College of Business Administration
Department of Management and Organizations
University of Iowa
Iowa City, IA 52242

FINAL REPORT

# Meta-Analysis of Integrity Tests: A Critical Examination of Validity Generalization and Moderator Variables

by

**Denise S. Ones**
**Chockalingam Viswesvaran**
**Frank L. Schmidt**



**June 1992**

*Prepared for:*

**Defense Personnel Security Research Center**

*Under the*

**Office of Naval Research**
**N00014-91-J-4168**

*Approved for Public Distribution: Distribution Unlimited*

SECURITY CLASSIFICATION OF THIS PAGE

# REPORT DOCUMENTATION PAGE

Form Approved
OMB No 0704-0188

| | |
|---|---|
| 1a REPORT SECURITY CLASSIFICATION | 1b RESTRICTIVE MARKINGS |
| 2a SECURITY CLASSIFICATION AUTHORITY | 3 DISTRIBUTION / AVAILABILITY OF REPORT |
| 2b DECLASSIFICATION / DOWNGRADING SCHEDULE | Unlimited |
| 4 PERFORMING ORGANIZATION REPORT NUMBER(S)<br>92-1 | 5 MONITORING ORGANIZATION REPORT NUMBER(S) |
| 6a NAME OF PERFORMING ORGANIZATION<br>University of Iowa<br>6b OFFICE SYMBOL (If applicable) | 7a NAME OF MONITORING ORGANIZATION<br>Defense Personnel Security Research and Education Center |
| 6c. ADDRESS (City, State, and ZIP Code)<br>Iowa City, Iowa 52242 | 7b ADDRESS (City, State, and ZIP Code)<br>990 Pacific St., Suite 455-E<br>Monterey, CA 93940 |
| 8a. NAME OF FUNDING / SPONSORING ORGANIZATION Defense Personnel Security Research & Education Center<br>8b OFFICE SYMBOL (If applicable) | 9 PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER<br>N00014-91-J-4168 |
| 8c. ADDRESS (City, State, and ZIP Code)<br>99 Pacific St., Suite 455-E<br>Monterey, CA 93940 | 10 SOURCE OF FUNDING NUMBERS<br>PROGRAM ELEMENT NO / PROJECT NO 34394 / TASK NO / WORK UNIT ACCESSION NO |

11 TITLE (Include Security Classification)
Meta-Analysis of Integrity Test Validities

12. PERSONAL AUTHOR(S)
Ones, D.S., Viswesvaran, C., Schmidt, F.

| 13a. TYPE OF REPORT | 13b TIME COVERED | 14 DATE OF REPORT (Year, Month, Day) | 15 PAGE COUNT |
|---|---|---|---|
| Technical | FROM 9-28-91 TO 6-30-92 | 1992 June | 77 |

16 SUPPLEMENTARY NOTATION
This report was prepared under an Office of Naval Research Contract

| 17 COSATI CODES | | | 18 SUBJECT TERMS (Continue on reverse if necessary and identify by block number) |
|---|---|---|---|
| FIELD | GROUP | SUB-GROUP | Personnel selection; personnel integrity & reliability; personnel security; meta-analysis; selection validity |
| | | | |
| | | | |

19 ABSTRACT (Continue on reverse if necessary and identify by block number)

A comprehensive meta-analysis was conducted to investigate whether integrity test validities are generalizable and to estimate differences in validity due to potential moderating influences. The database included 665 validity coefficients across 576,464 data points. Results indicate that integrity test validities are positive and in many cases substantial for predicting both job performance and counterproductive behaviors on the job such as theft, disciplinary problems, and absenteeism. Validities were found to be generalizable. The estimated mean operational predictive validity of integrity tests for supervisory ratings of job performance is .41. For the criterion of counterproductive behaviors, results indicate that use of concurrent validation study designs may overestimate the predictive criterion-related validity applicable in selection situations. Our results based on external criterion measures (i.e., excluding self reports) and predictive validity studies using applicants indicate that integrity tests

| | |
|---|---|
| 20 DISTRIBUTION / AVAILABILITY OF ABSTRACT<br>☒ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT ☐ DTIC USERS | 21 ABSTRACT SECURITY CLASSIFICATION<br>UNCLASSIFIED |
| 22a NAME OF RESPONSIBLE INDIVIDUAL<br>Ralph Carney, Ph.D. | 22c OFFICE SYMBOL<br>PERSEREC |

DD Form 1473, JUN 86 *Previous editions are obsolete*

S/N 0102-LF-014-6603

SECURITY CLASSIFICATION OF THIS PAGE

UNCLASSIFIED

predict the broad criterion of organizationally disruptive behaviors better than they predict the narrower criterion of employee theft alone. Our results also indicated substantial evidence for the construct validity of integrity tests. Perhaps the most important conclusion of this research is that despite the influence of moderators, integrity test validities are positive across situations and settings.

Abstract

A comprehensive meta-analysis was conducted to investigate whether integrity test validities are generalizable and to estimate differences in validity due to potential moderating influences. The database included 665 validity coefficients across 576,464 data points. Results indicate that integrity test validities are positive and in many cases substantial for predicting both job performance and counterproductive behaviors on the job such as theft, disciplinary problems, and absenteeism. Validities were found to be generalizable. The estimated mean operational predictive validity of integrity tests for supervisory ratings of job performance is .41. For the criterion of counterproductive behaviors, results indicate that use of concurrent validation study designs may overestimate the predictive criterion-related validity applicable in selection situations. Our results based on external criterion measures (i.e., excluding self reports) and predictive validity studies using applicants indicate that integrity tests predict the broad criterion of organizationally disruptive behaviors better than they predict the narrower criterion of employee theft alone. Our results also indicated substantial evidence for the construct validity of integrity tests. Perhaps the most important conclusion of this research is that despite the influence of moderators, integrity test validities are positive across situations and settings.

## Meta-Analysis of Integrity Test Validities

Over the last ten years, interest in and use of integrity testing has increased substantially. The publication of a series of literature reviews attests to the interest in this area and its dynamic nature (Guastello & Rieke, 1991; Sackett, Burris, & Callahan, 1989; Sackett & Decker, 1979; Sackett & Harris, 1984). Recently Sackett et al. (1989) and O'Bannon, Appleby, and Goldinger (1989) have provided extensive qualitative reviews and critical observations regarding integrity testing. In addition to these reviews, the US Congressional Office of Technology Assessment (OTA) (1990) and the American Psychological Association (APA) (Goldberg, Grenier, Guion, Sechrest, & Wing, 1991) have each released reports on integrity tests. The OTA report (1990) is short and somewhat superficial. The APA report (Goldberg et al., 1991) is more through and provides a generally favorable conclusion regarding the use of paper and pencil integrity tests in personnel selection. The aim of this paper is not to provide a qualitative overview, but to seek quantified answers to questions raised in these earlier reviews, and to test hypotheses that will help researchers and practitioners make sense of the validities of integrity tests.

The three meta-analyses that have previously been reported have each focused on a single integrity test. The first (Harris, undated) investigated the validity of the Stanton Survey. The second meta-analysis (McDaniel & Jones, 1986) examined the validity of the London House Employee Attitude Survey (London House, 1982). Lastly, McDaniel and Jones (1988) focused on the dishonesty scale of the Personnel Selection Inventory (PSI) (London House, 1980) in predicting employee theft. However, to date no comprehensive meta-

analysis of the validities of all integrity tests has been reported. The hypothesis that each test-criterion combination is unique and must be analyzed separately seems to have been implicitly assumed by the researchers in this field. One aim of this meta-analysis is to test this hypothesis and provide the required empirical evidence to confirm or refute the notion that validity is specific to particular types of instruments, criteria, or validation strategies (concurrent or predictive). That is, one purpose of this study is to use meta-analysis to investigate whether integrity test validities are generalizable across jobs, criteria, and tests, and to quantitatively document validity differences that may be due to moderating influences.

Sackett et al. (1989) classify honesty tests into two categories: "overt integrity tests" and "personality-based tests." Overt integrity tests (also known as clear purpose tests) are designed to directly assess attitudes regarding dishonest behaviors. Some overt tests specifically ask about past illegal and dishonest activities as well; although for several admissions are not a part of the instrument, but instead are used as the criterion. Overt integrity tests include the London House Personnel Selection Inventory (PSI) (London House Inc., 1975), Employee Attitude Inventory (EAI) (London House Inc., 1982), Stanton Survey (Klump, 1964), Reid Report (Reid Psychological Systems, 1951), Phase II Profile (Lousig-Nont, 1987), Milby Profile (Miller & Bradley, 1975), and Trustworthiness Attitude Survey (Cormack & Strand, 1970). According to Sackett et al. (1989), "...the underpinnings of all these tests are very similar..." (p. 493). Hence, they predict high correlations among all these overt integrity measures. On the other hand, personality-based measures (also

referred to as disguised purpose tests) aim to predict a broad range of counterproductive behaviors at work (e.g., disciplinary problems, violence on the job, excessive absenteeism and tardiness, drug abuse, in addition to theft) via personality dimensions, such as reliability, conscientiousness, adjustment, trustworthiness, and sociability. Personality-based measures have not been developed solely to predict theft or theft-related behaviors. Examples of personality-based measures that have been used in integrity testing include the Personal Outlook Inventory (Science Research Associates, 1983), the Personnel Reaction Blank (Gough, 1954), Employment Inventory of Personnel Decisions Inc. (Pajaanen, 1985), and the Hogan's Reliability Scale (Hogan, 1981). The similarity of these measures raises the question of whether they all measure primarily a single general construct. Different test publishers claim that their personality-based integrity tests measure different constructs, including responsibility, long term job commitment, consistency, proneness to violence, moral reasoning, hostility, work ethics, dependability, depression, and energy level (O'Bannon et al., 1989). Given the descriptions of these claimed constructs, we believe these tests may all measure the general construct of broadly defined "conscientiousness", one of the five dimensions of personality studied by Barrick and Mount (1991) (see also Digman (1990) and Goldberg (1990)). Conscientiousness reflects characteristics such as dependability, carefulness, and responsibility. In the integrity testing literature, this construct has been viewed from its negative pole (e.g., irresponsibility, carelessness, violation of rules). Inspection of items on several integrity tests confirms this notion. Therefore, we would anticipate high correlations among the

personality-based integrity tests. Detailed descriptions of all the above tests can be found in the 10th Measurement Yearbook (Conoley & Kramer, 1989) and/or in the extensive reviews of this literature (O'Bannon et al., 1989; Sackett et al., 1989; Sackett & Harris, 1984). Table 1 lists the integrity measures which contributed data to the analyses reported in this research.

- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -

Insert Table 1 About Here

- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -

Many researchers point to the diversity and the deficiencies of the criteria used in validation of integrity tests (McDaniel & Jones, 1986, 1988; Sackett & Harris, 1984). For the reasons enumerated in the most recent review on integrity testing (Sackett et al., 1989), correlations with the polygraph results, organizational level reductions in counterproductive behaviors (e.g., reductions in inventory losses due to theft) after an integrity test is introduced for personnel selection, and comparisons of criminal with noncriminal samples do not alone produce convincing evidence for the criterion-related validity of integrity tests in selection settings. Rather, findings of this sort are evidence of construct validity (Goldberg et al., 1991). The criteria of interest in integrity testing can be categorized into overall job performance and counterproductive behaviors on the job. In this research, Study 1 (described later) investigated criteria of overall job performance, while Study 2 examined criteria of counterproductive behaviors.

Counterproductive behaviors criteria can be classified into two categories. The first group includes actual theft, theft admissions, and dismissals for actual

theft. This category has been termed "narrow criteria" by Sackett et al. (1989). As opposed to narrow criteria, validation studies can use broad criteria of counterproductivity which usually consist of composite indexes of such behaviors as disciplinary problems, excessive tardiness and absenteeism, turnover, violence on the job, substance abuse, property damage, organizational rule breaking, theft, and other disruptive or irresponsible behaviors.

From a methodological perspective, the criteria can further be divided into external and self-report (admissions) criteria (Sackett et al., 1989). Lending support to this categorization are the meta-analysis results of McDaniel and Jones (1988) showing that the validity of the PSI is moderated by this distinction in criterion measurement method. In the external criteria category are all actual records of rule breaking incidents, disciplinary actions, supervisory ratings of disruptiveness, dismissals for theft, and so on. On the other hand, the self-report criteria include all admissions of theft, past illegal, and counterproductive behaviors.

If all integrity tests measure an overall general construct (Sackett et al., 1989, p. 493), then integrity test validities will generalize across different predictor measures. That is, all integrity tests will have at least moderate positive levels of validity, lending them some potential utility in personnel selection. If validity generalization results across all integrity tests show substantial variability in validities after correction for the effects of statistical artifacts, then potential influences of moderating variables on the validities will be explored. The proposed moderators of integrity test validities for predicting job performance are enumerated in Table 2.

Insert Table 2 About Here

The first set of proposed analyses involves examining the validities of overt integrity tests and personality-based tests separately (proposed analysis 1 Table 2). Currently, there is only one study in the literature comparing the effectiveness of an overt integrity test and a personality-based integrity test (Rafilson & Frost, 1989). O'Bannon et al. (1989, p. 29) state that "Until additional research is conducted, it is not possible to conclude superiority of one type of test over the other".

If the classification of the predictors into overt vs. personality-based categories is not found to explain sizable portions of the variance in the validities, then criteria characteristics can be explored as moderators. In traditional validation studies, the criterion of job performance has usually been measured via supervisory ratings. Another method of measuring job performance is via organizational production records. There is some evidence that the two methods of measuring worker performance are not exactly equivalent (Campbell, McHenry, & Wise, 1990; Nathan & Alexander, 1988). Specifically, recent research evidence on the construct of job performance indicates that supervisors take into consideration many factors when rating employees, including organizational citizenship behaviors in addition to the output or productivity of the employee (Borman, White, Pulakos, & Oppler, 1991; Orr, Sackett, & Mercer, 1989). The moderator analysis of job performance measurement method (supervisory ratings vs. production records)

will test the hypothesis that supervisory ratings of job performance lead to estimates of integrity test validities similar to those obtained using production records as criteria (proposed moderator analysis 2 in Table 2).

For the criterion of counterproductive behaviors on the job, we expect the measurement method used for criteria to moderate validity (proposed analyses 3 in Table 2). Because all thieves are not caught, or all illegal activities detected, lower correlations are expected with external criteria. But, if respondents provide socially desirable responses, the effect could be to depress the correlations based on self-report criteria relative to external criteria (because of decreased construct validity in self-reports of counterproductive behaviors). The present research cannot determine the extent to which the validities using external criteria are artificially depressed because of failure to detect theft, or the extent to which the validities using self-report criteria are artificially reduced because of social desirability bias. In the light of the results of an earlier meta-analysis (McDaniel & Jones, 1988), we hypothesize that the validity will be higher for self-report measures than for external criteria.

For the criterion of counterproductivity, the breadth of criteria can also be explored as a potential moderator (proposed analysis 4 in Table 2). For this purpose, narrow criteria (i.e., theft) can be analyzed separately from broad criteria (i.e., general disruptive, rule-breaking behaviors). It has been hypothesized that the validity of overt integrity tests in predicting theft (narrow criteria) will be greater than the validity of personality-based integrity tests with the same criterion because, "..conceptually, one might argue that when one's interest is in predicting a narrow theft criterion, the narrower overt integrity

tests are more appropriate..." (Sackett et al., 1989, p. 494). That is, they hypothesize that narrowly defined criteria such as theft might be predicted better by narrowly focused predictors. For example, "...tennis performance is better predicted by tennis ability than by general athletic ability" (Buss, 1989, p. 1385). In contrast, personality-based integrity tests may produce higher validity with broadly defined disruptiveness criteria than with theft (narrow criteria), because broader personality-based integrity tests measure a variety of attitudes, behaviors, and tendencies, and therefore might predict a broader range of behaviors better.

There are three other potential moderators that merit investigation. The first is the question of whether concurrent validities accurately estimate predictive validities (proposed analyses 5 in Table 2). In the ability and aptitude domain, concurrent validities have been found to accurately estimate predictive validities (Bemis, 1968; Society for Industrial and Organizational Psychology, 1987), but this question has not been systematically examined for integrity tests.

Another potential moderator of integrity test validities is the validation sample (proposed moderator 6 in Table 2). Two distinct groups have been used in validity research: applicants to jobs and current employees. In selection settings, the group of focal interest is applicants. The purpose of criterion-related validity studies in employment is to estimate the validity of the selection instrument when used to select applicants. Furthermore, one traditional criticism of personality related predictors (similar to integrity tests) has been the problem of potential response distortion. By examining the validities of

integrity tests for employee and applicant groups separately, it can be determined whether applicant responses result in validities comparable to validities obtained on employees.

Finally, another potential moderator of integrity test validities is the complexity of the jobs for which the validation has been conducted (proposed analyses 7 in Table 2). The moderating influences of job complexity on general mental ability test validities in predicting job performance is well established (Hunter & Hunter, 1984). For general ability tests, as the level of job complexity increases, the validities also increase. However, the opposite effect may hold for integrity test validities. It could be hypothesized that as the level of job complexity increases, estimated validities of integrity tests would systematically decline because of more successful dissimulation by incumbents and applicants for high complexity jobs, and/or because of greater difficulty in detecting dishonest behaviors in these jobs. The former would produce smaller actual validities, while the latter would bias validity estimates downward while not affecting true (operational) validities.

The proposed moderating effects enumerated in Table 2 for job performance and for counterproductive job behaviors could co-vary. Potential confounding of moderator variable effects could exist if, for example, most self-report criteria were also narrow criteria. The identification of the potentially confounded moderator effects involves the examination of the proposed moderators simultaneously. Availability of validities in each category may preclude an analysis of all combinations. However, to the extent feasible, we propose to

conduct a fully hierarchical moderator analysis (Hunter & Schmidt, 1990a, p. 527).

## Method

### Description of the Database

A massive search was conducted to locate all existing integrity test validities. All published empirical studies were obtained from published reviews of the literature (O'Bannon et al., 1989; Sackett et al., 1989; Sackett & Harris, 1984), the three other meta-analyses of integrity tests (Harris, undated; McDaniel & Jones, 1986, 1988), and a computerized search to locate the most recent studies in psychology and management related journals. According to O'Bannon et al. (1989), there are forty three integrity tests in use in the United States. All the publishers and authors of the forty three tests were contacted by telephone or in writing requesting validity, reliability, and range restriction information on their tests. In addition, we identified other integrity tests overlooked by O'Bannon et al. (1989); their publishers were also contacted. All the available unpublished technical reports reporting validities, reliabilities, or range restriction information were obtained from integrity test publishers and authors. Some integrity test authors and test publishers responded to our request for validity information on their test by sending us computer printouts that had not been written up as technical reports. These were included in the database.

We computed 126 validities using data sent by integrity test publishers or authors. These 126 validities included 122 cases where no correlations were reported, but using the information supplied we were able to calculate the phi

correlation, and then correct it for dichotomization (Hunter & Schmidt, 1990b). The corrected correlations were used in the meta-analysis. Sample sizes for these corrected correlations were adjusted to avoid underestimating the sampling error variance. First, the uncorrected correlation and the study sample size were used to estimate the sampling error variance for the observed correlation. This value was corrected for the effects of the dichotomization correction, and this corrected sampling error variance was then used with the uncorrected correlation in the standard sampling error formula to solve for the adjusted sample size, which was entered into the meta-analysis computer program. This process results in the correct estimate of the sampling error variance of the corrected correlation in the meta-analysis.

A total of 665 criterion-related validity coefficients contributed to the database. The total sample size across 665 validities was 576,464. For this meta-analysis over 700 pieces of literature and personal communications were reviewed. The validity data used in the analyses came from over 180 studies, technical reports and personal communications. A list of studies relevant to this meta-analysis can be obtained from Deniz Ones. Of the 665 validity estimates, 247 validities came from the published literature or the published reviews of integrity tests. To address the concern that there could be some kind of systematic difference in validities from the published sources compared to unpublished sources, we computed the correlation between the validity coefficients reported and the dichotomous variable of published vs. unpublished studies. This correlation was -.02. The negative sign of the correlation indicates that published studies reported negligibly higher validities. Hence in

our database, the published vs. unpublished distinction for the validities is trivial and inconsequential. The list of integrity tests contributing criterion-related validity coefficients, reliabilities, or range restriction information to this meta-analysis is presented in Table 1.

The 665 validities and other information were independently coded. For each validity coefficient predictor and criterion information, validation strategy, and validation sample information were coded. Across all coded validity coefficients, there was 89% full agreement. In coding 73 validities out of 665, there was at least one item of disagreement among all the pieces of information coded. Most of the disagreements between the coders resulted from vague reporting of information in technical reports and other unpublished sources. To resolve each disagreement, the test publishers were contacted to inquire about the item of disagreement. In 64 of the 73 disagreements, the new data obtained from the test publisher resolved the disagreement. In the 9 cases where even the test publisher did not have further information, the item of information in dispute was coded as missing.

The final database of 665 validities across 576,464 data points included 389 validities from overt integrity tests and 276 validities from personality-based integrity tests. Most of the validities came from service industries ($k = 503$), most notably from the retail industry (i.e., discount chains, department stores, supermarkets, grocery chains, convenience stores, drug stores). The increasing service orientation of the US Economy (Hudson Institute, 1987) makes the results of this meta-analysis more relevant. The validities were reported on a diverse range of occupations, including some from high complexity

jobs. Finally, of the 665 validities, 222 had job performance as the criterion and 443 had counterproductive behaviors as the criterion.

Artifact Distributions

Several sets of artifact distributions were compiled: 3 distributions for the reliability of the integrity tests, 4 distributions for the reliability of the criterion variables, and 1 distribution for range restriction. Descriptive information on the artifact distributions are provided in Table 3.

- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -

Insert Table 3 about here

- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -

A total of 124 integrity test reliability values were obtained from the published literature and the test publishers. The overall mean of the predictor reliability artifact distribution was .81 and the standard deviation was .11. The mean of the square roots of predictor reliabilities was .90 with a standard deviation of .06. Two other predictor reliability distributions were constructed: one for overt integrity tests and another for personality-based integrity tests. There were 97 reliabilities reported for overt tests. The mean of the overt test reliability artifact distribution was .83 and the standard deviation was .09. The mean of the square roots of overt test reliabilities was .91 with a standard deviation of .05. There were 27 reliabilities reported for personality-based tests. The mean of the personality-based test reliability artifact distribution was .72 and the standard deviation was .13. The mean of the square roots of the reliabilities was .85 with a standard deviation of .08. Each one of these predictor

reliability distributions were used in analyses with corresponding predictor categories.

Reliability estimates for the criterion variables were taken from the studies that contributed to the database for this meta-analysis and the published literature on counterproductivity and job performance. Four separate distributions were created, one each for: job performance, production records, supervisory ratings of job performance, and counterproductive behaviors on the job. The mean reliability values used in the corrections for criterion reliabilities are as follows: .54 for job performance (supervisory ratings and production records combined), .89 for production records, .52 for supervisory ratings of job performance (Rothstein, 1990); .69 for overall counterproductive behaviors. The mean criterion reliability for job performance represents the combination of supervisory ratings of overall job performance and production records. The reliability of supervisory ratings of overall job performance of .52 was assigned a frequency of 153 to match the number of validities for that criterion in our database and was combined with 10 reliabilities for production records to comprise the distribution of job performance reliabilities. The reliability of production records was obtained from Hunter, Schmidt, and Judiesch (1990) as .55 for a one week period. Using the Spearman-Brown formula, this value was adjusted to the appropriate time period in each study reporting validities for production records. There were 13 unique reliabilities reported for counterproductive behaviors. The mean reliability for externally measured counterproductive behaviors was similar to the mean reliability of admissions of counterproductivity. Each of the

reliabilities was assigned a frequency corresponding to the number of validities in the database using the criterion category for which the reliability was reported. There were no reliabilities reported for externally detected theft. The mean reliability for the distribution of counterproductive behaviors was .69.

Because integrity tests are used to screen applicants, the validity calculated using an employee sample may be affected by restriction in range. Also, dishonest employees may be terminated, creating a second source of range restriction. A distribution of range restriction values was constructed from the studies contributing to the database. There were 75 studies which reported both the study sample standard deviation and the applicant group standard deviation. The range restriction ratio was calculated as the ratio of study to reference group standard deviations (s/S). In four studies, correlations were reported for both the applicant and the employee groups. From these four studies range restriction ratios were calculated by taking the ratio of the two correlations reported and solving for the range restriction value using the standard range restriction formula (Case II formula; Thorndike, 1949, p. 173). Overall there were 79 range restriction values included in the artifact distribution. The mean ratio of the restricted sample's standard deviation to the unrestricted sample's standard deviation used is .81 and the standard deviation is .19. The mean of .81 indicates there is considerably less range restriction in this research domain than is the case for cognitive ability (Alexander, Carson, Alliger, & Cronshaw, 1989). Thus, range restriction corrections were much smaller in present research than in meta-analyses in the abilities domain.

Meta-Analytic Procedures

The hypotheses in this paper are tested using the Hunter-Schmidt (1990a, p. 185) psychometric meta-analytic procedure. Psychometric meta-analysis is a statistical technique used ( among other purposes) to estimate how much of the observed variance of findings across studies results from statistical artifacts. The artifact distributions described above were used to correct biases in the observed validities caused by statistical artifacts. The artifacts operating across studies include sampling error, unreliability in the predictor and the criterion, range restriction, dichotomization of variables, and so on. If the validity is strongly dependent on the situation or on moderators, statistical artifacts will not account for all or nearly all of the observed variation in the validities, and/or the standard deviation of the true validities will be relatively large. In addition to estimating the portion of the observed variance that is due to statistical artifacts, meta-analysis also provides the most accurate obtainable estimate of the mean true validity. In this study, the interactive meta-analysis procedure was used (Hunter & Schmidt, 1990a, p.165; Schmidt, Hunter, & Gast-Rosenberg, 1980). The program used incorporated refinements shown by computer simulation studies to increase accuracy (Law, Schmidt, & Hunter, 1992). These refinements include use of the mean observed correlation in the formula for sampling error variance and use of a nonlinear range restriction correction formula to estimate the standard deviation of true validities.

If all or a major portion of the observed variance in validities is due to statistical artifacts, one can conclude that the validities are constant or nearly so. If the 90% credibility value is greater than zero, indicating that 90% of the

estimates of true validity lie above that value, one can conclude that the presence of validity can be generalized to new situations (Hunter & Schmidt, 1990a). The lower credibility value is dependent on variance remaining after correction for statistical artifacts. In a meta-analysis, if the 90% credibility value is greater than zero, but there is a sizable variance in the validities after corrections, it can be concluded that validities are positive across situations, although the actual magnitude may vary across settings. However, the remaining variability may also be due to uncorrected statistical artifacts as well as methodological differences between studies. A final possibility is truly situationally specific test validities and/or the operation of moderator variables. In sum, the 90% credibility value is used to judge whether the validities are positive across situations (i.e., validity generalizes), while the variance accounted for by statistical artifacts and the estimated standard deviation of true validities are used to assess the moderating influences of the hypothesized factors.

The correlations cumulated cover a diverse range of occupations and organizations. Most of the studies on each integrity test were conducted on independent samples. Where more than one correlation was available on a single sample for the same criterion, the validities were averaged to avoid violations of the independence assumption (Hunter & Schmidt, 1990a, pp. 452-454). The sample size used was the average sample size.

The meta-analyses corrected the mean observed validity for mean attenuation due to criterion unreliability and range restriction (Hunter & Schmidt, 1990a, p. 165). No correction for predictor unreliability was applied to the mean validity because our interest was in estimating the operational

validities of integrity tests for selection purposes. However, the observed variance of validities was corrected for variation in predictor unreliabilities in addition to variation in criterion unreliabilities and range restriction values. For comparison purposes, we provide the percent variance due to sampling error alone in our results. Furthermore, mean observed validities without any artifact corrections are presented.[1]

## Analyses and Results

Table 4 summarizes the results of the meta-analyses conducted across all integrity test validities for predicting job performance and counterproductive behaviors.

- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -

Insert Table 4 about here

- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -

The first meta-analysis in Table 4 estimates the validity of all integrity tests combined, overt and personality-based, for predicting the criterion of overall job performance (Line 1 in Table 4). The total sample size across 222 studies reporting such a correlation was 63,500. This meta-analysis indicates that the proportion of the variance observed in validities due to statistical artifacts is 53%. The estimate of the mean operational validity of all integrity tests with the criterion of overall job performance is .34. The standard deviation of the true validity is .13. The 90% credibility value of .20 indicates that integrity test validities are positive across situations for the criterion of overall job performance.

The second meta-analysis was performed on the 443 correlations between integrity tests and counterproductive behaviors (Line 2 in Table 4). The 443 correlations were over a total sample size of 384,293 data points, and the criteria in this category included all measures of disruptive behaviors at work such as theft, illegal activities, absenteeism, tardiness, drug abuse, dismissals for theft, and violence on the job. Both self-report and external criteria were included. The lower 90% credibility value of .05 indicates that the validity of integrity tests as a group in predicting counterproductive behaviors is positive across situations. The mean operational validity for such tests is estimated at .47. For this category of integrity test validities the standard deviation of the true validity is .37, a fairly large value. In addition, sampling error, unreliability in the predictor, unreliability in the criteria, and range restriction combined account for only 9% of the variance observed in the correlations. These results indicate that all types of integrity tests are valid predictors of counterproductive behaviors. But the standard deviation of the true validity in analysis is large enough and the percent variance accounted for low enough to suggest that other statistical artifacts or potential moderators are operating. These results suggest that overall job performance and counterproductive behaviors on the job are not similarly predictable by integrity tests, confirming our decision to analyze validities for job performance and counterproductive behaviors separately.

Study 1: Analyses and Results for Predicting Job Performance

As is reported in Table 4, the mean operational validity of integrity tests in predicting overall job performance is .34. However, the SDρ of .13 and the

percent variance accounted for of 53% by all statistical artifacts we could correct for (i.e., sampling error, criterion and predictor unreliability, range restriction, and dichotomization) indicate that the validity may be moderated by other variables. The results of the moderator are analyses reported in Table 5.

- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -

Insert Table 5 About Here

- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -

The first potential moderator tested is the predictor type (overt vs. personality-based). The results across 84 validities and 27,768 data points (Line 1a in Table 5) show that the best estimate of overt integrity tests' validity in predicting overall job performance is .33. The worst case value of .16 indicates that the validity is positive across studies and situations. The percent variance accounted for by the corrected statistical artifacts is 40%, and the standard deviation of the true validity (SDρ) is .15. Personality-Based integrity tests show a mean validity of .35 (K = 138, N = 35,732) in predicting overall job performance, with 63% of the observed variance accounted for by the statistical artifacts we could correct for (Line 1b in Table 5). The SDρ for personality-based integrity tests was .11 and the lower credibility value was .23 indicating that the validities of personality-based integrity tests are also positive across studies and situations. These results suggest that test type is probably not a moderator of integrity test validities in predicting overall job performance; overt and personality-based integrity tests appear to have similar levels of operational validity when the criterion is job performance.

A second potential moderator of integrity tests validities, suggested by Nathan and Alexander (1988), is the criterion measurement method (supervisory ratings vs. production records). All available correlations between integrity tests and supervisory ratings of overall job performance were meta-analyzed. There were 153 such correlations obtained from a total sample size of 36,250 data points (Line 2a in Table 5). The operational validity of integrity tests in predicting supervisory ratings of job performance is .35. The worst case value is .20, indicating that the validity is positive across studies and situations. The percent variance accounted for by the corrected statistical artifacts is 55%, and the standard deviation of the true validity ($SD\rho$) is .13. For production records criteria, there were only 10 validities based on a total sample size of 2,210 (Line 2b in Table 5). The true validity for predicting production records is .28 and the standard deviation of true validity is .12. The lower credibility value and the percent variance accounted for by statistical artifacts are .15 and 47%, respectively. Although there were far more validities for supervisory ratings of overall job performance ($K = 153$) than for production records ($K = 10$), the meta-analytic results from these categories are somewhat similar (estimated true validities of .35 and .28, respectively). Therefore, we conclude that the criterion measurement method probably does not have large impact on integrity test validities in predicting job performance. This result mirrors the findings of Nathan and Alexander (1988) that studies using the criterion of supervisory ratings of job performance produce validity estimates similar to those from studies using production quantity as the criterion.

The third potential moderator studied is the validation strategy used in the primary studies. To determine whether concurrent validities estimate predictive validities accurately in this noncognitive domain, predictive and concurrent validities for predicting overall job performance were meta-analyzed separately (Lines 3a and 3b in Table 5). Predictive validities of integrity tests have mean true validity of .31, while concurrent studies have a mean true validity of .37 in predicting job performance. These results seem to suggest that concurrent validities of integrity tests may slightly overestimate predictive validities. However, in this set of analyses, there was one very large sample concurrent validation study contributing a validity coefficient much larger than the sample size weighted mean observed validity. In the concurrent validation moderator analysis the total sample size was 31,866 with a mean observed correlation of .22. This large sample concurrent study had a sample size of 9,819 and contributed an observed validity of .26 to the database. To counteract the potentially biasing effect of this one study, we calculated the unweighted mean observed validity for concurrent validities (unweighted mean r = .14). When the statistical artifact corrections were applied to the unweighted mean validity, the true validity obtained for the concurrent validation category was .23, a substantially smaller value than .37 (mean $\rho$ using the sample size weighted mean validity). In the analysis of predictive validities, there was also a very large sample validation study. However, the validity coefficient in this case was much smaller than the observed sample size weighted mean validity of the predictive validation category. In the predictive validation moderator analysis the total sample size was 30,150 with a mean observed correlation of .19. The

large sample predictive study had a sample size of 6,884 and contributed the observed validity of .15 to the database. To counteract the potentially biasing effect of this one study, we calculated the unweighted mean observed validity for predictive validities (unweighted mean r = .27). When the statistical artifact corrections were applied to the unweighted mean validity, the true validity obtained was .43, a substantially larger value than the .31 in Table 5. When the estimated true validities calculated using the unweighted mean validities are compared for the concurrent and predictive validation strategies, it seems that predictive validity ($\rho$ = .43) is almost twice as large as concurrent validity ($\rho$ = .23). This contradicts the conclusions reached using mean $\rho$s based on sample size weighted means. Because it cannot be determined in which set of analyses, if either, the large sample studies are biasing the results, the conclusion regarding the moderating influences of validation strategy on validities when the criterion is job performance is inconclusive. Other analyses reported in Study 2 of this paper will examine whether concurrent and predictive validities are similar for the other major criteria category, counterproductive behaviors. On a positive note, in both the concurrent and predictive validation categories the 90% credibility values indicate that validity of integrity tests for predicting job performance is positive (lower credibility values of .22 and .17, respectively).

The fourth potential moderator studied is the validation sample used in the studies (applicant sample vs. employee sample) (lines 4a and 4b in Table 5).This analysis is not redundant with the analysis of predictive vs. concurrent studies because there were some predictive studies conducted with employees (K = 63); in these studies, the criterion data were not gathered until a considerable

time after administration of the test. There was also one predictive study conducted on applicants using the criterion of supervisory rating of work sample performance. In selection settings, the optimal method for estimating operational selection validities is predictive validation based on applicants. Although the predictive validities of tests using employee samples can be informative, for personnel selection research that value is important only to the extent that it approximates the applicant sample validity. For studies using the criterion of overall job performance, the mean true validity estimate obtained using an applicant sample is .40. When employees constitute the sample, the mean true validity estimate is .29. The standard deviations of true validity for applicant and employee samples are 0 and .18, respectively. Hence, in studies in which applicants constitute the sample, 100% of the variance is explained by statistical artifacts. On the other hand, in validity studies in which employees constitute the sample, 42% of the variance is explained by the statistical artifacts, and the lower credibility value is .08, indicating that the validity is positive across studies and situations. But the large standard deviation of true validity and the low percent variance accounted for in employee samples suggests that other statistical artifacts or potential moderators may be operating. Validation sample (applicants vs. employees) seems to be a moderator of integrity tests in predicting job performance.

A fifth potential moderator of integrity test validities for predicting job performance is job complexity. Three job complexity levels were used: high, medium, and low, as defined by Hunter et al. (1990). Several studies reported too little information to determine with certainty whether the sample was of

high, medium, or low complexity. For the criterion of job performance, only 110 validation studies reported the information necessary to look up the DOT code for the job on which the validation was undertaken. For the other 112 studies providing validity coefficients with job performance, either no data was available on the jobs constituting the sample or the studies indicated a heterogeneous sample comprised of several jobs (e.g., retail employees). Among the studies which supplied information on the jobs studied, most were conducted on medium complexity jobs. Of the 110 studies, 80 were reported on medium complexity jobs. Only 19 studies reported validities for low complexity jobs, and only 11 reported validities on high complexity jobs. The meta-analysis results for this moderator are provided on lines 5a, 5b, and 5c of Table 5. The meta-analysis results indicate that for low complexity jobs, the mean true validity across 1,633 people is .45, and the standard deviation of the true validity is zero. For low complexity jobs, the artifacts that we correct for explain all the observed variation in integrity test validities in predicting job performance. For medium complexity jobs, the mean true validity across 14,701 people is .32; and the standard deviation of the true validity is .15. Statistical artifacts account for 50% of the variance. For high complexity jobs on this set of validities the mean true validity across 754 people and 11 validities is .46, and the standard deviation of the true validity is 0. Given the small sample size and the small number of correlation in the high complexity category, the results may not be robust. However, from these results an interesting pattern emerges suggesting that even for high complexity jobs,

integrity tests are valid in predicting job performance at a level comparable to their validity for low complexity jobs.

In personnel selection, supervisory ratings of job performance are a widely used and hence important criterion measure. Most validation studies of other predictors used in personnel selection use the criterion of supervisory ratings of job performance. Furthermore, most validity generalization studies have been conducted based on studies using that criterion. In addition, supervisory ratings of job performance rarely concentrate on only one aspect of performance such as quality or quantity of production. Instead supervisory ratings of job performance constitute an overall evaluation of an individual's work performance (Orr et al., 1989). The validities coded for this database were ratings of overall job performance and not partial performance ratings. Finally, utility analysis as typically conducted requires the use of a criterion of overall job performance. For this reason, integrity test validities based on the criterion of supervisory ratings of job performance were analyzed separately for moderating influences. These results are reported in Table 6.

- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -

Insert Table 6 About Here

- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -

For the most part, results are similar to the results reported for job performance in Table 5. Test type does not seem to be a strong moderator of the integrity test validities. Overt integrity tests predict supervisory ratings with a true validity of .30 and personality-based integrity tests predict supervisory ratings with a true validity of .37 (lines 1a and 1b of Table 6).

The mean true validity estimate across studies which used a concurrent validation strategy is .39, with an SDρ value of .11 (Line 2a of Table 6). The true validity across studies which used a predictive validation strategy is .32, with an SDρ value of .13. These results suggest that when the criterion of interest is supervisory ratings of overall job performance, concurrent validities may overestimate predictive validities in the domain of integrity testing. However, as was noted in the similar moderator analysis for all measures of job performance, among predictive studies included here, there was a very large sample study (N = 6,884) reporting an observed validity of .15. For the predictive validities, the total sample size was 22,657 with a mean observed correlation of .19. To counteract the potentially biasing effect of this one study, we calculated the unweighted mean observed validity for predictive studies (Unweighted mean correlation = .28). When the statistical artifact corrections were applied to this unweighted mean validity, the true validity obtained for the predictive validation category was .46. A similar re-analysis was not necessary for the concurrent validation category as there was no large sample single study in this category. However, for comparison purposes, the sample size weighted mean observed validity for concurrent studies was .23 and the unweighted mean observed validity was .26, which became .43 after correction for statistical artifacts. Thus, the moderating influence of validation strategy on validities for the criterion of supervisory ratings of job performance is inconclusive. Other analyses reported in Study 2 of this paper will examine whether concurrent and predictive validities are similar for integrity tests for other types of criterion measures (counterproductive behaviors).

For the potential moderators of validation sample (applicant vs. employee) and job complexity (low vs. medium vs. high), the same conclusions are reached for the criterion of supervisory ratings of overall job performance as were reached earlier for the combined criteria of job performance (Lines 3a through 4c in Table 6). Studies conducted on applicant samples seem to yield higher estimated operational validities than those conducted on employee samples ($\rho$=.42 and .33, respectively). Integrity tests also seem to be at least as valid for high complexity jobs as for low complexity jobs ($\rho$ = .51 and $\rho$ = .46, respectively).

The moderator analyses reported for job performance and supervisory ratings of job performance may give a distorted picture if the moderator variables are not independent. In order to determine the relationships among the moderators, intercorrelations of the moderator variables were calculated. The results are reported in Table 7.

- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -

Insert Table 7 About Here

- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -

Job complexity is not highly correlated with the other potential moderators (average correlation = -.06). Type of test (overt vs. personality-based) does not seem to be highly correlated with the other potential moderators (average correlation = -.11). However, validation strategy is substantially correlated with the sample used, applicants vs. employees (r = -.58). Predictive studies more frequently used applicant samples, and concurrent studies more frequently used employee samples, as concurrent criterion data is typically not available on applicant samples. This finding is consistent with expected practice in

traditional personnel psychology research. Earlier moderator analyses for all job performance criteria and for the supervisory ratings of job performance (Tables 5 and 6, respectively) resulted in the conclusion that validation strategy and validation sample may moderate the integrity test validities. Because these two moderators seem to be highly correlated, a hierarchical moderator analysis is needed to assess the potential impact of confounding on the moderator analyses. To accomplish this, all integrity test validities for supervisory ratings of overall job performance were broken down by validation strategy first and then within the concurrent and predictive validation categories, a moderator analysis by validation sample (applicants vs. employees) was undertaken. These results are reported in Table 8.

- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -

Insert Table 8 About Here

- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -

In personnel selection the purpose of the criterion-related validity coefficient is to estimate how the predictor will operate when applicants are administered the instrument and the results are used to predict job performance at some future point in time. The upper left cell in Table 8 indicates that when integrity tests are administered to applicants and the scores are used to predict later supervisory ratings of job performance, the mean operational validity is .41. This result is based on 6,674 individuals and 23 validity coefficients. The standard deviation of the true validity is 0, indicating that all the variance across studies and situations observed in this cell is due to statistical artifacts and the true validity of .41 is invariant across settings. When employees make up the

sample of predictive studies (upper right cell in Table 8), the operational validity is much lower, $\rho = .26$ across a total sample size of 6,118 and 20 validity coefficients. In addition, the standard deviation of true validity is .21, with only 24% of the variance accounted for. Concurrent validation conducted on employees (lower right cell) produces an operational validity of .37 across 8,264 individuals and 63 validity coefficients. The standard deviation of the true validity is .14, and 61% of the observed variance is accounted for by statistical artifacts. One study reported a validity coefficient for a concurrent validation strategy using an applicant sample. In that case the criterion was supervisory ratings of performance on a work sample administered to applicants, a very nontraditional criterion. However, given the extremely small sample size of that study ($N = 27$), little weight should be given to this validity coefficient. The overall results from Table 8 seem to indicate that concurrent validities overestimate predictive validities. For employees, the estimated mean true concurrent validity is .37; while the estimated mean true predictive validity is .26. Second, when the validation strategy is controlled for, validities from applicant samples are higher than validities from employee samples. For predictive validities, the applicant group mean true validity is .41, and the employee group mean true validity is .26. Although both validation strategy and validation sample seem to affect estimates of integrity test validities for predicting supervisory ratings of overall job performance, the highest mean operational validity estimate is obtained in applicant samples using predictive validation strategies ($\rho = .41$). This is the type of validity estimate that is most relevant in personnel selection.

## Study 2: Analyses and Results for Predicting Counterproductive Behaviors

As was reported in Table 4, the mean operational validity across all integrity tests for predicting counterproductive behaviors on the job is .47. However, the large standard deviation of the validity (.37) and low percent variance accounted for by the statistical artifacts (9%) indicate that there might be potential moderators affecting this category of validities. The results of the moderator analyses for predicting counterproductive behaviors are reported in Table 9.

- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -

Insert Table 9 About Here

- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -

The first potential moderator tested is the predictor type (overt vs. personality-based). All available correlations between overt integrity tests and disruptive behaviors on the job were used. The results across 305 correlations and 242,967 data points (Line 1a in Table 9) show that the best estimate of the mean validity of overt tests in predicting disruptive behaviors is .55. The worst case value of .07 indicates that the validity is positive across studies and situations. However, the percent variance accounted for by corrected statistical artifacts is low at 9%, and the standard deviation of the true validity ($SD_\rho$) is large at .41. The meta-analysis of personality-based integrity test validities shows a mean validity of .32 in predicting counterproductive behaviors with 44% of observed variance accounted for by the statistical artifacts that we could correct for (Line 1b in Table 9). $SD_\rho$ for personality-based integrity tests was .11, much smaller than the value of .41 for overt tests. The lower credibility

value of .20 indicates that validities of personality-based integrity tests are positive across studies and situations. Overall, these results suggest that overt integrity tests may be better in predicting counterproductivity ($\rho$ = .55) than personality-based tests ($\rho$ = .32).

The second moderator analysis involves testing for moderators by criterion measurement method (admissions of counterproductivity vs. external measures). In their meta-analysis of the validities of one integrity test, McDaniel and Jones (1988) found that validities against self-report measures were higher than those against external criteria. We therefore separated integrity test validities into those using admissions criteria and those using external criteria, such as supervisory ratings of theft, cash shortages, actual theft, and organizational records of other counterproductive behaviors. Results are shown in lines 2a and 2b in Table 9. They support the McDaniel and Jones (1988) findings, and indicate that admissions criteria yield a mean true validity estimate of .58, while for predicting external criteria, the mean true validity estimate is .32. The SD$\rho$ values in the two categories are .40 and .22, respectively. Only 10% of the variance is accounted for by artifacts with admissions criteria, and 16% with external criteria. The fairly large standard deviations of the true validities and relatively small percent variances accounted for indicate that validities of integrity tests may be affected by other moderators. However, the positive 90% credibility values indicate that the integrity test validities can be expected to be positive across situations for both the criteria of admissions of counterproductivity and externally measured counterproductivity.

We next examined criterion breadth as a potential moderator of validity for counterproductive behaviors criteria. As seen in line 3a of Table 9, integrity test validities against theft criteria yield an estimated mean operational validity of .52 and a 90% credibility value of .06 with 10% percent of the variance accounted for. The SDρ for this analysis is .39. As shown on line 3b in Table 9, validities against broad criteria (general disruptive behaviors) have an estimated mean corrected validity of .45, with a 90% credibility value of .04 and 9% of variance accounted for by the statistical artifacts. In this case, the SDρ was .36, again a fairly large value. The difference in operational validities for theft criteria ($\rho$ = .52) vs. other disruptive behaviors ($\rho$ = .45) indicate that criterion breadth may be a moderator of integrity test validities.

The fourth potential moderator studied for the criterion of counterproductivity is the validation strategy used in the studies. To determine whether concurrent validities estimate predictive validities accurately in this noncognitive domain, predictive and concurrent studies were separately analyzed (lines 4a and 4b in Table 9). Predictive validities have a mean of .36, while concurrent studies have a mean of .56. These results suggest that concurrent validities may overestimate predictive validities in this research domain. The utility of a selection test depends on its predictive validity; the only purpose of concurrent validity is to estimate predictive validity. Thus, the present finding is potentially important. The percent variance accounted for with both concurrent and predictive validities is 10%. SDρ is higher for concurrent than for predictive validities (.39 for concurrent validities and .28 for predictive

validities). However, in both cases the 90% credibility values indicate validity is likely to be greater than zero, regardless of the validation strategy used.

The next potential moderator tested was the validation sample (applicant vs. employee). This analysis is not redundant with the analysis of predictive vs. concurrent studies, for two reasons. First, some concurrent (K = 87) studies were conducted on applicants; these were studies that used criteria of admissions, and the admissions were obtained from applicants. Second, some predictive studies were conducted with employees (K = 39); in these studies, the criterion data were not gathered until a considerable time after administration of the test. The mean estimated operational validity is .44 in applicant samples and .54 in employee samples (Lines 5a and 5b in Table 9). Thus, employee samples appear to yield larger validity estimates, a finding consistent with the results of the analysis of predictive vs. concurrent studies. The SDρs for these two categories were .35 and .47, respectively. For both types of samples, the lower 90% credibility interval is positive indicating that the validities are positive across all situations and settings.

A sixth potential moderator of integrity test validities in predicting counterproductive job behaviors is job complexity. As in the job complexity analysis in Study 1, three job complexity levels were used: high, medium, and low (as defined by Hunter et al., [1990]). Three hundred studies reported too little information to determine with certainty whether the sample was of high, medium, or low complexity. For example, some studies indicated only that the sample consisted of "retail employees" without identifying the jobs included in the sample. Among the studies which supplied information on the jobs studied

most were conducted on medium complexity jobs. Of the 143 correlations indicating specific jobs used in validation, 78 were reported on medium complexity jobs. Only 21 studies reported validities for high complexity jobs, and 44 studies reported validities for low complexity jobs. The results indicate that for low complexity jobs, the mean true validity of integrity tests across 9,654 people is .43, the standard deviation of the true validity is .25, and the artifacts that we correct for explain 23% of the observed variation in integrity test validities. For medium complexity jobs, the estimated mean true validity across 19,866 people is .40, the standard deviation of the true validity is .24, and statistical artifacts account for 24% of the variance. For high complexity jobs, the mean true validity across 2,246 people is .68, and the standard deviation of the true validity is .20. The percent variance accounted for by the statistical artifacts is 45%. Because our classification of the validities into the three categories has resulted in the loss of approximately 68% of the validities in the database, perhaps no definitive conclusions can be reached for this hypothesized moderator. Yet an interesting trend does emerge: As the level of job complexity increases, the mean true validity may increase. There seems to be some evidence that the mean validity of integrity tests is highest for high complexity jobs. This was an unexpected result. One possible explanation for this trend may be that in high complexity jobs, less supervision is received and consequently there is more opportunity to be dishonest and display other counterproductive behaviors, making these behaviors easier to measure. But this is purely speculative.

As was the case in Study 1, the results reported above and in Table 9 may be difficult to interpret if the hypothesized moderators are intercorrelated. To explore this possibility for Study 2, we correlated dummy coded hypothesized moderators of integrity tests using only those studies which reported validities for counterproductivity. The results are reported in Table 10.

- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -

Insert Table 10 About Here

- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -

Results indicate that the moderators of job complexity and validation sample (applicants vs. employees) are not highly correlated with the other moderators. Most other moderators seem to be substantially correlated with each other. Predictor type (overt vs. personality-based) correlates substantially with criterion measurement method (admissions vs. external criteria), criterion breadth (theft vs. broad criteria), and validation strategy (predictive vs. concurrent). This means that overt tests tended to be used with admissions criteria, narrow criteria (theft only), and in concurrent studies. Similarly, criterion measurement method correlates very highly with validation strategy (observed r = .74), meaning that studies using admissions criteria tended to be concurrent studies. Because some of the correlations between the potential moderators in Study 2 are substantial, a fully hierarchical moderator analysis was conducted for all potential moderators except job complexity.

In a fully hierarchical moderator analysis, the dataset of correlations is broken down by one key potential moderator variable first, and then within each subgroup subsequent moderator analyses are undertaken one by one in an

hierarchical manner (Hunter & Schmidt, 1990a, p. 527). First, the validities for counterproductive behaviors were divided into two categories by predictor type (Overt vs. Personality-Based). Within each predictor subgroup, validities were then sorted into the external criteria or the admissions criteria. Next, the validities in each subgroup were further grouped by theft criteria vs. broad criteria, predictive vs. concurrent validation and applicant vs. employee sample. The fully hierarchical moderator analysis takes all the moderators being taken into consideration simultaneously: five moderators with two levels each resulting in $2^5 = 32$ combinations. The results of the fully hierarchical analysis are reported in Table 11.

- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -

Insert Table 11 About Here

- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -

Due to lack of information on some potential moderators in some studies, the breakdown of our database to 32 cells, as presented in Table 11, resulted in the loss of about one third of the validity data from the analyses. The major reason for the loss of data is that many studies did not report whether the predictor data was collected from current employees or applicants.

Overt Tests

The results in upper half of Table 11 indicate that validities for overt tests are in general lower for applicant samples than for employee samples. The respective true estimated validities are .13 vs. .16 for predictive validation using external theft criteria; .32 vs. .94 for concurrent validation using externally measured broad counterproductivity criteria; .42 vs. .54 for

concurrent validation using theft admissions criteria, and .46 vs. .99 for concurrent validation using admissions of broad counterproductivity criteria. The exception to this trend is the higher predictive validity obtained for applicant samples ($\rho$ = .39) than for employee samples ($\rho$ = .09) when overt tests are used to predict externally measured broad counterproductivity on the job. There is no ready explanation for this exception. For unknown reasons, predictive validities for this criterion are quite small for overt tests.

The operational selection validity of a test can best be estimated by its predictive validity computed using applicants. In light of this, the estimated true predictive validity of .39 for overt integrity tests in predicting externally measured broad counterproductivity when the predictor is administered to applicants indicates substantial potential utility in using overt tests in selection. However, when the criterion is the much narrower one of (externally measured) theft alone, the mean estimated validity from predictive studies conducted on applicants is a considerably smaller .13. The relatively low validity estimates for externally measured theft criteria may be underestimates to some degree. The reliability estimates used in these meta-analyses were for counterproductive behaviors in general (See Table 3), rather than reliability values for externally detected theft per se. No reliability estimates of the latter measures were found. It is possible that the reliability of external theft measures is lower on average than the reliability of all counterproductive behaviors. However, if external theft measures had a true average reliability of only .30, the mean true validity estimate of .13 in Table 11 would rise to only .20. Thus the relatively low validities for externally measured theft are

unlikely to be explainable solely on grounds of undercorrection for criterion unreliability.

For the criterion of broad counterproductive behaviors externally measured, concurrent validities computed using present employees substantially overestimate the predictive validity of overt integrity tests derived from applicant samples. The mean operational validity of .94 is 2.41 times larger than the .39 that we believe is the best estimate of operation selection validity of overt tests for this criterion measure. Although the concurrent validity estimate of .32 derived on applicants does not overestimate predictive validity, this figure is based on only two studies and a total N of only 213. For this reason, it should receive little weight in the interpretation of the findings. In addition, as discussed in the next section, concurrent validities conducted on applicants are very atypical validity studies.

The results in Table 11 indicate that no matter what the content of the criterion measure (theft or broadly defined disruptive behaviors), self-reported criteria tend to result in higher estimates of validities for integrity tests. Many may judge that correlations with self-report criteria are not acceptable as estimates of the operational validity of integrity tests; however, it is not entirely clear that external measures of counterproductive behaviors are more valid than admissions of such behaviors. Many thefts and other counterproductive behaviors may go undetected, limiting the validity of external measures. In addition, there is considerable evidence from research on juvenile delinquency that the correlation between admissions and actual behavior is substantial (about .50; Viswesvaran, Ones, & Schmidt, 1992). In any event,

validities against admissions criteria can be taken as evidence of construct validity. All studies using admissions criteria have been concurrent; Table 11 contains no predictive validities for this criterion. The meta-analyses of overt test correlations with admissions criteria indicate that correlations are higher for employees than for applicants. For self-reports of theft, the true estimated mean correlation is .54 for the N = 2,917 employee sample and .42 for the N = 67,618 applicant sample. In both cases the $SD_\rho$ is large enough to indicate additional moderators may be operating. However, the positive lower credibility values mean that a positive correlation can be expected between honesty test scores and admissions of theft in studies with concurrent design for both employee and applicant samples regardless of the setting and situation. When the admissions criteria include other disruptive behaviors such as tardiness, violence on the job, absenteeism, drug abuse, and alcohol abuse in addition to only theft, mean correlations of overt tests increase to .99 for employee samples (N = 27,887) and .46 for applicant samples (N = 85,824). In both these cases, self-report criteria were collected concurrently with the predictor data. The pattern of mean correlations for both theft and broad counterproductive criteria suggest that employees are more willing to admit negative behaviors than are applicants. Under this interpretation, the lower correlations for applicants are due to response distortion by applicants. (Here the focus is on response distortion on the (self-report) criterion measure, but there may also be response distortion on the predictor by applicants.) A much larger portion of the variance in the observed correlations is accounted for by statistical artifacts when the sample is comprised of employees rather than applicants (67% of the

variance in the employee sample; 9% in the applicant sample). In both cases the positive lower credibility value indicates that the concurrent correlations of overt integrity tests with self-reported broad counterproductivity criteria are positive. Taken together, the results for self-report criteria support the construct validity of overt integrity tests.

Summarizing across both admissions criteria and externally measured criteria, it is noteworthy that overt tests predict broad disruptive behaviors better than they predict theft alone. This pattern of findings suggests that the construct being measured by these tests is not theft-proneness per se (as Ash, 1985 and others have hypothesized), but a broader construct which includes theft among many other disruptive behaviors on the job. We suspect that this broad construct is general conscientiousness.

Personality-Based Tests

For personality-based tests, the estimated true validities from applicant samples are equal to or higher than validities obtained using employee samples, controlling for all other moderators. The respective mean validities for externally measured broad counterproductivity criteria are .29 vs. .26 (predictive), and .77 vs. .29 (concurrent). In contrast to overt tests, the true standard deviation of personality-based tests is zero or negligibly small (i.e., .02). For personality-based tests virtually all the variance in the observed validities is accounted for by statistical artifacts. The mean true validities obtained for personality-based tests do not appear to vary across organizations or situations. One odd category of analysis for personality-based integrity tests is concurrent studies done on applicants with external criteria (K = 6, N =

4,261). These studies used reference checks from previous employers, police reports obtained, interviewer evaluations, and in one case disruptive behaviors observed during a one day assessment center. This constellation of broad disruptive behaviors criteria is not representative of the other broad counterproductive behaviors criteria, and appears to be responsible for the extraordinarily large ρ obtained for this category (.77). These studies can be taken as supportive of the construct validity of personality-based integrity tests. The key validity estimate in Table 11 for personality-based tests is the mean true validity of .29 from the 62 predictive studies conducted on 76,835 applicants using broad measures of counterproductive job behaviors externally assessed. This is the best estimate of the operational validity of these tests in selection for the criterion they were designed to predict. As noted earlier, the comparable value for overt tests is .39.

## Critical Summary of Findings

### Job Performance

In selection settings, the best estimate of integrity test validities for predicting job performance would be based on (a) predictive studies (b) conducted on samples of applicants. To obtain such an estimate of the mean validity of integrity tests for selection, we meta-analyzed predictive validities calculated on applicant samples (Table 8). There were 23 such validities for predicting supervisory ratings of job performance. Across 6,674 people, the best estimate of the mean true validity was .41. The SDρ was 0, and the percent variance accounted for was 100%. These findings imply that the average validity that integrity tests may be expected to have in selection settings is .41, and that

this value is constant across settings. The meta-analysis results presented in this research also show that overt and personality-based tests produce fairly similar operational validities when the criterion of interest is supervisory ratings of job performance.

Counterproductive Behaviors

Generally, validities for integrity tests for predicting counterproductive behaviors on the job appear to be fairly substantial. However, several moderators were identified for this type of criterion: type of test (overt vs. personality based), criterion measurement method (admissions vs. external), criterion breadth (theft vs. broad counterproductivity), validation strategy (predictive vs. concurrent), and validation sample (applicants vs. employees). When the effects of these moderators are controlled (see Table 11), the standard deviations of true validity ($SD\rho$) for integrity tests appear to be comparable to those of ability tests in predicting job performance (e.g., Pearlman, Schmidt, & Hunter, 1980; Schmidt, Hunter, Pearlman, & Shane, 1979). Some exceptions to this conclusion are concurrent studies of overt tests conducted on employees using externally measured broad counterproductivity criteria ($SD\rho$ = .29 in Table 11), and concurrent studies of overt tests conducted on applicants using admissions of theft and broad counterproductive behaviors ($SD\rho$ = .33 and $SD\rho$ = .35, respectively in Table 11).

For the criterion of counterproductive behaviors, admissions produce much higher correlations than external criteria, and concurrent studies often seem to overestimate predictive validity. The utility of a selection test depends on its predictive validity; the only purpose of concurrent validity is to estimate

predictive validity. Thus, the finding that in this research domain concurrent validity estimates overestimate predictive validity is potentially important. Theft appears to be less predictable than broad counterproductive behaviors, although this comparison could be made only for overt integrity tests.

In selection settings, the best estimate of integrity test validities for predicting theft would be based on predictive studies conducted on applicants. In addition, as noted earlier, many would argue for reliance on external criteria in preference to admissions criteria, although the relative construct validity of these two criterion measures is unclear at present. Considering externally measured theft as the criterion in predictive studies, we find that the mean operational validity of overt integrity tests is estimated at .13 (Table 11). For reasons explained earlier, this value may be an underestimate. For personality-based tests, no validity estimates for the prediction of theft alone were available. Considering externally measured broad counterproductive behaviors as the criterion in predictive studies conducted on applicants, we find that the mean operational validity of overt integrity tests is .39 (Table 11). For personality-based tests, the estimated operational validity for predicting broad counterproductive behaviors is .29 (Table 11).

In sum, integrity tests predict overall job performance with moderate and generalizable validity. They also predict counterproductive behaviors such as theft, absenteeism, tardiness, and disciplinary problems, but that validity seems to be affected by several simultaneously operating moderators. All in all, the validity of integrity tests is positive and in useful ranges for both overall job performance criteria and counterproductive behaviors criteria.

## Implications of Findings

### Implications for Incremental Validity

A key unanswered question is the size of the increment in validity from adding integrity tests to general mental ability tests in predicting overall job performance in personnel selection. Many studies suggest that the correlations between integrity measures and ability measures are extremely low and negligible. For example: when Jones and Terris (1983) investigated the correlation between an overt integrity test and a measure of general mental ability, the correlations were -.02 for theft admissions and -.03 for theft attitudes; Gough (1972) reported that a vocabulary test correlated -.05 with the Personnel Reaction Blank; Werner, Jones, and Steffy (1989) reported that integrity test scores are unrelated to educational level (an arguable proxy for ability); Hogan and Hogan (1989) reported correlations of .07 and -.09 between the Hogan Reliability Scale and the quantitative and verbal portions of the Armed Services Vocational Aptitude Battery (ASVAB), respectively. Thus if we assume that the correlation between ability and integrity measures is zero, based on these studies, the expected maximum incremental validity of integrity tests can be calculated. Table 12 presents the predicted incremental validity of integrity tests for each of the five job complexity levels used by Hunter (1980).

- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -

Insert Table 12 about here

- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -

In Table 12, the first column of multiple correlations shows the combined validity of integrity and general mental ability test scores. For example, for medium complexity jobs (complexity level 3), the multiple correlation is .65. This is an

increase in validity of 27% compared to ability alone, and an increase in validity of 59% compared to integrity alone. The second column of multiple correlations in Table 12 reports the combined validity of general mental ability, psychomotor ability, and integrity. The correlations between general mental ability and psychomotor ability necessary to calculate the multiple correlations were obtained from Hunter (1980); they are about .30 across each of the various job complexity levels. The multiple correlation for predicting overall job performance is .64 for the lowest complexity jobs (level 5), .67 for medium complexity jobs (level 3) and .72 for highest complexity jobs (level 1). These preliminary results appear to indicate that using integrity tests in conjunction with measures of ability can lead to substantial incremental validity for all job complexity levels. We now have research underway to more exactly estimate the relationship between measures of integrity and measures of ability in order to obtain more precise estimates of the magnitude of the incremental validity of integrity tests.

Implications for Adverse Impact

Hunter and Hunter (1984) indicate that it may be possible to identify other predictors that will add to the validity of general mental ability and at the same time reduce adverse impact. Integrity test publishers have devoted considerable research to examining the question of adverse impact. No differences have been found in mean test scores of minorities and whites (e.g., Arnold, 1989; Bagus, 1988; Cherrington, 1989; Moretti & Terris, 1983; Strand & Strand, 1986; Terris & Jones; 1982). Sackett et al. (1989, p. 499) concluded "... minority groups are not adversely affected by either overt integrity tests or personality oriented measures". Integrity test scores and race appear to be uncorrelated. From the ability testing literature,

we know that blacks average about one standard deviation below whites on tests of general mental ability. Given this information, the mean difference between blacks and whites on an equally weighted composite of ability and integrity test scores and race is .67 standard deviations. Thus, when ability and integrity test scores are equally weighted, the black-white difference is reduced approximately by 36.4% in comparison to ability tests used alone. This reduction can be expected to translate into a greater reduction in adverse impact (reduction in adverse impact depends on the selection ratio as well). By way of example, suppose all those above the white mean were selected (i.e., a selection ratio of .50 for whites). In this case, the percentage of blacks selected based solely on ability, without an integrity test, would be 15.9%. However, if an integrity and an ability test were used together, with scores equally weighted, the percentage of blacks selected would increase to 25.1%. This is an increase in hiring rate of blacks by 58.3%.

Even though the use of integrity tests alone should produce no adverse impact, it can be expected to result in loss in utility of at least 37% in comparison to use of ability and integrity tests in combination. Stated alternately, using a composite of ability and integrity tests in selection can be expected to result in improved utility of at least 58% compared to integrity alone. These calculations are based on the figures in Table 12. Hence, the implication is that employers should use integrity tests in addition to measures of general mental ability. This combination has the potential for reducing adverse impact and enhancing validity and utility at the same time. Questions related to adverse impact and utility of integrity tests are explored in detail in Ones, Viswesvaran, and Schmidt (1992).

## Discussion

One question we have repeatedly pondered since beginning our research on integrity tests, has been the question of potential response distortion, including the possibility of faking, responding in a socially desirable manner, or otherwise responding inaccurately. The conclusion we infer from our meta-analytic results is that response distortion, to the extent that it exists, does not seem to destroy the criterion-related validities of these tests. Substantial validities were found for studies conducted on applicants. Applicants in these studies experienced all the usual inducements for response distortion, yet substantial estimated mean validities were nevertheless observed.

Some concerns have been raised regarding integrity tests generally. One concern involves the absence of strong empirical evidence for choosing any particular base rate for honesty in studies of overt tests used to predict theft. Base rate refers to the proportion of test takers in the referent population who are actually dishonest by some criterion. But the absence of an established base rate for honesty has no relevance for the validity of integrity tests. In exploring this question, we first note that usage of the terms false positive and false negative in integrity testing is the reverse of the regular usage of these terms in personnel selection. In an integrity test setting, a false positive error is the rejection of an applicant who would be honest if hired, and a false negative error is the acceptance of an employee who is dishonest. Some have argued that integrity test usage results in high false positive rates (that is rejection of applicants who would be honest if hired) because the associated base rates are low (US OTA, 1990). This argument implicitly assumes all applicants would be accepted if an integrity test were not

used. Such an assumption is untenable in a selection setting, and the failure to use any valid selection predictor will result in a higher false positive rate than its use. High overall false positive rates are primarily the result of having more applicants than positions (Martin & Terris, 1990). False positive rates depend on the validity of the selection procedure used. As validity increases, both types of decision errors decline. Therefore, any improvement in validity of the selection process will reduce both the probability of rejecting a qualified applicant and the probability of accepting an unqualified one. Hence, no matter what the actual base rate is for honesty, the validity of integrity tests cannot be challenged on the grounds of low base rates. However, the utility of integrity tests to the organization does depend on the base rate of dishonesty in the applicant pool. The larger this base rate (up to 50%), the greater will be the utility, other things being equal. Therefore when overt integrity tests are used to predict only employee theft, the question of base rates is important in determining utility.

Some limitations of the present study need to be pointed out. First, in some cells of the fully hierarchical moderator analyses, the number of existing studies is small enough to raise concerns about the stability of the estimates. Any empirical study of validity generalization is limited by the number of available validation studies with particular criterion-predictor combinations. This has implications for second order sampling error in meta-analyses (Hunter & Schmidt, 1990a, pp. 411-450). But even with this limitation, a meta-analytic review based on a reasonable conceptual or theoretical framework provides sounder conclusions than other approaches to understanding the data, including the traditional narrative review.

A second limitation of this study is the inability to conclusively determine the validities of integrity tests as a function of job complexity. Nonetheless, a preliminary exploratory moderator analysis suggested that the mean validity of integrity tests is highest for high complexity jobs. This result may imply increased opportunity to be dishonest in higher complexity jobs. This increased opportunity could result from less supervision and control coupled with increased access to resources. Another implication of this finding is that the expectation that applicants to high complexity jobs may engage in more response dissimulation or show more of other forms of response distortion on integrity tests than other individuals may be incorrect. Future research should explore job complexity further as a moderator of integrity test validities.

It is our hope that future criterion-related validity studies on integrity tests will discontinue the practice of pooling data across jobs differing in level of complexity and will provide full information on reliabilities, range restriction, and other artifacts. Another problem in this literature is that only a small proportion of the available validity studies of integrity tests have been published in the professional journals, and many of the unpublished reports are sketchy, often omitting important information. Perhaps as the potentially important implications of this sort of research become work widely known, journals will be more likely to publish studies in this area and researchers will be more willing submit them for publication.

This validity generalization effort is noteworthy in two respects: (a) most of the studies reporting criterion-related validities for integrity tests came from service jobs (the largest sector of the US economy), although some validities for

manufacturing jobs were reported; (b) the meta-analysis of integrity tests is based on one of the largest data bases in the literature (665 validity coefficients based on 576,464 data points). Even in the domain of mental abilities, few data bases have been this large. Before beginning this research, we would not have estimated that the extant data base for integrity tests was this large.

The finding that selection instruments can predict externally measured composite measures of irresponsible or counterproductive behaviors (e.g., disciplinary problems, disruptiveness on the job, tardiness, absenteeism) with substantial validity seems remarkable. Industrial psychologists have long been concerned with such behaviors and their negative impact on individual and organizational performance. There is evidence indicating that employers are even more concerned about such behaviors. For example, the Michigan Employability Survey (Michigan Department of Education, 1989) found that of 86 employee qualities ranked for importance in entry level employment by over 3000 employers, seven of the top eight qualities were related to integrity, trustworthiness, conscientiousness and related qualities. The other quality in the top eight (ranked 5th) referred to general mental ability.

The implications of these findings are substantial. For example, the most commonly used selection procedure could become a combination of general mental ability scores and an integrity test. Also, these findings raise the question of whether general conscientiousness is in actuality the motivation variable that has been so elusive in personnel psychology (Schmidt & Hunter, in press; Schmidt, Ones, & Hunter, 1992). That is, conscientiousness may be the most important trait motivation variable. Across jobs in general, mental ability and conscientiousness

may be the two most important determinants of job performance (Schmidt & Hunter, in press). Considerably more research on this question will be needed in the future.

Additional research is needed on the construct validity of integrity tests. With the exception of Woolley and Hakstian (in press) and Collins and Schmidt (1992), there is relatively little research aimed at determining what constructs are measured by integrity tests. We currently have work underway investigating construct validity questions about integrity tests. Research in this area was recommended by the APA Task force report on integrity tests (Goldberg et al., 1991).

When we started our research on integrity tests, we, like many other industrial psychologists, were skeptical of integrity tests used in industry. Now, based on a database across more than 500,000 individuals and more than 600 validity coefficients, we conclude that integrity tests substantial evidence of generalizable validity. Our findings indicate that both overt and personality-based measures of integrity correlate substantially with supervisory ratings of job performance and with both self-reported and externally measured counterproductive behaviors. Our meta-analyses confirm many of our moderator hypotheses. However, perhaps the most significant conclusion of this research is that integrity test validities are positive across situations and settings despite moderating influences on their exact magnitudes.

## References

Alexander, R. A., Carson, K. P., Alliger, G., & Cronshaw, S. F. (1989). Empirical distributions of range restricted $SD_X$ in validity studies. Journal of Applied Psychology, 74, 253-258.

Arnold, D. W. (1989). The Reid Report's lack of adverse impact (Research Memorandum #22). Chicago, IL: Reid Psychological Systems.

Ash, P. (1985). Predicting Dishonesty with the Reid Report. Journal of the American Polygraph Association, 5(2), 139-145.

Bagus, K. K. (1988). Industry specific affirmative action studies. Stanton Survey Report. Charlotte, NC: Stanton Corporation.

Barrick, M. R., & Mount, M. K. (1991). The big five personality dimensions and job performance: A meta-analysis. Personnel Psychology, 44, 1-25.

Bemis, S. E. (1968). Occupational validity of the general aptitude test battery. Journal of Applied Psychology, 52, 240-244.

Borman, W. C., White, L. A., Pulakos, E. D., & Oppler, S. H. (1991). Models of supervisory job performance ratings. Journal of Applied Psychology, 76, 863-872.

Buss, A. H. (1989). Personality as traits. American Psychologist, 44, 1378-1388.

Campbell, J. P., McHenry, J. J., & Wise, L. L. (1990). Modeling performance in a population of jobs. Personnel Psychology, 43, 313-333.

Cherrington, D. (1989). Validation of the Applicant Review: Analysis of adverse impact (Research Report). Standfordville, NY: CHC Forecast Inc.

Collins, J. M., & Schmidt, F. L. (1992). Integrity, personality, and white-collar crime. Unpublished Manuscript.

Conoley, J. C., & Kramer, J. J. (Eds.). (1989). The Tenth Mental Measurements Yearbook. Lincoln, NE: Buros Institute of Mental Measurements.

Cormack, R. W., & Strand, A. L. (1970). Trustworthiness Attitude Survey. Oakbrook, IL: Personnel Systems Corporation.

Digman, J. M. (1990). Personality structure: Emergence of the Five Factor Model. Annual Review of Psychology, 41, 417-440.

Goldberg, L. R. (1990). An alternative "Description of Personality": The Big Five structure. Journal of Personality and Social Psychology, 59, 1216-1219.

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.

Gough, H. G. (1954). Personnel Reaction Blank. Palo Alto, CA: Consulting Psychologists Press.

Gough, H. G. (1972). Manual for the Personnel Reaction Blank. Palo Alto, CA: Consulting Psychologists Press.

Guastello, S. J., & Rieke, M. L. (1991). A review and critique of honesty test research. Behavioral Sciences and the Law, 9, 501-523.

Harris, W. G. (Undated). An investigation of the Stanton Survey using a validity generalization model. Charlotte, NC: Stanton Corporation.

Hogan, J., & Hogan, R. (1989). How to measure employee reliability. Journal of Applied Psychology, 74, 273-279.

Hogan, R. (1981). Hogan Personality Inventory. Minneapolis, MN: National Computer Systems Inc.

Hudson Institute. (1987). Workforce 2000: Work and workers for the twenty-first century. Indianapolis, IN.

Hunter, J. E. (1980). Validity generalization for 12,000 jobs: An application of synthetic validity and validity generalizations to the General Aptitude Test Battery (GATB). Washington, DC: US Employment Service, US Department of Labor.

Hunter, J. E., & Hunter, R. F. (1984). Validity and utility of alternative predictors of job performance. Psychological Bulletin, 96, 72-95.

Hunter, J. E., & Schmidt, F. L. (1990a). Methods of meta-analysis: Correcting error and bias in research findings. Newbury Park, CA: Sage

Hunter, J. E., & Schmidt, F. L. (1990b). Dichotomization of continuous variables: The implications for meta-analysis. Journal of Applied Psychology, 75, 334-349.

Hunter, J. E., Schmidt, F. L., & Judiesch, M. K. (1990). Individual differences in output variability as a function of job complexity. Journal of Applied Psychology, 75, 28-42.

Jones, J. W. (1980). Attitudinal correlates of employee deviance: theft alcohol use and nonprescribed drug use. Psychological Reports, 47, 71-77.

Jones, J. W.,& Terris, W. (1983). Personality correlates of theft and drug abuse among job applicants. Proceedings of the Third International Conference on the 16PF Test, pp. 85-94.

Klump, C. S. (1964). Stanton Survey. Charlotte, NC: Stanton Corporation.

Law, K. S., Schmidt, F. L., & Hunter, J. E. (1992). Non-Linearity of range corrections in meta-analysis: A test of an improved procedure. Unpublished manuscript.

London House Inc. (1975). Personnel Selection Inventory. Park Ridge, IL: Author.

London House Inc. (1982). Employee Attitude Inventory. Park Ridge, IL: Author.

London House Press. (1980). Personnel Selection Inventory Manual, Park Ridge, IL: London House Press.

London House Press. (1982). The Employee Attitude Inventory. Park Ridge, IL: London House Press.

Lousig-Nont, G. M. (1987). Phase II Profile. Las Vegas, NV: Author.

Martin, S. L., & Terris, W. (1990). The four-cell classification table in personnel selection: A heuristic gone awry. Industrial-Organizational Psychologist, 27, 49-55.

McDaniel, M. A., & Jones, J. W. (1986). A meta-analysis of the employee attitude inventory theft scales. Journal of Business and Psychology, 2, 327-345.

McDaniel, M. A., & Jones, J. W. (1988). Predicting employee theft: A quantitative review of the validity of a standardized measure of dishonesty. Journal of Business and Psychology, 2, 327-345.

Michigan Department of Education. (1989). The Michigan Employability Survey.

Miller, J. F., & Bradley, P. (1975). Milby Profile. Minneapolis, MN: Milby Systems Inc.

Moretti, D. M. (1986). The prediction of employee productivity through attitude assessment. Journal of Business and Psychology, 1, 134-147.

Moretti, D. M., & Terris, W. (1983). The Personnel Selection Inventory: A predictive validation study. Park Ridge, IL: London House Inc.

Nathan, B. R., & Alexander, R. A. (1988). A Comparison of criteria for test validation: A meta-analytic investigation. Personnel Psychology, 41, 517-536.

O'Bannon, R. M., Goldinger L. A., & Appleby, G. S. (1989). Honesty and Integrity Testing. Atlanta, GA: Applied Information Resources.

Ones, D. S., Viswesvaran, C., & Schmidt, F. L. (1992). Incremental validity, utility and adverse impact of integrity tests. Unpublished manuscript.

Orr, J. M., Sackett, P. R., & Mercer, M. (1989). The role of prescribed and nonprescribed behaviors in estimating the dollar value of performance. Journal of Applied Psychology, 74, 34-40.

Pajaanen, G. (1985). PDI Employment Inventory. Minneapolis, MN: Personnel Decision Inc.

Pearlman, K., Schmidt, F. L., & Hunter, J. E. (1980). Validity generalization results for tests used to predict job proficiency and training success in clerical occupations. Journal of Applied Psychology, 65, 373-406.

Rafilson, F. M., & Frost, A. G. (1989). Overt integrity tests vs. personality based measures of delinquency: An empirical comparison. Journal of Business and Psychology, 3, 269-277.

Reid Psychological Systems. (1951). Reid Report. Chicago, IL: Author.

Rothstein, H. R. (1990). Interrater reliability of job performance ratings: Growth to asymptote level with increasing opportunity to observe. Journal of Applied Psychology, 75, 322-327.

Sackett, P. R., & Decker, P. J. (1979). Detection of deception in the employment context: A review and critical analysis. Personnel Psychology, 32, 487-506.

Sackett, P. R., & Harris, M. M. (1984). Honesty testing for personnel selection: A review and critique. Personnel Psychology, 37, 221-246.

Sackett, P. R., Burris, L. R., & Callahan, C. (1989). Integrity testing for personnel selection: An update. Personnel Psychology, 42, 491-529.

Schmidt, F. L., & Hunter, J. E. (in press). Development of causal models of processes determining job performance. Current Directions in Psychological Science.

Schmidt, F. L., Hunter, J. E., & Gast-Rosenberg, I. (1980). Validity generalization results for computer programmers. Journal of Applied Psychology, 65, 643-661.

Schmidt, F. L., Hunter, J. E., Pearlman, K., & Shane, G. S. (1979). Further tests of the Schmidt-Hunter Bayesian validity generalization procedure. Personnel Psychology, 32, 257-281.

Schmidt, F. L., Ones, D. S., & Hunter, J. E. (1992). Personnel selection. Annual Review of Psychology, 43, 627-670.

Science Research Associates. (1983). Personal Outlook Inventory. Parkridge, IL: Author.

Society for Industrial and Organizational Psychology Inc. (1987). Principles for the validation and use of personnel selection procedures (3rd edition). College Park, MD.

Strand, A. L., & Strand, M. L. (1986). Validity, reliability, compliance, and efficacy studies of the T. A., A. I., E. S., D. A., and P. A. S. S. Surveys. Oak Brook, IL: Psychological Surveys Corporation.

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

Thorndike, R. L. (1949). Personnel selection. NY: Wiley.

US Congress Office of Technology Assessment. (1990). The use of integrity tests for pre-employment screening, OTA-SET-442. Washington, DC: US Government Printing Office.

Viswesvaran, C., Ones, D. S., & Schmidt, F. L. (1992, May). Appropriateness of self reported criteria for integrity test validation: Evidence from beyond. Paper presented at the Annual Meeting of the Society for Industrial and Organizational Psychology, Montreal, Canada.

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.

Woolley, R. M., & Hakstian, A. R. (in press). An examination of the construct validity of personality-based and overt measures of integrity. Educational and Psychological Measurement.

## Footnote

[1] To examine the robustness of the results in our meta-analyses to the artifact distributions used, all the analyses were re-conducted correcting only for sampling error. None of the conclusions about the presence and generalizability of validity changed.

## Author Notes

The order of authorship is arbitrary; all three contributed equally to this manuscript. None of the authors are or have ever been associated with any integrity test publishers. Communications regarding this manuscript should be directed to Deniz Ones at the Department of Management and Organizations, University of Iowa, Iowa City, IA 52242. Some of the research for this manuscript was supported by the US Navy, Office of the Chief of Naval Research (Contract No: N00014-91-G-4168 to Frank Schmidt). The content of this manuscript does not necessarily reflect the position or the policy of the Government, and no official endorsement should be inferred.

Earlier versions of this research were presented at the following conferences: Annual Conference of the Society for Industrial and Organizational Psychology, St. Louis, 1991; Annual Conference of the American Psychological Society, Washington DC, 1991; Annual Conference of the Society for Industrial and Organizational Psychology, Montreal, 1992; 26th International Congress of Psychology, Brussels, 1992. This manuscript supersedes all previous versions including the working manuscript cited in the APA Task Force Report (Goldberg et al., 1991).

We thank numerous test publishers, authors, and colleagues who sent us data on integrity tests. We also thank Paul Sackett for many informative and constructive discussions, Neal Schmitt for having inadvertently directed us to enlarge our database from 77 validities to 665, two anonymous reviewers, and many colleagues for their comments on earlier versions of our work.

Table 1

Tests Contributing Data to the Meta-Analyses

| Test Name |
|---|
| 1. Accutrac Evaluation System[a] |
| 2. Applicant Review[a] |
| 3. Compuscan[a,c] |
| 4. Employee Attitude Inventory (London House)[a] |
| 5. Employee Reliability Inventory |
| 6. Employment Productivity Index[b] |
| 7. Hogan Personnel Selection Series (Reliability Scale)[b] |
| 8. Integrity Interview[a] |
| 9. Inwald Personality Inventory[b] |
| 10. Orion Survey[a,c] |
| 11. P.E.O.P.L.E. Survey[a] |
| 12. Personnel Decisions Inc. Employment Inventory[b] |
| 13. Personnel Outlook Inventory[b] |
| 14. Personnel Reaction Blank[b] |
| 15. Personnel Selection Inventory (London House)[a] |
| 16. Phase II Profile[a] |
| 17. P.O.S. Preemployment Opinion Survey[a,c] |
| 18. Preemployment Analysis Questionnaire[a] |
| 19. Reid Report and Reid Survey[a] |
| 20. Rely[a] |
| 21. Safe-R[a,c] |
| 22. Stanton Survey[a] |
| 23. True Test[a] |
| 24. Trustworthiness Attitude Survey; PSC Survey; Drug Attitudes/Alienation Index[a] |
| 25. Wilkerson Preemployment Audit[a,c] |

Note. The list of publishers and authors of these tests are available in O'Bannon et al. (1989).

[a]Overt integrity test. [b]Personality-Based integrity test. [c]No validity data was reported, but the test contributed to the statistical artifact distributions.

Table 2

Proposed Moderator Analyses for Integrity Test Validities in Predicting Job Performance and Counterproductive Behaviors

1. Predictor type (overt vs. personality-based).[a,b]
2. Job performance measurement method (supervisory ratings vs. production records)[a]
3. Counterproductive behaviors measurement method (admissions vs. external).[b]
4. Breadth of criteria (narrow vs. broad counterproductivity).[b]
5. Validation strategy (predictive vs. concurrent).[a,b]
6. Validation sample (applicants vs. employees).[a,b]
7. Job complexity (high, medium, low).[a,b]

[a]Proposed moderator applicable to the criterion of job performance. [b]Proposed moderator applicable to the criterion of counterproductive behaviors.

Table 3

Descriptive Information on Statistical Artifact Distributions Used to Correct Validities

| | No. of Values | Mean | Standard deviation | Mean of the square roots of reliabilities | Standard deviation of the square roots of reliabilities |
|---|---|---|---|---|---|
| **Integrity test reliabilities** | | | | | |
| Overall distribution | 124 | .81 | .11 | .90 | .06 |
| Overt | 97 | .83 | .09 | .91 | .05 |
| Personality-Based | 27 | .72 | .13 | .85 | .08 |
| **Criterion reliabilities** | | | | | |
| Job performance | 163[a] | .54 | .09 | .73 | .05 |
| Production records | 10 | .89 | .05 | .94 | .03 |
| Supervisory ratings of overall job performance | 1 | .52 | - | .72 | - |
| Counterproductive behaviors | 171[b] | .69 | .09 | .83 | .05 |
| **Artifact distribution for range restriction correction** | | | | | |
| U[c] | 79 | .81 | .19 | - | - |

[a] The reliability of supervisory ratings of overall job performance of .52 was assigned a frequency of 153 and was combined with 10 reliabilities for production records; [b] 13 unique reliabilities for counterproductive behaviors were assigned frequencies corresponding to the number of validities in the database using the same criterion ; [c] U refers to the ratio of the selected group standard deviation to the referent group standard deviation.

Table 4

Overall Meta-Analyses of the Validity of Integrity Tests

| Analyses categories | Total N | K | $r_{mean}$ | $SD_r$ | $\sigma_{res}$ | $\rho$ | $SD\rho$ | % Var. S.E. | % Var. acc for | 90% CV |
|---|---|---|---|---|---|---|---|---|---|---|
| 1.All integrity tests predicting overall job performance[a] | 63,500 | 222 | .21 | .1019 | .0701 | **.34** | .13 | 30.9 | 52.6 | .20 |
| 2.All integrity tests predicting counterproductive behaviors[b] | 384,293 | 443 | .33 | .2463 | .2345 | **.47** | .37 | 1.5 | 9.4 | .05 |

Note. K = number of correlations; $r_{mean}$ = mean observed correlation; $SD_r$ = observed standard deviation; $\sigma_{res}$ = residual standard deviation; $\rho$ = true validity; $SD\rho$ = true standard deviation; % Var. S.E. = % variance due to sampling error; % Var. acc. for =% variance due to all corrected statistical artifacts; 90% CV = lower 90% credibility value. [a]The criteria for validation include supervisory ratings of overall job performance, production records, and commendations. [b]The criteria include narrow and broad criteria of disruptive behaviors, such as actual theft, admitted theft, dismissals for actual theft, illegal activities, absenteeism, tardiness, and violence.

Table 5

Meta-Analyses of the Validity of Integrity Tests for Predicting Overall Job Performance: All Performance Criteria

| Analyses categories | Total N | K | $r_{mean}$ | $SD_\rho$ | $s_{res}$ | $\rho$ | SDr | % Var.S.E | % Var.acc. for | 90% CV |
|---|---|---|---|---|---|---|---|---|---|---|
| 1. a. Overt integrity tests[a] | 27,768 | 84 | .20 | .1093 | .0844 | **.33** | .15 | 23.3 | 40.5 | .16 |
| b. Personality-Based tests[a] | 35,732 | 138 | .22 | .0976 | .0591 | **.35** | .11 | 37.0 | 63.3 | .23 |
| 2. a. Supervisory ratings of overall job performance | 36,250 | 153 | .21 | .1039 | .0699 | **.35** | .13 | 35.9 | 54.7 | .20 |
| b. Production records | 2,210 | 10 | .22 | .1163 | .0846 | **.28** | .12 | 30.4 | 47.1 | .15 |
| 3. a. Concurrent validation[a] | 31,866 | 135 | .22 | .1051 | .0683 | **.37** | .12 | 34.8 | 57.7 | .22 |
| b. Predictive validation[a] | 30,150 | 79 | .19 | .0951 | .0687 | **.31** | .12 | 26.9 | 47.9 | .17 |
| 4. a. Applicant sample[a,b] | 24,264 | 43 | .24 | .0617 | 0 | **.40** | 0 | 41.3 | 100 | .40 |
| b. Employee sample[a] | 24,354 | 135 | .17 | .1274 | .0970 | **.29** | .18 | 32.3 | 42.0 | .08 |
| 5. a. Low complexity jobs[a] | 1,633 | 19 | .28 | .0902 | 0 | **.45** | 0 | 100 | 100 | .45 |
| b. Medium complexity jobs[a] | 14,701 | 80 | .19 | .1180 | .0831 | **.32** | .15 | 36.4 | 50.3 | .14 |
| c. High complexity jobs[a] | 754 | 11 | .28 | .1215 | 0 | **.46** | 0 | 85.0 | 100 | .46 |

Note. K = number of correlations; $r_{mean}$ = mean observed correlation; $SD_r$ = observed standard deviation; $s_{res}$ = residual standard deviation; $\rho$ = true validity; $SD_\rho$ = true standard deviation; % Var. S.E. = % variance due to sampling error; % Var. acc. for = % variance due to all corrected statistical artifacts; 90% CV = lower 90% credibility value.

[a]The criteria for validation include supervisory ratings of overall job performance, production records, and commendations. [b]These studies are predictive, with the exception of one study N = 27.

Table 6

Meta-Analyses of the Validity of Integrity Tests for Predicting Job Performance: Supervisory Ratings Only

| Analyses categories | Total N | K | $r_{mean}$ | $SD_r$ | $\sigma_{res}$ | $\rho$ | SD$\rho$ | % Var.S.E | % Var.acc. for | 90% CV |
|---|---|---|---|---|---|---|---|---|---|---|
| 1. a. Overt integrity tests | 10,045 | 51 | .18 | .1430 | .1189 | **.30** | .22 | 23.4 | 30.8 | .05 |
| b. Personaltiy-Based tests | 26,205 | 102 | .22 | .0811 | .0259 | **.37** | .05 | 53.6 | 89.8 | .32 |
| 2. a. Concurrent validation | 12,109 | 88 | .23 | .1109 | .0577 | **.39** | .11 | 53.2 | 73.0 | .27 |
| b. Predictive validation | 22,657 | 57 | .19 | .0968 | .0727 | **.32** | .13 | 24.9 | 43.6 | .17 |
| 3. a. Applicant sample | 6,955 | 25 | .25 | .0814 | .0252 | **.42** | .05 | 47.7 | 90.4 | .37 |
| b. Employee sample | 15,660 | 90 | .20 | .1278 | .0958 | **.33** | .18 | 32.7 | 43.8 | .13 |
| 4. a. Low complexity jobs | 1,333 | 16 | .28 | .0850 | 0 | **.46** | 0 | 100 | 100 | .46 |
| b. Medium complexity jobs | 7,014 | 45 | .22 | .1272 | .0905 | **.36** | .17 | 36.2 | 49.4 | .17 |
| c. High complexity jobs | 619 | 10 | .31 | .1185 | 0 | **.51** | 0 | 95.7 | 100 | .51 |

Note. K = number of correlations; $r_{mean}$ = mean observed correlation; $SD_r$ = observed standard deviation; $\sigma_{res}$ = residual standard deviation; $\rho$ = true validity; SD$\rho$ = true standard deviation; % Var. S.E. = % variance due to sampling error; % Var. acc. for = % variance due to all corrected statistical artifacts; 90% CV = lower 90% credibility value.

Table 7

Intercorrelations Between Moderators of Integrity Tests in Predicting Job Performance

| Moderator | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| 1. Test (overt vs. personality-based) | – | .15 | -.40 | -.09 |
| | | (214) | (179) | 109 |
| 2. Strategy (concurrent vs. predictive) | | – | -.58 | -.27 |
| | | | (171) | (106) |
| 3. Sample (applicants vs. employees) | | | – | .18 |
| | | | | (105) |
| 4. Job complexity (high, medium, low) | | | | – |

Note. The number of studies used in calculating the correlations (i.e., sample size) is indicated in parentheses below the correlation coefficient. All the moderators were dummy coded as follows: 1 Type of test (overt = 1, personality-based = 2); 2 Validation strategy (concurrent = 1, predictive = 2); 3 Validation sample (applicants = 1, employees = 2); 4 Job complexity (high = 1,2, medium = 3, low = 4,5).

Table 8

Hierarchical Moderator Analyses of the Integrity Test Validities for Predicting Supervisory Ratings of Overall Job Performance

| | | Applicants | Employees |
|---|---|---|---|
| Total N | | 6,674 | 6,118 |
| K | | 23 | 20 |
| $r_{mean}$ | | .25 | .15 |
| $SD_r$ | | .0753 | .1318 |
| $\sigma_{res}$ | **Predictive** | 0 | .1146 |
| $\rho$ | | **.41** | **.26** |
| $SD_\rho$ | | 0 | .21 |
| % Var | | 100 | 24.4 |
| 90% C.V. | | .41 | .01 |
| Total N | | 27 | 8,264 |
| K | | 1 | 63 |
| $r_{mean}$ | | .29 | .22 |
| SDr | | - | .1227 |
| $\sigma_{res}$ | **Concurrent** | - | .0766 |
| $\rho$ | | **.48** | **.37** |
| $SD_\rho$ | | - | .14 |
| % Var | | - | 61.0 |
| 90% C.V. | | - | .21 |

Note. N=Total sample size; K = number of correlations; $r_{mean}$ = mean observed correlation; $SD_r$ = observed standard deviation; $\sigma_{res}$ = residual standard deviation; $\rho$ = true validity; $SD\rho$ = true standard deviation; % Var = % variance due to all corrected statistical artifacts; 90% CV = lower 90% credibility value.

Table 9

Moderator Analyses for Predicting Counterproductive Behaviors

| Analyses categories | Total N | K | $r_{mean}$ | $SD_r$ | $\sigma_{res}$ | $\rho$ | $SD_\rho$ | % Var. S.E. | % Var. acc. for | 90% CV |
|---|---|---|---|---|---|---|---|---|---|---|
| 1.a. Overt integrity tests[a] | 242,967 | 305 | .39 | .2835 | .2710 | .55 | .41 | 1.1 | 8.6 | .07 |
| b. Personality-Based tests[a] | 141,326 | 138 | .22 | .0884 | .0663 | .32 | .11 | 11.3 | 43.7 | .20 |
| 2. a.Admissions of counterproductivity[a] | 219,479 | 255 | .41 | .2730 | .2589 | .58 | .40 | 1.1 | 10.1 | .11 |
| b. Externally measured counterproductivity[a] | 164,674 | 187 | .22 | .1490 | .1369 | .32 | .22 | 4.5 | 15.6 | .07 |
| 3. a. Theft (narrow criteria)[b] | 103,258 | 152 | .36 | .2654 | .2523 | .52 | .39 | 1.6 | 9.6 | .06 |
| b.Broad counterproductivity[c] | 279,805 | 290 | .32 | .2382 | .2267 | .45 | .36 | 1.5 | 9.4 | .04 |
| 4. a.Concurrent validation[a] | 212,385 | 295 | .39 | .2680 | .2539 | .56 | .39 | 1.4 | 10.2 | .10 |
| b.Predictive validation[a] | 166,404 | 138 | .25 | .1885 | .1785 | .36 | .28 | 2.1 | 10.4 | .03 |
| 5. a.Applicant samples[a] | 258,034 | 183 | .30 | .2314 | .2207 | .44 | .35 | 1.1 | 9.0 | .04 |
| b.Employee samples[a] | 93,914 | 153 | .38 | .3120 | .3003 | .54 | .47 | 1.2 | 7.4 | .02 |
| 6. a.Low complexity jobs[a] | 9,654 | 44 | .30 | .1836 | .1607 | .43 | .25 | 11.3 | 23.4 | .13 |
| b.Medium complexity jobs[a] | 19,866 | 78 | .28 | .1731 | .1513 | .40 | .24 | 11.2 | 23.6 | .13 |
| c.High complexity jobs[a] | 2,246 | 21 | .49 | .1751 | .1295 | .68 | .20 | 17.9 | 45.3 | .45 |

Note. K = number of correlations; $r_{mean}$ = mean observed correlation; $SD_r$ = observed standard deviation; $\sigma_{res}$ = residual standard deviation; $\rho$ = true validity; $SD\rho$ = true standard deviation; % Var. S.E. = % variance due to sampling error; % Var. acc. for = % variance due to all corrected statistical artifacts; 90% CV = lower 90% credibility value.

[a]The criteria include narrow and broad criteria of disruptive behavior such as actual theft, admitted theft, dismissals for actual theft, illegal activities, absenteeism, tardiness, and violence. [b]The criteria include admissions of theft, actual theft, and dismissals for actual theft. [c]The criteria include violence on the job, tardiness, absenteeism, and other disruptive behaviors not included in b.

Table 10

Intercorrelations Between Moderators of Integrity Tests in Predicting Counterproductive Behaviors

| Moderator | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| 1. Test | – | .56 | .44 | .42 | -.19 | .16 |
| (overt vs. personality-based) | | (442) | (443) | (433) | (409) | (143) |
| 2. Criterion measurement method | | – | .38 | .74 | .22 | .00 |
| (admissions vs. external) | | | (442) | (433) | (408) | (142) |
| 3. Criterion breadth | | | – | .24 | .00 | .05 |
| (theft vs. broad counterproductivity) | | | | (433) | (409) | (143) |
| 4. Strategy | | | | – | .10 | .04 |
| (concurrent vs. predictive) | | | | | (402) | (142) |
| 5. Sample | | | | | – | -.11 |
| (applicants vs. employees) | | | | | | (138) |
| 6. Job complexity | | | | | | – |
| (high, medium, low) | | | | | | |

Note. The number of studies used in calculating the correlations (i.e., sample size) is indicated in parentheses below the correlation coefficient. All the moderators were dummy coded as follows: 1 Type of test (overt = 1, personality-based = 2); 2 Validation strategy (concurrent = 1, predictive = 2); 3 Validation sample (applicants = 1, employees = 2); 4 Job complexity (high = 1,2, medium = 3, low = 4,5).

Table 11

Fully Hierarchical Moderator Analyses of the Validity of Integrity Tests for Predicting Counterproductive Behaviors

| | | External criteria | | | | | | | | Admissions criteria | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Theft criteria | | | | Broad criteria | | | | Theft criteria | | | | Broad criteria | | | |
| | | Predictive | | Concurrent | | Predictive | | Concurrent | | Predictive | | Concurrent | | Predictive | | Concurrent | |
| | | App | Ees | App | Ees | App | Ees | App | Ees | App | Ees | App | Ees | App | Ees | App | Ees |
| Total N | | 1,031 | 5,695 | | | 1,916 | 10,842 | 213 | 6,813 | | | 67,618 | 2,917 | | | 85,824 | 27,887 |
| K | | 7 | 11 | | | 10 | 23 | 2 | 14 | | | 63 | 34 | | | 24 | 46 |
| $r_{mean}$ | | .09 | .11 | | | .27 | .06 | .22 | .71 | | | .30 | .38 | | | .32 | .76 |
| $SD_r$ | | .1152 | .1049 | | | .1218 | .1192 | .1597 | .2336 | | | .2235 | .1644 | | | .2336 | .1346 |
| $\sigma_{res}$ | **Overt** | .0781 | .0923 | | | .0837 | .1091 | .1208 | .2072 | | | .2128 | .1125 | | | .2233 | .0771 |
| $\rho$ | | **.13** | **.16** | | | **.39** | **.09** | **.32** | **.94** | | | **.42** | **.54** | | | **.46** | **.99** |
| $SD_\rho$ | | .12 | .15 | | | .13 | .17 | .19 | .29 | | | .33 | .17 | | | .35 | .11 |
| % Var | | 54.0 | 22.6 | | | 52.7 | 16.2 | 42.7 | 21.3 | | | 9.3 | 53.2 | | | 8.6 | 67.2 |
| 90% C.V. | | -.01 | -.01 | | | .23 | -.11 | .10 | .59 | | | .04 | .34 | | | .06 | .86 |
| Total N | | | | | | 76,835 | 37,415 | 4,261 | 1,511 | | | | | | | | 210 |
| K | | | | | | 62 | 5 | 6 | 12 | | | | | | | | 2 |
| $r_{mean}$ | | | | | | .20 | .18 | .57 | .20 | | | | | | | | .16 |
| $SD_r$ | | | | | | .0555 | .0118 | .0519 | .1033 | | | | | | | | .1000 |
| $\sigma_{res}$ | **Personality** | | | | | .0115 | 0 | 0 | .0339 | | | | | | | | 0 |
| $\rho$ | **Based** | | | | | **.29** | **.26** | **.77** | **.29** | | | | | | | | **.23** |
| $SD_\rho$ | | | | | | .02 | 0 | 0 | .06 | | | | | | | | 0 |
| % Var | | | | | | 95.7 | 100 | 100 | 89.3 | | | | | | | | 100 |
| 90% C.V. | | | | | | .27 | .26 | .77 | .23 | | | | | | | | .23 |

Table 11 (Continued)

Fully Hierarchical Moderator Analyses of the Validity of Integrity Tests for Predicting Counterpreoductive Behaviors

Note. App = applicants; Ees = employees; K = number of correlations; $r_{mean}$ = mean observed correlation; $SD_r$ = observed standard deviation; $\sigma_{res}$ = residual standard deviation; $\rho$ = true validity; $SD_\rho$ = true standard deviation; % Var = % variance due to all corrected statistical artifacts; 90% CV = lower 90% credibility value. This table represents the following moderators being taken into consideration simultaneously; predictor type, criterion measurement method, breadth of criteria, validation strategy, and validation sample.

[a]The criteria include admissions of theft, dismissals for actual theft using self-report measures. [b]The criteria include violence on the job, tardiness, absenteeism, and other behaviors not included in a using self-report measures. [c]The criteria are actual theft and dismissals for theft using external measures. [d]The criteria include violence on the job, tardiness, absenteeism, and other disruptive behaviors not included in c using external measures.

Table 12

Effect of Combining Integrity Tests with Measures of Ability

| Job Complexity Level [a] | Validity of General Mental Ability[b] (GMA) | Validity of Psychomotor Ability[b] (PA) | Validity of Integrity[c] (I) | Multiple R's | |
|---|---|---|---|---|---|
| | | | | GMA + I | GMA + PA + I |
| Complexity level 1 (Highest) | .56 | .30 | .41 | **.69** | **.72** |
| Complexity level 2 | .58 | .21 | .41 | **.71** | **.71** |
| Complexity level 3 | .51 | .32 | .41 | **.65** | **.67** |
| Complexity level 4 | .40 | .43 | .41 | **.57** | **.65** |
| Complexity level 5 (Lowest) | .23 | .48 | .41 | **.47** | **.64** |

Note. The multiple correlations reported in this table are computed assuming that both general mental ability and psychomotor ability correlate zero with integrity.

[a]Job complexity levels are those used by Hunter (1980). [b]Validities are from Hunter (1980). [c]Predictive validity of integrity tests for supervisory ratings of overall job performance calculated using applicants.

## Studies Coded for the Meta-analysis of Integrity Tests


Alvord, G. (1985). Validity of the Personnel Selection Inventory as a screening test for bus operators. Journal of Security Administration, 8(1), 37-47.

Ash, P. (1970). The validation of an instrument to predict the likelihood of employee theft. Proceedings of the 78th Convention of the American Psychological Association, (pp. 579-580).

Ash, P. (1985). Predicting dishonesty with the Reid Report. Journal of the American Polygraph Association, 5, 139-145.

Ash, P. (1986). Reid Report Inventory reliability (Research Memorandum 3). Chicago: Reid Psychological Systems.

Ash, P. (1987). Honesty test scores, biographical data, and delinquency indicators. Paper presented in the Annual Meeting Of the Academy of Criminal Justice Sciences, St. Louis, MO.

Ash, P., Baehr, M.E., Joy, D.S., & Orban, J.A. (in press). Employment testing for the selection and evaluation of bus drivers. Applied Psychology: An International Review.

Berland, A.C. (1990). Psychometric assessment of honesty testing for personnel selection. Unpublished masters thesis, Iowa State university, Ames, IA.

Berte, D. L., Moretti, D. M., Jusko, R., & Leonard, J. (1984). An investigation of a combined withdrawal and counterproductive behavior decision model. Paper presented at the 2nd Annual National Graduate Student Convention in Industrial/ Organizational Psychology-Organizational Behavior. East Lansing, Michigan. (Also published as Technical Report No. 13 Park ridge, IL: London House Press.)

Borofsky, G.L. (1990). User's manual for the Employee Reliability Inventory. Boston, MA: Bay State Psychological Associated Inc.

Borofsky, G.L. Personal Communication. April 10, 1991.

Bradley, P. (1980). The Milby system: Validation study 1980-M., Minneapolis, MN: Milby Systems, Inc.

Bradley, P.A. (1984). Milby Profile Validity Study. Minneapolis: Milby Systems, Inc.

Brannick J. Personal Communication. April 29, 1991.

Brooks, P. (1990). Predictive relationship of the Reid Report to measures of employee work performance in a retail setting (Research Report). Chicago, IL: Reid Psychological Systems.

Brooks, P. (1990). Relationship of the Reid Report to performance ratings in a retail environment (Research Report no: 34). Chicago, IL: Reid Psychological Systems.

Brooks, P., & Arnold, D. (1989). Integrity attitudes as predictors of public safety organizational criteria (Research Report no.: 29). Chicago, IL: Reid Psychological Systems.

Brown, T.S., & Joy, D.S., (1985). The predictive validity of the Personnel Selection Inventory in the grocery industry. (Technical Report No. 48) Park Ridge, IL: London House Inc.

Carpenter, B.N. (undated). Reliability and motivational distortion in the Wilkerson pre-employment audit (Technical Report). Houston, TX: Team Building Systems.

Cherrington, D. (1989). The Applicant Review, (Administrator's Manual). Orem, UT: CHC Forecast Inc.

Collins, J.M. (1991). Personality, integrity, and white collar criminality. Unpublished manuscript.

Conoley, J.C., & Kramer, J.J. (Eds.). (1989). The Tenth Mental Measurements Yearbook. Lincoln, NE: Buros Institute of Mental Measurements.

Cooke, D.K. Bermardin, J.J., & Salerno, L.J. (1991, August). The predictive validity of an integrity test for convenience store employees. Paper presented at the Academy of Management meetings, Miami Beach, FL.

Cotton, J.R. (1990). A study to evaluate the use of the Phase II profile as an indicator of employment longevity for hospitality industry employees. Unpublished masters thesis. Johnson and Wales University, Providence, Rhode Island.

Cunningham, M.R. (1988). Summary statistics for the Reid Drug Scale across five validation samples (Research Memorandum 18). Chicago: Reid Psychological Systems.

Cunningham, M.R. (1989). Test taking motivations and outcomes on a standardized measure of on the job integrity. Journal of Business and Psychology, 4, 119-127.

Cunningham, M.R., Trucott, M., & Wong, D.T. (1990, August). An experimental investigation of integrity: Testing the predictive validity of the Reid Report. Paper presented at the Annual meeting of the American Psychological Association, Boston, MA.

Debro, J. & Murty, K.S. (undated). Validity and reliability of test scores of measuring integrity and honesty. Georgia: The Atlanta University Criminal Justice Institute.

Durbrow, B.R. (1983). Comparative summary of results of concurrent and predictive studies using the Accutrac Evaluation System. Cincinnati, OH: Barbrisons Management Systems, Inc.

Durbrow, B.R. (1985). Accutrac Evaluation System predictive validation study and minority adverse impact study for critical potentials measures. Cincinnati, OH: Barbrisons Management Systems, Inc.

Durbrow, B.R. (undated) Summary of study results. Cincinnati, OH: Barbrisons Management Systems, Inc.

Frost, A.G., & Rafilson, F.M. (1989). Overt integrity tests versus personality-based measures of delinquency: An empirical comparison. Journal of Business and Psychology, 3(3), 269-277.

Gerhardt, V. (undated). Research and statistics: Pre-employment analysis questionnaire "H", Dallas,TX: Psychometric Behavioral Group.

Gerstein, L.H., Broorke, C.R., & Johnson, S.D. (1989). Internal validity studies of a telephone pre-employment measure. Journal of Employment Counselling, 26, 77-83.

Gough, H.G. (1971). The assessment of wayward impulse by means of the Personnel Reaction Blank. Personnel Psychology, 24, 669-667.

Gough, H.G. (1972). Manual for the Personnel Reaction Blank. Palo Alto, CA: Consulting Psychologist Press.

Grimsely, D.L. (1986). Frequency of admitted termination of employment on a pre employment honesty test. Paper presented at the Annual meeting of the North Carolina Psychological Association, Myrtle Beach, SC.

Grimsley, D.L. (1986). Terst-retest reliability of the Stanton Survey integrity measure. Charlotte, NC: Stanton Corporation.

Guier, L.R. (1984). Validation of personality measures for selection of psychiatric counselors. Tulsa, University of Tulsa.

Harris, M.M. (1980). A research statement on the Reid report. Unpublished manuscript, University of Illinois at Chicago, Chicago.

Harris, W.G. (1987). A components analysis of a pre-employment integrity measure: A replicated study. Psychological Reports, 60, 1051-1055.

Harris, W.G. (1988). The Stanton Profile: Test manual. Charlotte, NC: Stanton Corporation.

Harris, W.G. (undated). An investigation of the Stanton Survey using a validity generalization model. Charlotte, NC: Stanton Corporation.

Helm and Associates. (1985). Validation of the RELY: An inventory designed to measure job applicants attitudes toward good work ethic, absenteeism, and punctuality (Technical Report no. 1409). Helm and Associates.

Hogan, J. & Hogan, R. (1986). Hogan Personnel Selection Series manual. Minneapolis, MN: National Computer Systems.

Hogan, J., & Hogan, R. (1989). How to measure employee reliability. Journal of Applied Psychology, 74, 273-279.

Hogan, J., Arneson, S., Hogan, R., & Jones, S. (1986). Development and validation of personnel selection procedures for the job rehabilitation therapist. Tulsa, OK: Hogan Assessment Systems.

Hogan, J., Hogan, R., & Briggs, S. (1984). Development and validation of employee selection tests for combination drivers. Unpublished manuscript, University of Tulsa, Tulsa:OK.

Hogan, J., Peterson, S., Hogan, R., Jones, S. (1985). Development and validation of a line haul driver selection inventory. Tulsa, OK University of Tulsa.

Hogan, R. (1991). [Correlations Between the HPI Reliability Scale (REL) and Various Scales]. Unpublished raw data.

Hogan, R. (undated). Profiles interpretive manual. Tulsa, OK: Hogan Assessments Systems.

Hogan, R., Jacobson, G.K., Hogan, J., Thompson, B. (1985). Development and validation of a service operations dispatcher selection battery. Tulsa, OK: Hogan Assessment Systems.

Hough, L.M. (1990). The Personnel Reaction Blank (1988 ed.). In J. Hogan and R. Hogan (eds.), Business and Industry Testing (pp. 434-440). Austin, TX: Proed.

Inwald, Robin E. & Brockwell, Albert L. (1989). Predicting the Performance of Government Security Personnel with the IPI & MMPI. New York: Hilson Research, Inc.

Inwald, Robin E. & Knatz, Hilary F. (1988). Seven-Year Follow-up of Officer Terminations Predicted by Psychological Testing. New York: Hilson Research, Inc.

Inwald, Robin E. & Patterson, Terence. (1989). Use of Psychological Testing to Predict Performance of Law Enforcement Trainees. New York: Hilson Research, Inc.

Inwald, Robin E. & Shusman, Elizabeth J. (1984). The IPI and MMPI as Predictors of Academy Performance for Police Recruits. Journal of Police Science and Administration, 12(1), 1-11.

Inwald, Robin E. (1988). Five-Year Follow-up Study of Departmental Terminations as Predicted by 16 Pre-employment Psychological Indicators. Journal of Applied Psychology, 73(4), 703-710.

Inwald, Robin E. (1990b). Predicting the Performance of Private Security Personnel with the 153 & HPP/SO (research abstract). New York: Hilson Research, Inc.

Inwald, Robin E. (1990c). IPI Predictions of Private Security Personnel Performance (research abstract). New York: Hilson Research, Inc.

Inwald, Robin E., Knatz, Hilary F. & Shusman, Elizabeth J. (1982). Inwald Personality Inventory Technical Manual. New York: Hilson Research, Inc.

Jones J.W. (Ed). (1991). Pre-employment honesty testing. NY: quorum books.

Jones J.W., Joy, D.S., & Martin S.L. (1990). A multidimensional approach for selecting childcare workers. Psychological Reports, 67, 543-553.

Jones, J. W. & Fay, L. (1987). Predicting child abuse potential with the Personnel Selection Inventory for childcare workers. St. Paul, MN: The St. Paul Companies

Jones, J. W. (1979). Employee Deviance: Attitudinal Correlates of theft and on the job alcohol abuse. Paper presented at the 5th Annual Convention of the Society of Police and Criminal Psychology. Chicago, IL. (Also published as Technical Report No. 5 Park Ridge: IL: London House Press.)

Jones, J. W. (1980). Attitudinal correlates of employee deviance: Theft, alcohol use and nonprescribed drug use. Psychological Reports, 47, 71-77.

Jones, J.W. (1980). Attitudinal correlates of employee violence. Journal of Security Administration, 3(1), 42-47.

Jones, J. W. (1981a). Attitudinal Correlates of employee theft of drugs and hospital supplies among nursing personnel. Nursing Research, 30, 349-351.

Jones, J. W. (1981b). Dishonesty, burnout and unauthorized work break extensions. Personality and Social Psychology Bulletin, 7, 406-409.

Jones, J.W. (1981c). Dishonesty, staff burnout, and employee theft. Paper presented at the 89th Annual Conference of the American Psychological Association, the Division of Industrial Psychology, Los Angeles.

Jones, J.W. (1981d). Personality profiles of endorsers of nuclear crime. Paper presented at the 89th Annual convention of the APA, The Division of Military Psychology, Los Angeles, CA.

Jones, J.W. (1982a). Correlates of police misconduct: Violence and alcohol use on the job. Paper presented at the annual meeting of the Society and Criminal Psychology. Atlanta, GA.

Jones, J.W. (1982b). Dishonesty, staff burnout, and employee theft. (Technical Report No. E2). Park Ridge, IL: London House Press.

Jones, J.W. (1982c). Psychological predictors of employee theft. Paper presented at the 90th Annual Conference of the American Psychological Association, Washington, D. C. (Also published as Technical Report No. 17. Park Ridge, IL: London House Press.)

Jones, J.W. (1983a). Validation of the employee attitude inventory with retail employees. Technical Research Report, Park Ridge, IL: London House Press.

Jones, J.W. (1983b). The employee attitude inventory: A validity study on employees' theft in the auto parts industry. Technical Research Report, Park Ridge, Illinois: London House Press.

Jones, J.W. (1984). Psychologically identifying current employee thieves (Technical Report 36). Park Ridge, IL: London House, Inc.

Jones, J.W. (1985). Dishonesty, staff burnout, and employee theft. In W. Terris (Ed.), Employee theft: Research, theory, and applications, Park Ridge, IL: London House Press.

Jones, J.W. (1989). Predicting police behavior with the Personnel Selection Inventory. Park Ridge, IL: London House, Inc.

Jones, J.W. (in press). Psychological predictors of employee theft. Journal of Security Administration.

Jones, J. W., & Joy, D.S. (1987). Use of the Personnel Selection Inventory for childcare workers: A summary. (Technical Report No. 66). Park Ridge, IL: London House Press.

Jones, J.W. & Terris, W. (1982a). Attitudes toward theft as predictors of employee theft among salesclerks. Paper presented at the 90th Annual Convention of the American Psychological Association, Washington, D.C.

Jones, J.W. & Terris, W. (1982b). The use of the Personnel Selection Inventory to predict turnover in a group of security guards: A pilot study (Technical Report 21). Park Ridge, IL: London House, Inc.

Jones, J.W. & Terris, W. (1984). Psychological correlates of employee theft in department stores. (Technical Report 20). Park Ridge, IL: London House, Inc.

Jones, J.W. & Wuebker, L.J. & Dubois, D. (1984). The Safety Locus of Control Scale (SLC): A validity study. Paper presented at the American Industrial Hygiene Association Research Conference, Upper Midwest Section, Roseville, MN.

Jones, J.W. & Wuebker, L.J. (1988a). Accident prevention through personnel selection. Journal of Business and Psychology, 3(2), 187-198.

Jones, J.W. & Wuebker, L.J. (1988b). Pre-employment screening to control hospital crime and counterproductivity. Readings in Health Care Risk Management, 93-96.

Jones, J.W., & Scruggs, D. P. (1981). Psychologically profiling endorsers of nuclear crime and sabotage. Paper presented at the Annual Meeting of the Midwestern Psychological Association, Detroit, MI.

Jones, J.W., & Terris, W. (1981). Predictive validation of a dishonesty test that measures theft proneness. Paper presented at the 18th Inter American Congress of Psychology, Santo Domingo, The Dominican Republic.

Jones, J.W., & Terris, W. (1982a). The Employee Attitude Inventory: A validity study on theft by current employees. Paper presented at the Annual Conference of the Society of Police and Criminal Psychology, Nashville,TN.

Jones, J.W., & Terris, W. (1982b). Using the PSI to reduce employee theft in department stores: A program evaluation study. (Technical Report No. 28) Park Ridge, IL London House Press.

Jones, J.W., & Terris, W. (1983a). Personality correlates of theft and drugs abuse among job applicants. Proceedings of the 3rd Annual Conference on the 16PF Test. pp. 85-94.

Jones, J.W., & Terris, W. (1983b). Predicting employees' theft in home improvement centers. Psychological Reports, 52. 187-201.

Jones, J.W., & Terris, W. (1983c). Use of the Employee Attitudinal Inventory to assess dishonesty among hotel employees. Technical Research Report, Park Ridge, Illinois: London House Press.

Jones, J.W., Joy, D.S., Werner, S.H., & Orban, J.A. (1991). Criterion related validity of a pre-employment integrity inventory: A large scale between groups comparison. Perceptual and motor skills, 72, 131-136.

Jones, J.W., Terris, W., & Allen, D.C. (1984). Use of the Employee Attitude Inventory for re-employment.Technical Research Report, Park Ridge, Illinois: London House Press.

Joy, D.S. & Frost, A.G. (1987). Validation of the Personnel Selection Inventory: Form 7S. Park Ridge, IL: London House, Inc.

Joy, D.S., & Frost, A.G. (1987). The prediction of employee performance using the employment productivity index. Park Ridge, IL: London House Inc.

Joy, D.S., & Frost, A.G., & Boomhower, D. (1987). Using the PSI (PSR-7S) to predict successful performance of employees in the automobile rental industry. Technical Research Report No.55, Park Ridge, Illinois: London House Press.

Joy, D.S., & Frost, A.G., & Cook, M.S. (1987). Predicting counterproductive behaviors at a major metropolitan transit authority using the PSI.Technical Research Report No.63, Park Ridge, Illinois: London House Press.

Joy, D.S., & Rafilson, F.M. (1988a). Validation of the PSI for consumer electronics stores. Technical Research Report No.71, Park Ridge, Illinois: London House Press.

Joy, D.S., & Rafilson, F.M. (1988b). Validation of the PSI (PSI-7S) for the grocery industry. Technical Research Report No.68, Park Ridge, Illinois: London House Press.

Joy, D.S., & Werner, S.H. (1988). The concurrent validity of the Personnel Selection Inventory in a major drug store chain. Technical Research Report No.72, Park Ridge, Illinois: London House Press.

Joy, D.S., & Werner, S.H. (1990). Effectiveness of the PSI in the identification of current drug use: A comparison with a medical drug screening procedure. Unpublished Technical Report, Park Ridge, Illinois: London House Press.

Kamp, J. (1988). Relationship of the Reid Report to inventory and turnover among convenience store managers (Research Report no. 12). Chicago, IL: Reid Psychological Systems.

Klump, C.(1980). The Stanton Survey Manual: Description and Validation. Chicago, IL: Stanton Corporation.

Knatz, Hilary F., Inwald Robin E. Brockwello, Albert L. & Tran, Linn N. (1989). IPI and MMPI Predictions of Counterproductive Job Behaviors by Group. New York: Hilson Research, Inc.

Kpo, W.R., & Harris, W.G. (1986). Generalizability of an attitudinal behavioral measure of integrity. Paper presented at the Annual Meeting of the New England Psychological Association, Boston: MA.

Lavelli, M., Terris, W., & Jones, J.W. (1985). Psychological predictors of employee counterproductivity. Unpublished master's thesis, DePaul University, Chicago, IL.

London House Press. (1980). Personnel Selection Inventory Manual, Park Ridge, IL: London House Press.

London House Press. (1982). The Employee Attitude Inventory Manual. Park Ridge, IL: London House Press.

Loudermilk, K.M. (1966). Prediction of efficiency of lumber and paper mill employees. Personnel Psychology, 19, 301-310.

Lousig-Nont, G.M. (1982b). Phase II Profile statistical validation studies. Las Vegas, NV: Lousig-Nont & Associates.

Malin, Sharon Z., Luria, Jeffrey & Morgenbesser, Leonard I. (1987). New York State Pre-Employment Psychological Screening Program: Longitudinal Validation Study. Paper presented at the American Psychological Association, New York, NY.

Martelli, T.A. (1988). Pre employment screening for honesty: The construct validity, criterion related validity and test-retest reliability of a written integrity test. Unpublished doctoral dissertation, Ohio University, Columbus, OH.

McDaniel, M.A., & Jones, J.W. (1986). A meta-analysis of the Employee Attitude Inventory theft scales. Journal of Business and Psychology, 1, 31-50.

McDaniel, M.A., & Jones, J.W. (1988). Predicting employee theft: A quantitative review of the validity of a standardized measure of dishonesty. Journal of Business and Psychology, 2, 327-345.

Montgomery, J, Butler, S., & McPhail, S.M. (1987). Development and validation of a selection test battery radiation protection technicians and chemical technicians. Houston, TX: Jeanneret and Associates.

Moretti, D. M. (1980). Employee counterproductivity: Attitudinal predictors of industrial damage and waste. Paper presented at the Annual Meeting of the Society of Police and Criminal Psychology, Atlanta, GA.

Moretti, D. M. (1982). The prediction of employee counterproductivity through attitude assessment. (Technical Report No. 34) Park Ridge, IL: London House Press.

Moretti, D. M. (1983). A note on the validity of the Personnel Selection Inventory distortion scale (Technical Report 33). Park Ridge, IL: London House, Inc.

Moretti, D. M. (1984). The prediction of employee performance in a New England retail chain. Park Ridge, IL: London House Press.

Moretti, D. M. (1986). The prediction of employee productivity through attitude assessment. Journal of Business and Psychology,1, 134-147.

Moretti, D.M., & Terris, W. (1983). The personnel selection inventory:A predictive validation study. Park Ridge IL: London House Press.

Morey, L. (1981). Statistical properties of the Wilkerson pre employment audit. Tulsa, OK: Wilkerson and Associates Ltd.

Muchinsky, P. (1987). Validation documentation for the development of personnel selection test batteries for telecommunications service jobs.. Ames: Iowa State University.

Nolan, K.A. (1991). The relationship between socially desirable responding and personal integrity: the unique effect of self-deception and impression management. Unpublished doctoral dissertation. Georgia State University, GA.

O'Bannon, R.M., Goldinger, L.A., & Appleby, G.S. (1989). Honesty and integrity testing. Atlanta, GA: Applied Information Resources.

Paajanen, G. E. (1987). The prediction of counterproductive behavior by individual and organizational variables. Unpublished doctoral dissertation, University of Minnesota, Minneapolis, MN.

Paajanen, G. E. (1987). The prediction of counterproductive behavior by individual and organizational variables. Unpublished doctoral dissertation, University of Minnesota.

Paajannen, G. E. (1990). Validity of the PDI employment inventory for drug stores. (Technical Report). Minneapolis, MN: Personnel Decisions Inc.

Paajannen, G. E. Personal Communication. June 1, 1991.

Personnel Decisions Incorporated (1985). Development and validation of the PDI Employment Inventory. Minneapolis: Author.

Personnel Decisions Incorporated (1986a). Summary of findings - retail industry, 1986. Minneapolis: Author.

Personnel Decisions Incorporated (1986b). Validation of the PDI Employment Inventory for store security officers. Minneapolis: Author.

Personnel Decisions Incorporated (1986c). Validation of the PDI Employment Inventory for transit commission. Minneapolis: Author.

Personnel Decisions Incorporated (1987a). Retail department store evaluation of the PDI employment Inventory. Minneapolis, MN: Personnel Decisions Inc.

Personnel Decisions Incorporated (1987b). Validation of the PDI Employment Inventory for fast food restaurants. Minneapolis: Author.

Personnel Decisions Incorporated (1987c). Validation of the PDI Employment Inventory for health care center. Minneapolis: Author.

Personnel Decisions Incorporated (1987d). Validation of the PDI Employment Inventory for retail department stores. Minneapolis: Author.

Personnel Decisions Incorporated (1988a). Validation of the PDI Employment Inventory for airline baggage employees. Minneapolis: Author.

Personnel Decisions Incorporated (1988b). Validation of the PDI Employment Inventory for amusement parks. Minneapolis: Author.

Personnel Decisions Incorporated (1988c). Validation of the PDI Employment Inventory for bank tellers. Minneapolis: Author.

Personnel Decisions Incorporated (1988d). Validation of the PDI Employment Inventory for courier service. Minneapolis: Author.

Personnel Decisions Incorporated (1988e). Validation of the PDI Employment Inventory for department stores. Minneapolis: Author.

Personnel Decisions Incorporated (1988f). Validation of the PDI Employment Inventory for discount department stores. Minneapolis: Author.

Personnel Decisions Incorporated (1988g). Validation of the PDI Employment Inventory for food wholesaler. Minneapolis: Author.

Personnel Decisions Incorporated (1988h). Validation of the PDI Employment Inventory for home improvement centers. Minneapolis: Author.

Personnel Decisions Incorporated (1988i). Validation of the PDI Employment Inventory for Mexican restaurants. Minneapolis: Author.

Personnel Decisions Incorporated (1988j). Validation of the PDI Employment Inventory for seafood restaurants. Minneapolis: Author.

Personnel Decisions Incorporated (1988k). Validation of the PDI Employment Inventory for supermarkets. Minneapolis: Author.

Personnel Decisions Incorporated (1988l). Validation of the PDI Employment Inventory for toy stores. Minneapolis: Author.

Personnel Decisions Incorporated (1988m). Validation of the PDI Employment Inventory for video stores. Minneapolis: Author.

Personnel Decisions Incorporated (1988n). Validation of the PDI Employment Inventory for petroleum station stores. Minneapolis: Author.

Personnel Decisions Incorporated (1988o). Validation of the PDI Employment Inventory for national discount fashion retailer. Minneapolis: Author.

Personnel Security Corporation (undated). Survey Description Manual: T.A. Survey. Oak Brook, IL: Personnel Security Corporation.

Personnel Systems Corporation, (1987). PASS III Data Survey implementation Manual. Oak Brook : IL.

Phase II Profile. (1982). Statistical Validation Studies. Peoria, IL: Lousing Nont and Associates.

Psychological Surveys Corporation. Validity, reliability, compliance, and efficacy studies of the TA, AI, ES, DA, and PASS Surveys (Technical Report). OakBrook, IL: Psychological Surveys Corporation.

Rafilson, F.M. & Callans, M.C. (1991). Validity, classification, and response distortion in attitudinally based and criterion based predictors of honest workplace behavior: A preliminary study. Unpublished manuscript.

Rafilson, F.M. (1988). Development of a standardized measure to predict employee productivity. Journal of Business and Psychology, 3, 199-213.

Rafilson, F.M., & Frost, A.G. (1989). Overt integrity tests versus personality-based measures of delinquency: An empirical comparison. Journal of Business and Psychology, 3, 269-277.

Raza, S., Metz, D., Dyer, P., Coan, T., & Hogan, J. (1986). Development and validation of personnel selection procedures for hospital service personnel. Tulsa, OK: University of Tulsa.

Reed, H. (1982). The Stanton Survey: Description and Validation manual. Chicago, IL: Stanton Corporation.

Reid Psychological Systems. (1989). Reid Report: Examiner's Manual (3rd ed.). Chicago, IL: Reid Psychological Systems.

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, 248-256.

Sackett, P.R., & Harris, M.M. (1984). Honesty testing for personnel selection: A review and critique. Personnel Psychology, 37, 221-246.

Sackett, P.R., Burris, L.R., & Callahan, C. (1989). Integrity testing for personnel selection: An update. Personnel Psychology, 42, 491-529.

Science Research Associates, Inc. (1987). Personal Outlook Inventory (POI) examiner's manual. Chicago: Author.

Selection Research Publishing. (1983). Development and validation of the Personnel Outlook Inventory. Chicago, IL: Selection Research Publishing.

Sevy, B.A. (1987). Validity of the PDI employment inventory for bus driver selection. Minneapolis, MN: Personnel Decisions Inc.

Shusman, Elizabeth J. & Inwald, Robin E. (1984). A Longitudinal Validation Study of Correction Officer Job Performance as Predicted by the IPI and MMPI. Paper presented at the American Psychological Association Meeting, Toronto, August.

Shusman, Elizabeth J., Inwald, Robin E. & Knatz, Hilary F. (1987). A Cross-Validation Study of Police Recruit Performance as Predicted by the IPI and MMPI. Journal of Police Science and Administration. 15(2), 162-169.

Shusman, Elizabeth J., Inwald, Robin E. & Landa, Beth. (1984). Correction Officer Job Performance as Predicted by the IPI and MMPI: A Validation and Cross Validation Study. Criminal Justice and Behavior. 11(3), 309-329.

Shyman, J.H. (1990). Fakability of pre-employment paper and pencil honesty tests. Unpublished masters thesis, Radford University, VA.

Slora, K. B. , Joy, D.S. , Jones, J. W. ,& Terris, W. (1991). The prediction of on the job violence. In J.W. Jones (Ed.), Preemployment Honesty Testing: Current Research and Future Directions. Westport, CT: Quorum Books.

Springborn, B.A. (undated). [Concurrent validity study of the Phase II Profile ADdendum]. Lakeland, FL: Kelly/Springborn Associates, Inc.

Statistical Consulting Associates. (1986). Preliminary validation of the drug and alcohol usage items from the Wilkerson Pre-Employment Audit. (Available from Team Building Systems, Houston, Texas).

Terris, W. (1979). Attitudinal correlates of theft, violence and drug use. Paper presented at the 17th Congress of Psychology, Lima, Peru.

Terris, W. (1985). Attitudinal correlates of employment integrity. Journal of Police and Criminal Psychology, 1(1), 60-68.

Terris, W. (1986). The development and validation of EPI-3. Park Ridge IL: London House Inc.

Terris, W., & Jones, J.W. (1980). Attitudinal and personality correlates of theft among supermarket employees. Journal of Security Administration, 3, 65-78.

Terris, W., & Jones, J.W. (1982a). Pre employment screening to reduce employee theft in department stores: Three separate studies. Park Ridge, IL: London House Management Consultants.

Terris, W., & Jones, J.W. (1982b). Psychological correlates of employee theft in department stores. (Technical Report No. 20) Park Ridge IL: London House Inc.

Terris, W., & Jones, J.W. (1982c). Validation of the PSI with hourly employees in a fast food company. (Technical Report No. 22) Park Ridge IL: London House Inc.

U.S. Congress, Office of Technology Assessment. (1990). The use of integrity tests for pre-employment screening, OTA-SET-442. Washington, D.C.: U.S. Government Printing Office.

Weichman D. J., Bae, R.P. (1983). The Personnel Selection Inventory: An empirical examination. Technical Report No. 14. Park Ridge , IL: London House Inc.

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.

Wilkerson, O.M.(1980). The Wilkerson Pre-employment audit. Tulsa, OK: Wilkerson and Associates Ltd.

Woolley, R. (1991). An examination of the construct and criterion-related validity of overt and personality oriented predictors of counterproductivity. Unpublished masters thesis, University of British Columbia, Vancouver.