of conscientiousness in predicting counter-productive behavior, and finding out why gender differences exist in integrity test results, there are other directions research could take as well. For example, it would be of interest to see how subjects that were given no instructions, then grouped by what they assumed to be the reason for testing, responded to the same integrity tests used in this research. Also, a comparative study on a larger scale, using greater numbers of the existing integrity tests would be very informative. Finally, it would be fruitful to search for other methods of identifying people that commit counter-productive acts.

References

- American Tesco, Inc. (1985). Validation Manual - Tesco Survey. Portland, OR: Author.
- American Tesco, Inc. (1994). Validation Research Report. Portland, OR: Author.
- Association of Personnel Test Publishers. (1990). Model Guidelines for Pre-employment Integrity Testing Programs. Washington, DC: Author.
- Brodsky, S.L. (1978). A review of the Reid Report. In Buros O.C. (Ed.), 8th Mental Measurement Yearbook (p. 658). Lincoln, NE: Buros Institute of Mental Measurements.
- Brown, T.S., Jones, J.W., Terris, W., & Steffy, B.D. (1987). The impact of pre-employment integrity testing on employee turnover and inventory shrinkage losses. Journal of Business and Psychology, 2(2), 36-149.
- Bullard, P. (1992). Tesco Survey Validation. Portland, OR: American Tesco, Inc.
- Camara, W.J., & Schneider, D.L. (1994). Integrity tests: Facts and unresolved issues. American Psychologist, 49(2), 112-119.
- Chatman, J.A. (1989). Improving interactional organizational research: A model of person-organization fit. Academy of Management Review, 14(3), 333-349.
- Dunnette, M.D., McCartney, J., Carlson, H.C., & Kirchner, W.K. (1962). A study of faking behavior on a forced-choice self-description checklist. Personnel Psychology, 15, 13-24.

- Dyer, C.O. (1985). Jackson Personality Inventory. Test Critiques, II, 369-375.
- Goldberg, L.R., Grenier, J.R., Guion, 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. (1985). A work orientation scale for the California Psychological Inventory. Journal of Applied Psychology, 70(3), 505-513.
- Hefter, R. (1986). The crippling crime. Security World, 23, 36-38.
- 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 on those validities. Journal of Applied Psychology, 75(5), 581-595.
- House of Representatives. Employee Polygraph Protection Act of 1988. 100th Congress, 2nd Session. Report 100-659, pp. 1-16.
- Jackson, D.N. (1976). Jackson Personality Inventory Manual. Port Huron, MI: Research Psychologists Press.
- Jones, J.W., Slora, K.B., & Boye, M.W. (1990). Theft reduction through personnel selection: A control group design in the supermarket industry. Journal of Business and Psychology, 5(2), 275-279.
- Jones, J.W., & Terris, W. (1991). Integrity testing for personnel selection: An overview. Special Issue: Integrity Testing. Forensic Reports, 4(2), 117-140.

- London House Press, Inc. (1980). The Personnel Selection Inventory (PSI). Park Ridge, IL.
- Myers, A. & Hansen, C. (1993). Experimental Psychology. Belmont, CA: Brooks/Cole Publishing Company.
- Nicholson, R.A., & Hogan, R. (1990). The construct validity of social desirability. American Psychologist, 45, 290-292.
- Ones, D.S., Viswesvaran, C., & Schmidt, F.L. (1993). Comprehensive meta-analysis of integrity test validities: Findings and implications for personnel selection and theories of job performance. Journal of Applied Psychology, 78:4, 679-703.
- Reid, J.E. (1967). The Reid Report. Chicago: John E. Reid and Associates.
- Ryan, A.M., & Sackett, P.R. (1987). Pre-employment honesty testing: Fakability, reactions of test takers, and company image. Journal of Business and Psychology, 1(3), 248-256.
- Sackett, P.R., Burris, L.R., & Callahan, C. (1989). Integrity testing for personnel selection: An update. Personnel Psychology, 42, 491-527.
- Sackett, P.R., & Harris, M.M. (1984). Honesty testing for personnel selection: A review and critique. Personnel Psychology, 37, 221-245.
- Sauser, W.I. (1985). Review of the London House Personnel Selection Inventory. In Mitchell J.V. (Ed.), 9th Mental Measurements Yearbook (pp. 870-871). Lincoln, NE: Buros Institute of Mental Measurements.
- Schneider, B. (1983). Interactional psychology and organizational behavior. In B.M. Staw & L.L. Cummings (Eds.), Research In

Organizational Behavior (Vol. 5, pp. 1-31), Greenwich, CT: JAI Press.

Terris, W. (1979). Attitudinal correlates of employee integrity: Theft-related admissions made in pre-employment polygraph examinations. Journal of Security Administration, 2, 30-39.

Terris, W., & Jones, J.W. (1982). Psychological factors related to employees' theft in the convenience store industry. Psychological Reports, 51, 1219-1238.

Willis, C.G. (1985). Review of the Reid Report/Reid Survey. Test Critiques, Vol. II., p. 631-636.