JOURNAL OF APPLIED PSYCHOLOGY MONOGRAPH

Comprehensive Meta-Analysis of Integrity Test Validities: Findings and Implications for Personnel Selection and Theories of Job Performance

Deniz S. Ones, Chockalingam Viswesvaran, and Frank L. Schmidt

The authors conducted a comprehensive meta-analysis based on 665 validity coefficients across 576,460 data points to investigate whether integrity test validities are generalizable and to estimate differences in validity due to potential moderating influences. Results indicate that integrity test validities are substantial for predicting job performance and counterproductive behaviors on the job, such as theft, disciplinary problems, and absenteeism. The estimated mean operational predictive validity of integrity tests for predicting supervisory ratings of job performance is .41. Results from predictive validity studies conducted on applicants and using external criterion measures (i.e., excluding self-reports) indicate that integrity tests predict the broad criterion of organizationally disruptive behaviors better than they predict employee theft alone. Despite the influence of moderators, integrity test validities are positive across situations and settings.

Over the last 10 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, Goldinger, and Appleby (1989) provided extensive qualitative reviews and critical observations regarding integrity testing. In addition to these reviews, the U.S. Congressional Office of Technology Assessment (U.S. OTA; 1990) and the American Psychological Association (APA; Goldberg, Grenier, Guion, Sechrest, & Wing, 1991) each released papers on integrity tests. In comparison with the U.S. OTA paper, the APA report is more thorough and provides a generally favorable conclusion regarding the use of paper-and-pencil integrity tests in personnel selection. The purpose of this article 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, No date) investigated the validity of the Stanton Survey. The second meta-analysis (McDaniel & Jones, 1986) examined the validity of the London House Employee Attitude Inventory (London House Press, 1982). Last, McDaniel and Jones (1988) focused on the validity of the Dishonesty scale of the Personnel Selection Inventory (PSI; London House Press, 1975) 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 in the three previous meta-analyses in this field. Our meta-analysis tests this hypothesis and provides the empirical evidence required to confirm or refute the notion that validity is specific to particular types of instruments, criteria, or validation strategies (concurrent or predictive). That is, this study uses meta-analysis to investigate whether integrity test validities are generalizable and quantitatively documents validity differences that may be the result of moderating influences.

The similarity of the different integrity tests raises the ques-

Deniz S. Ones, Chockalingam Viswesvaran (now at Department of Psychology, Florida International University), and Frank L. Schmidt, Department of Management and Organizations, University of Iowa.

The order of authorship is arbitrary; all three of us contributed equally to this article. None of us is or has ever been associated with any integrity test publishers.

Partial and earlier results from this research were presented at the following conferences: annual conference of the Society for Industrial and Organizational Psychology, St. Louis, Missouri, 1991; annual conference of the American Psychological Society, Washington, DC, 1991; annual conference of the Society for Industrial and Organizational Psychology, Montreal, Quebec, Canada, 1992; 26th International Congress of Psychology, Brussels, Belgium, 1992. This article supersedes all previous versions, including the working manuscript cited in Goldberg, Grenier, Guion, Sechrest, & Wing (1991).

This research was supported in part by the Office of Naval Research (Contract No. N00014-91-G-4168 to Frank L. Schmidt). We thank Howard Timm and Ralph Carney of the Defense Personnel Security Research and Education Center for their encouragement and support of this research. The content of this article does not necessarily reflect the position or the policy of the U.S. government, and no official endorsement should be inferred.

We thank numerous test publishers, authors, and colleagues who sent us data on integrity tests. We also thank John Hunter and Paul Sackett for many informative and constructive discussions, two anonymous reviewers, and many colleagues for their comments on earlier versions of our work.

Correspondence concerning this article should be addressed to Deniz Ones, who is now at the Department of Management, University of Houston, Houston, Texas 77206-6283.

Journal of Applied Psychology, 1993, Vol. 78, No. 4, 679–703
Copyright 1993 by the American Psychological Association, Inc. 0021-9010/93/$3.00

tion of whether they all measure primarily a single general construct. Different test publishers claim that their 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 that these tests may all measure the general construct of broadly defined *conscientiousness,* one of the five dimensions of personality hypothesized in the Big Five theory of personality (Barrick & Mount, 1991; Digman, 1990; Goldberg, 1990). Conscientiousness reflects such characteristics as dependability, carefulness, and responsibility. In the integrity-testing literature, this construct appears to have been viewed and measured from its negative pole (e.g., irresponsibility, carelessness, and violation of rules). Inspection of items on several integrity tests confirms this notion. Therefore, high correlations might be anticipated among the different integrity tests. Detailed descriptions of all integrity tests can be found in the *Tenth Mental Measurements Yearbook* (Conoley & Kramer, 1989) and 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 that contributed data to the analyses reported in the present research.

If all integrity tests measure an overall general construct, then integrity test validities will generalize across different predictor measures. That is, all integrity tests may have at least moderate positive levels of validity, lending them some potential utility in personnel selection. If validity generalization results across all integrity tests showed substantial variability in validities after correction for the effects of statistical artifacts, then we explored potential influences of moderating variables on the validities. The proposed moderators of integrity test validities are enumerated in Table 2.

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[a]
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. PEOPLE Survey[a]
12. Personnel Decisions, Inc., Employment Inventory[b]
13. Personal Outlook Inventory[b]
14. Personnel Reaction Blank[b]
15. Personnel Selection Inventory (London House)[a]
16. Phase II Profile[a]
17. 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 and Alienation Index[a]
25. Wilkerson Preemployment Audit[a,c]

Note. The list of publishers and authors of these tests is available in O'Bannon, Goldinger, & Appleby (1989).
[a] Overt integrity test. [b] Personality-based integrity test. [c] No validity data were reported, but the test contributed to the statistical artifact distributions.

Sackett et al. (1989) classified 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 of these tests, admissions are not a part of the instrument, but are instead used as a criterion measure in validity studies. Overt integrity tests include the PSI (London House Press, 1975), the Employee Attitude Inventory (EAI; London House Press, 1982), the Stanton Survey (Klump, 1964), the Reid Report (Reid Psychological Systems, 1951), the Phase II Profile (Lousig-Nont, 1987), the Milby Profile (Miller & Bradley, 1975), and the Trustworthiness Attitude Survey (Cormack & Strand, 1970). 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, and drug abuse, in addition to theft) using composite measures of 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), the Employment Inventory of Personnel Decisions, Inc. (PDI; Paajanen, 1985), and the Hogan Personality Inventory's Reliability Scale (R. Hogan, 1981). Thus, in our first set of proposed analyses we examined the validities of overt integrity tests and personality-based tests separately (Proposed Analysis 1 in Table 2).

If the classification of the predictors into overt versus personality-based categories is not found to explain sizable portions of variance in the validities, then criteria characteristics can be explored as moderators. Many researchers have pointed to the diversity and the deficiencies of the criteria used in validating

Table 2
Proposed Moderator Analyses for Integrity Test Validities in Predicting Job Performance and Counterproductive Behaviors

Analysis no.	Moderator analysis
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.

integrity tests (McDaniel & Jones, 1986, 1988; Sackett & Harris, 1984). The criteria of interest in integrity testing can be categorized into overall job performance and counterproductive behaviors on the job. In this research, in Study 1 (described later) we investigated criteria of overall job performance, whereas in Study 2 we examined criteria of counterproductive behaviors.

In traditional validation studies, the criterion of job performance has usually been measured using supervisory ratings. Another method of measuring job performance is with organizational production records. There is some evidence that these two methods of measuring worker performance are not 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). In the moderator analysis of job-performance measurement method (supervisory ratings vs. production records) we tested 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 Analysis 2 in Table 2).

For the criterion of counterproductive behaviors on the job, we expected that the measurement method used for criteria would moderate validity (Proposed Analysis 3 in Table 2). From a methodological perspective, measures of counterproductive behavior can 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), which show 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 of theft, and so on. On the other hand, the self-report criteria include all admissions of theft, past illegal activities, and counterproductive behaviors. Because not all thieves are caught or all illegal activities detected, we expected lower correlations with external criteria. On the other hand, if respondents were to provide socially desirable responses, the effect could be to depress the correlations based on self-report criteria in relation 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 light of the results of an earlier meta-analysis (McDaniel & Jones, 1988), we hypothesized that the validity would 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, we analyzed narrow criteria (i.e., theft) separately from broad criteria (i.e., general disruptive or rule-breaking behaviors). 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 the 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. Sackett et al. (1989) hypothesized that the validity of overt integrity tests in predicting theft (narrow criteria) would 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" (p. 494). That is, they hypothesized that narrowly defined criteria, such as theft, might be better predicted by narrowly focused predictors. In contrast, Sackett et al. also hypothesized that 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 better predict a broader range of behaviors.

There are three other potential moderators that merit investigation. The first is the question of whether concurrent validities accurately estimate predictive validities (Proposed Analysis 5 in Table 2). In selection research, although concurrent validities may shed light on the question of construct validity, the major use of concurrent validity is to estimate predictive validity. 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 Analysis 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 the applicants. The purpose of criterion-related validity studies in employment is to estimate the validity of the selection instrument when used to select applicants. By examining the validities of integrity tests for employee and applicant groups separately, we hoped to determine 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 Analysis 7 in Table 2). The moderating influences of job complexity on general-mental-ability-test validities in predicting job performance are well established (Hunter & Hunter, 1984). For general mental ability tests, as the level of job complexity increases, the validity of the tests also increases. However, the opposite effect may hold for integrity test validities. It could be hypothesized that as the level of job complexity increases, estimated validity of integrity tests would systematically decline because of more successful dissimulation by incumbents and applicants for high-complexity jobs, because of greater difficulty in detecting dishonest behaviors in these jobs, or for both reasons.

The proposed moderating effects enumerated in Table 2 could covary. Potential confounding of moderator variable effects could exist if, for example, most self-report criteria were also narrow criteria. The identification of potentially confounded moderator effects involves the simultaneous examination of the proposed moderators. Availability of validities in each category may preclude an analysis of all combinations. However, to the extent feasible, our intention was to conduct a fully hierarchical moderator analysis (Hunter & Schmidt, 1990b, p. 527).

Method

Description of the Database

We conducted a thorough search to locate all existing integrity test validities. We obtained all published empirical studies 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, No date; 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., there are 43 integrity tests in use in the United States. All of the publishers and authors of the 43 tests were contacted by telephone or in writing to request validity, reliability, and range-restriction information on their tests. Of these, 36 responded by sending research reports. In addition, we identified other integrity tests overlooked by O'Bannon et al.; the publishers of these tests were also contacted. All unpublished and published technical reports reporting validities, reliabilities, or range-restriction information were obtained from integrity test publishers and authors. Some integrity test authors and publishers responded to our request for validity information by sending us data and correlational analysis results in computer printouts that had not yet 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 in which 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, 1990a). We used these corrected correlations in the meta-analysis. We adjusted sample sizes for these corrected correlations to avoid underestimating the sampling-error variance. First, we used the uncorrected correlation and the study sample size to estimate the sampling-error variance for the observed correlation. We corrected this value for the effects of the dichotomization correction and then used this corrected sampling-error variance with the uncorrected correlation in the standard sampling-error formula to solve for the adjusted sample size, which we 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 composed the database. The total sample size (number of test takers) across 665 validities was 576,460. 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 is provided in the Appendix. The studies listed in the Appendix were often incomplete in reporting various aspects of the validities, and additional information had to be obtained through personal communications with test publishers. Of the 665 validity estimates, 247 validity coefficients came from the published literature or from the published reviews of integrity tests; of these, 67 were published. To address the concern that there could be some systematic difference in validities between the published sources and the unpublished sources, we computed the correlation between the validity coefficients reported and the dichotomous variable of validities reported in published versus unpublished studies. This correlation was −.02. The negative correlation indicates that published studies reported higher validities, but the low absolute value of the correlation indicates that any differences in validities are negligible. Hence, in our database, the published versus unpublished distinction for the validities was 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 criterion-related 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. For the 9 cases in which 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 576,460 data points included 389 validities from overt integrity tests and 276 validities from personality-based integrity tests. Most of the validities were computed on samples from service industries ($K = 503$), most notably from the retail industry (i.e., discount chains, department stores, supermarkets, grocery chains, convenience stores, and drug stores). The increasing service orientation of the U.S. economy (Johnston & Packer, 1987) makes the results of this meta-analysis more relevant. The validities were reported on a diverse range of occupations across high-, medium-, and low-complexity levels. Finally, of the 665 validities, 222 had job performance as the criterion, and 443 had counterproductive behaviors as the criterion.

Artifact Distributions

We compiled several sets of artifact distributions: three distributions for the reliability of the integrity tests, four distributions for the reliability of the criterion variables, and one distribution for range restriction. Descriptive information on the artifact distributions is provided in Table 3.

A total of 124 integrity test reliability values were obtained from the published literature and the test publishers. Of the 124 values, 68 were alpha coefficients (55%), and 47 were test–retest reliabilities over periods of time ranging from 1 to 1,825 days ($M = 111.4$ days, $SD = 379.7$ days). The mean of the coefficient alphas was .81 ($SD = .10$), and the mean of the test–retest reliabilities was .85 ($SD = .10$). There were 9 reliabilities reported for which the type of reliability was not given. The ideal estimate of reliability for purposes of this meta-analysis is coefficient alpha or the equivalent. However, test–retest reliability estimates over relatively short time periods provide reasonably close approximations to alpha coefficients. Furthermore, in this case the means of the two reliability types were similar: The overall mean of the predictor reliability artifact distribution was .81 ($SD = .11$) and the mean of the square roots of predictor reliabilities was .90 ($SD = .06$). We constructed two other predictor reliability distributions, 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, ($SD = .09$). The mean of the square roots of overt test reliabilities was .91 ($SD = .05$). There were 27 reliabilities reported for personality-based tests. The mean of the personality-based-test reliability artifact distribution was .72 ($SD = .13$). The mean of the square roots of the reliabilities was .85 ($SD = .08$). Each of these predictor reliability distributions was used in analyses with corresponding predictor categories. That is, when validities of overt tests were being cumulated, the predictor reliability distri-

Table 3
Descriptive Information on Statistical Artifact Distributions Used to Correct Validities

Artifact distribution	No. of values	*M*	*SD*	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
Range-restriction correction values					
U[c]	79	.81	.19	—	—

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

bution for overt tests was used; but when validities of personality-based tests were being meta-analyzed, the predictor reliability distribution for personality-based tests was used. Finally, when the analyses involved both overt and personality-based tests, the overall predictor reliability distribution was used.

Reliability estimates for the criterion variables were taken from the studies that contributed to the database for this meta-analysis and from the published literature on counterproductivity and job performance. We created four separate distributions, 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 were 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); and .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. We assigned the .52 reliability for supervisory ratings of overall job performance a frequency of 153 to match the number of validities for that criterion in our database, and we combined it with 10 reliabilities for production records to compose the distribution of job performance reliabilities. We obtained the reliability of production records from Hunter, Schmidt, and Judiesch (1990) as .55 for a 1-week period. Using the Spearman-Brown formula, we adjusted this value 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. We assigned each of the reliabilities a frequency corresponding to the number of validities in the database by 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. We constructed a distribution of range-restriction values from the studies contributing to the database. There were 75 studies that reported both the study-sample standard deviation and the applicant-group standard deviation. We calculated the range-restriction ratio as the ratio of study-group(s) to reference-group (S) standard deviations (s:S). In four studies, correlations were reported for both the applicant and the employee groups. From these four studies we calculated range-restriction ratios 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 we used was .81 (*SD* = .19); this indicates that there is considerably less range restriction in this domain than is the case for cognitive ability (Alexander, Carson, Alliger, & Cronshaw, 1989). Thus, range-restriction corrections were much smaller in the present research than in meta-analyses in the abilities domain.

Meta-Analytic Procedures

We tested the hypotheses in this article using the Hunter and Schmidt (1990b, 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. We used the artifact distributions described previously to correct biases in the observed validities that were caused by statistical artifacts. The artifacts operating across studies included 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, or the standard deviation of the true validities will be relatively large, or both. 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, we used the interactive meta-analysis procedure (Hunter & Schmidt, 1990b, p. 165; Schmidt, Gast-Rosenberg, & Hunter, 1980). The program we used incorporated refinements shown by computer simulation studies to increase accuracy (Hunter & Schmidt, in press; Law, Schmidt, & Hunter, 1992a, 1992b). These refinements included 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, 1990b). 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. Final possibilities for variability are truly situationally specific test validities, the operation of moderator variables, or both. In summary, we used the 90% credibility value to judge whether the validities are positive across situations (i.e., whether validity generalizes), whereas we used the variance accounted for by statistical artifacts and the estimated standard deviation of true validities 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. When more than one correlation was available on a single sample for the same criterion, we averaged the validities to avoid violations of the independence assumption (Hunter & Schmidt, 1990b, pp. 452–454). The sample size we used was the average sample size.

In our meta-analyses we corrected the mean observed validity for mean attenuation due to criterion unreliability and range restriction (Hunter & Schmidt, 1990b, p. 165). We did not apply a correction for predictor unreliability to the mean validity because our interest was in estimating the operational validities of integrity tests for selection purposes. However, we did correct the observed variance of validities for variation in predictor unreliabilities in addition to correcting for variation in criterion unreliabilities and range-restriction values. For comparison purposes, we provide the percentage of variance due to sampling error alone in our results. Furthermore, we present the mean observed validities without any artifact corrections.[1]

Analyses and Results

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

The first meta-analysis estimated the validity of all integrity tests combined for predicting the criterion of overall job performance. The total sample size across 222 studies reporting such a correlation was 68,772. This meta-analysis indicates that the proportion of the variance observed in validities because of statistical artifacts was 53%. The estimate of the mean true validity (ρ)[2] of all integrity tests with the criterion of overall job performance is .34 (SD_ρ = .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 (K) between integrity test scores and counterproductive behaviors. The 443 correlations were over a total sample size of 507,688, 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, and the standard deviation is .37, a fairly large value. In addition, sampling error, unreliability in the predictor, unreliability in the criteria, and range restriction combined accounted for only 9% of the variance observed in the correlations. These results indicated that all types of integrity tests were valid predictors of counterproductive behaviors. But the standard deviation of the true validity in this analysis was large enough, and the percentage of variance accounted for was low enough, to suggest that other statistical artifacts or potential moderators were operating. Finally, 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 was .34. However, the standard deviation of the true validity (.13) and the percentage of variance accounted for (53%) by all statistical artifacts we could correct for (i.e., sampling error, criterion and predictor unreliability, range restriction, and dichotomization) indicated that the validity could have been moderated by other variables. The results of the moderator analyses are reported in Table 5.

The first potential moderator tested was the predictor type (overt vs. personality based). The results across 84 validities and 31,089 data points showed that the best estimate of overt integrity tests' validity in predicting overall job performance was .33. The lower 90% credibility value of .16 indicated that the validity was positive across studies and situations. The variance accounted for by the corrected statistical artifacts was 40%, and the standard deviation of the true validity was .15. Personality-based integrity tests showed a mean validity of .35 (K = 138, N = 37,683) in predicting overall job performance, with 63% of the observed variance accounted for by the statistical artifacts we could correct for. The standard deviation of the true validity for personality-based integrity tests was .11, and the lower 90% credibility value was .23, indicating that the validities of personality-based integrity tests were 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

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

[2] The phrase *true validity* is a shorthand expression for the longer phrase *estimated true (operational) criterion-related validity for this particular criterion.*

Table 4
Overall Meta-Analyses of the Validity of Integrity Tests

Category of analysis	N	K	mean r	SD_r	SD_{res}	ρ	SD_ρ	% variance SE	% variance acc. for	90% CV
All integrity tests predicting overall job performance[a]	68,772	222	.21	.1019	.0701	.34	.13	30.9	52.6	.20
All integrity tests predicting counterproductive behaviors[b]	507,688	443	.33	.2463	.2345	.47	.37	1.5	9.4	.05

Note. K = number of correlations; mean r = mean observed correlation; SD_r = observed standard deviation; SD_{res} = residual standard deviation; ρ = true validity; SD_ρ = standard deviation of the true validity; % variance SE = percentage of variance due to sampling error; % variance acc. for = percentage of variance due to all corrected statistical artifacts; 90% CV = lower 90% credibility value.
[a] The criteria for validation included supervisory ratings of overall job performance, production records, and commendations. [b] The criteria for validation included narrow and broad criteria of disruptive behaviors, such as actual theft, admitted theft, dismissals for actual theft, illegal activities, absenteeism, tardiness, and violence.

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 test validities, suggested by Nathan and Alexander (1988), is the criterion measurement method (supervisory ratings vs. production records). We meta-analyzed all available correlations between integrity tests and supervisory ratings of overall job performance. There were 153 such correlations obtained from a total sample size of 40,013 data points. The operational validity of integrity tests in predicting supervisory ratings of job performance was .35. The lower 90% credibility value was .20, indicating that the validity was positive across studies and situations. The variance accounted for by the corrected statistical artifacts was 55%, and the standard deviation of the true validity was .13. For production records criteria, there were only 10 validities based on a total sample size of 2,210. The true validity for predicting production records was .28 (SD_ρ = .12). The lower credibility value and the percentage of variance accounted for by statistical artifacts were .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 were somewhat similar (ρ = .35 and .28, respectively). Therefore, we concluded that the criterion measurement method probably does not have a large impact on integrity test validities in predicting job performance. This result mirrors the findings of Nathan and Alexander that studies using the criterion of supervisory ratings of

Table 5
Meta-Analyses of the Validity of Integrity Tests for Predicting Overall Job Performance: All Performance Criteria

Category of analysis	N	K	mean r	SD_r	SD_{res}	ρ	SD_ρ	% variance SE	% variance acc. for	90% CV
Predictor type[a]										
Overt	31,089	84	.20	.1093	.0844	.33	.15	23.3	40.5	.16
Personality based	37,683	138	.22	.0976	.0591	.35	.11	37.0	63.3	.23
Criterion measure										
Supervisory ratings	40,013	153	.21	.1039	.0699	.35	.13	35.9	54.7	.20
Production records	2,210	10	.22	.1163	.0846	.28	.12	30.4	47.1	.15
Validation strategy[a]										
Concurrent	31,877	135	.22	.1051	.0683	.37	.12	34.8	57.7	.22
Predictive	35,411	79	.19	.0951	.0687	.31	.12	26.9	47.9	.17
Validation sample										
Applicants[a,b]	26,215	43	.24	.0617	.0000	.40	.00	41.3	100.0	.40
Employees[a]	27,675	135	.17	.1274	.0970	.29	.18	32.3	42.0	.08
Job complexity[a]										
Low	1,633	19	.28	.0902	.0000	.45	.00	100.0	100.0	.45
Medium	16,200	80	.19	.1180	.0831	.32	.15	36.4	50.3	.14
High	858	11	.28	.1215	.0000	.46	.00	85.0	100.0	.46

Note. K = number of correlations; mean r = mean observed correlation; SD_r = observed standard deviation; SD_{res} = residual standard deviation; ρ = true validity; SD_ρ = standard deviation of the true validity; % variance SE = percentage of variance due to sampling error; % variance acc. for = percentage of variance due to all corrected statistical artifacts; 90% CV = lower 90% credibility value.
[a] Criteria for validation included supervisory ratings of overall job performance, production records, and commendations. [b] These studies were predictive, with the exception of one study (N = 27).

job performance produce validity estimates similar to those from studies using production quantity as the criterion.

The third potential moderator studied was the validation strategy used in the primary studies. To determine whether concurrent validities estimate predictive validities accurately in this noncognitive domain, we separately meta-analyzed predictive and concurrent validities for predicting overall job performance. Predictive validities of integrity tests had a mean true validity of .31 in predicting job performance whereas, concurrent studies had a mean true validity of .37. These results seemed 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 31,877, 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 (the mean true validity using the sample-size weighted-mean validity). In the analysis of predictive validities, there was also a validation study with a very large sample. 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 35,411, 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 were compared for both the concurrent and predictive validation strategies, it seemed that predictive validity ($\rho = .43$) was almost twice as large as concurrent validity ($\rho = .23$). This contradicted the conclusions reached using mean true validities based on sample-size weighted means. Because it could not be determined in which set of analyses, if either, the large-sample studies were biasing the results, the conclusion about the moderating influences of validation strategy on validities when the criterion is job performance was inconclusive. Other analyses, reported in Study 2, examined whether concurrent and predictive validities were 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 indicated that validity of integrity tests for predicting job performance was positive (90% CV = .22 and .17, respectively). Furthermore, the substantial concurrent validity provided evidence of construct validity for integrity tests.

The fourth potential moderator studied was the validation sample used in the studies (applicant sample vs. employee sample). This analysis was not redundant with the analysis of predictive versus 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 1 predictive study conducted on applicants using the criterion of supervisory ratings of performance on a work sample. 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 applied personnel-selection research that value is important only to the extent that it approximates the validity of the applicant sample. For studies using the criterion of overall job performance, the mean true validity estimate obtained using an applicant sample was .40; when employees constituted the sample, the mean true validity estimate was .29 ($SD_\rho = .00$ and .18, respectively). Hence, in studies in which applicants constituted the sample, 100% of the variance was explained by statistical artifacts. On the other hand, in validity studies in which employees constituted the sample, 42% of the variance was explained by the statistical artifacts and the lower credibility value was .08, indicating that the validity was positive across studies and situations. But the large standard deviation of true validity and the low percentage of variance accounted for in employee samples suggest 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 was job complexity. We used three job-complexity levels: high, medium, and low (as defined by Hunter et al., 1990). Several studies reported too little information for us 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 *Dictionary of Occupational Titles* (1991) code for the job on which the validation was undertaken. For the other 112 studies providing validity coefficients with job performance, either no data were available on the jobs constituting the sample or the studies indicated a heterogeneous sample comprising several jobs (e.g., retail employees). Among the 110 studies that supplied information on the jobs studied, the percentages conducted at each of the three complexity levels were as follows: low, 17%; medium, 73%; and high, 10%. In the U.S. economy as a whole, these percentages are 20%, 63%, and 17%, respectively (Hunter, 1980); therefore, the distribution here was in rough proportion to that in the economy, although medium-complexity jobs were somewhat overrepresented and low-complexity jobs were somewhat underrepresented. The meta-analysis results for this moderator indicated that for low-complexity jobs the mean true validity across 1,633 people was .45 ($SD_\rho = .00$; see Table 5). For low-complexity jobs, the artifacts that we corrected for explained all the observed variation in integrity test validities in predicting job performance. For medium-complexity jobs, the mean true validity across 16,200 people was .32 ($SD_\rho = .15$), and statistical artifacts accounted for 50% of the variance. For high-complexity jobs the mean true validity across 858 people and 11 validities was .46 ($SD_\rho =$

.00). Given the (relatively) small sample size ($N = 858$) and the small number of correlations in the high-complexity category, we expect that these results may not be robust. However, from these results an interesting pattern emerges, suggesting that even for high-complexity jobs, integrity tests are valid for 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 on the basis of 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, we separately analyzed integrity test validities that used the criterion of supervisory ratings of job performance to determine moderating influences. These results are reported in Table 6.

For the most part, results were similar to those reported for job performance in Table 5. Test type did not seem to be a strong moderator of the integrity test validities. Overt integrity tests predicted supervisory ratings with a true validity of .30, and personality-based integrity tests predicted supervisory ratings with a true validity of .37. However, further moderator analyses are necessary before any definite conclusions can be reached regarding these two types of tests.

The mean true validity estimate across studies that used a concurrent validation strategy was .39 ($SD_\rho = .11$). The true validity across studies that used a predictive validation strategy was .32 ($SD_\rho = .13$). These results suggested that when the criterion of interest was supervisory ratings of overall job performance, concurrent validities could overestimate predictive validities somewhat in the domain of integrity testing. However, as we noted for the similar moderator analysis for all measures of job performance, among predictive studies included here there was a study with a very large sample ($N = 6{,}884$) reporting an observed validity of .15. For the predictive validities, the total sample size was 26,409, 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 $r = .28$). When we applied the statistical artifact corrections to this unweighted mean validity, we obtained a true validity of .46 for the predictive validation category. A similar reanalysis 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 was inconclusive. Other analyses, reported in Study 2, examined whether concurrent and predictive validities were 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 general conclusions were reached for the criterion of supervisory ratings of overall job performance as were reached earlier for the combined criteria of job performance: Studies conducted on applicant samples seemed to yield higher estimated operational validities than those conducted on employee samples (ρ = .42 and .33, respectively). Integrity tests also seemed to be at least as valid for high-complexity jobs as for low-complexity jobs (ρ = .51 and .46, respectively). The percentages of studies at each complexity level were as follows: low, 23%; medium, 63%; and high, 14%. These percentages were very

Table 6
Meta-Analyses of the Validity of Integrity Tests for Predicting Overall Job Performance: Supervisory Ratings Only

Category of analysis	N	K	mean r	SD_r	SD_{res}	ρ	SD_ρ	% variance SE	% variance acc. for	90% CV
Predictor type										
Overt	12,932	51	.18	.1430	.1189	.30	.22	23.4	30.8	.05
Personality based	27,081	102	.22	.0811	.0259	.37	.05	53.6	89.8	.32
Validation strategy										
Concurrent	12,120	88	.23	.1109	.0577	.39	.11	53.2	73.0	.27
Predictive	26,409	57	.19	.0968	.0727	.32	.13	24.9	43.6	.17
Validation sample										
Applicants[a]	7,831	25	.25	.0814	.0252	.42	.05	47.7	90.4	.37
Employees	18,499	90	.20	.1278	.0958	.33	.18	32.7	43.8	.13
Job complexity										
Low	1,333	16	.28	.0850	.0000	.46	.00	100.0	100.0	.46
Medium	7,438	45	.22	.1272	.0905	.36	.17	36.2	49.4	.17
High	723	10	.31	.1185	.0000	.51	.00	95.7	100.0	.51

Note. K = number of correlations; mean r = mean observed correlation; SD_r = observed standard deviation; SD_{res} = residual standard deviation; ρ = true validity; SD_ρ = standard deviation of the true validity; % variance SE = percentage of variance due to sampling error; % variance acc. for = percentage of variance due to all corrected statistical artifacts; 90% CV = lower 90% credibility value.
[a] These studies were predictive, with the exception of one study (N = 27).

similar to those for jobs in the economy as a whole (Hunter, 1980; Hunter & Hunter, 1984).

The moderator analyses reported for job performance and supervisory ratings of job performance could have given a distorted picture if the moderator variables were not independent. To determine the relationships between the moderators, we calculated intercorrelations of the moderator variables. The results are reported in Table 7.

Job complexity was not highly correlated with the other potential moderators (mean $r = -.06$). Type of test (overt vs. personality based) did not seem to be highly correlated with the other potential moderators (mean $r = -.11$). However, validation strategy was substantially correlated with the sample used, that is, applicants versus employees ($r = -.58$). Predictive studies more frequently used applicant samples, and concurrent studies more frequently used employee samples, as concurrent criterion data were not typically available on applicant samples. This finding was 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 seemed to be highly correlated, a hierarchical moderator analysis was needed to assess the potential impact of confounding on the moderator analyses. To accomplish this, we first broke down all integrity test validities for supervisory ratings of overall job performance by validation strategy, and then, within the concurrent and predictive validation categories, we undertook a moderator analysis by validation sample (applicants vs. employees). These results are reported in Table 8.

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 test results are used to predict job performance at some future point in time. Table 8 shows that when integrity tests were administered to applicants and the scores were used to predict later supervisory ratings of job performance the mean operational validity was .41. This result was based on 7,550 individuals and 23 validity coefficients. The standard deviation of the true validity was .00, indicating that all of the variance across studies and situations observed in this analysis was due to statistical artifacts and that the true validity of .41 was invariant across settings. When employees made up the sample of predictive studies (column 2 in Table 8), the operational validity was much lower ($\rho = .26$ across a total sample size of 8,994 and 20 validity coefficients). In addition, the standard deviation of the true validity was .21, with only 24% of the variance accounted for. Concurrent validation conducted on employees (Table 8, lower half of column 2) produced an operational validity of .37 across 8,275 individuals and 63 validity coefficients. The standard deviation of the true validity was .14, and 61% of the observed variance was 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, because of the extremely small sample size of that study ($N = 27$), we gave little weight to this validity coefficient. The results in Table 8 seemed to indicate that concurrent validities overestimated predictive validities when the comparison was limited to the employee group. For employees, the estimated mean true concurrent validity was .37, whereas the estimated mean true predictive validity was .26. However, concurrent studies calculated on employees did not overestimate predictive validities calculated on applicants (.37 vs. .41). This finding is important because concurrent studies based on em-

Table 7
Intercorrelations Between Moderators of Integrity Tests in Predicting Overall Job Performance

Moderator	1	2	3	4
1. Predictor type	—	.15	−.40	−.09
N		214	179	109
2. Validation strategy		—	−.58	−.27
N			171	106
3. Validation sample			—	.18
N				105
4. Job complexity				—

Note. N = number of studies used in calculating the correlations (i.e., sample size). All of the moderators were dummy coded as follows: for predictor type, overt = 1 and personality based = 2; for validation strategy, concurrent = 1 and predictive = 2; for validation sample, applicants = 1 and employees = 2; for job complexity, high = 1, 2, medium = 3, and low = 4, 5.

Table 8
Hierarchical Moderator Analyses of the Integrity Test Validities for Predicting Supervisory Ratings of Overall Job Performance

Statistic	Applicants	Employees
	Predictive	
N	7,550	8,994
K	23	20
mean r	.25	.15
SD_r	.0753	.1318
SD_{res}	.0000	.1146
ρ	.41	.26
SD_ρ	.00	.21
% variance acc. for	100	24.4
90% CV	.41	.01
	Concurrent	
N	27	8,275
K	1	63
mean r	.29	.22
SD_r	—	.1227
SD_{res}	—	.0766
ρ	.48	.37
SD_ρ	—	.14
% variance acc. for	—	61.0
90% CV	—	.21

Note. K = number of correlations; mean r = mean observed correlation; SD_r = observed standard deviation; SD_{res} = residual standard deviation; ρ = true validity; SD_ρ = standard deviation of the true validity; % variance acc. for = percentage of variance due to all corrected statistical artifacts; 90% CV = lower 90% credibility value.

ployees are frequently used to estimate the predictive validity these tests will have when they are used with applicants. When the validation strategy was controlled for, validities from applicant samples were higher than validities from employee samples. For predictive validities, the applicant group mean true validity was .41 and the employee group mean true validity was .26. Although both validation strategy and validation sample seemed to affect estimates of integrity test validities for predicting supervisory ratings of overall job performance, the highest mean operational validity estimate was obtained in applicant samples using predictive validation strategies (ρ = .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 percentage of variance accounted for by the statistical artifacts (9%) indicated 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.

The first potential moderator tested was the predictor type (overt vs. personality based). We used all available correlations between overt integrity tests and disruptive behaviors on the job. The results across 305 correlations and 349,623 data points showed that the best estimate of the mean validity of overt tests in predicting disruptive behaviors was .55. The lower 90% credibility value of .07 indicated that the validity was positive across studies and situations. However, the percentage of variance accounted for by corrected statistical artifacts was low at 9%, and the standard deviation of the true validity was large at .41. The meta-analysis of personality-based integrity test validities showed 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. The standard deviation of the true validity for personality-based integrity tests was .11, much smaller than the value of .41 for overt tests. The lower credibility value of .20 indicated that validities of personality-based integrity tests were positive across studies and situations. These results appeared to suggest that overt integrity tests may be better at predicting counterproductivity (ρ = .55) than are personality-based tests (ρ = .32); however, this conclusion was premature without an examination of other potential moderator variables.

The second moderator analysis involved 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. The results supported the

Table 9
Moderator Analyses for Predicting Counterproductive Behaviors

Category of analysis	*N*	*K*	mean *r*	SD_r	SD_{res}	ρ	SD_ρ	% variance SE	% variance acc. for	90% CV
Predictor type[a]										
Overt	349,623	305	.39	.2835	.2710	.55	.41	1.1	8.6	.07
Personality based	158,065	138	.22	.0884	.0663	.32	.11	11.3	43.7	.20
Criterion measure[a]										
Admissions of counterproductivity	309,831	255	.41	.2730	.2589	.58	.40	1.1	10.1	.11
Externally measured counterproductivity	197,717	187	.22	.1490	.1369	.32	.22	4.5	15.6	.07
Criterion breadth										
Theft[b]	193,631	152	.36	.2654	.2523	.52	.39	1.6	9.6	.06
Broad counterproductivity[c]	312,827	290	.32	.2382	.2267	.45	.36	1.5	9.4	.04
Validation strategy[a]										
Concurrent	219,640	295	.39	.2680	.2539	.56	.39	1.4	10.2	.10
Predictive	282,544	138	.25	.1885	.1785	.36	.28	2.1	10.4	.03
Validation sample[a]										
Applicants	369,581	183	.30	.2314	.2207	.44	.35	1.1	9.0	.04
Employees	105,369	153	.38	.3120	.3003	.54	.47	1.2	7.4	.02
Job complexity[a]										
Low	14,301	44	.30	.1836	.1607	.43	.25	11.3	23.4	.13
Medium	32,764	78	.28	.1731	.1513	.40	.24	11.2	23.6	.13
High	2,372	21	.49	.1751	.1295	.68	.20	17.9	45.3	.45

Note. *K* = number of correlations; mean *r* = mean observed correlation; SD_r = observed standard deviation; SD_{res} = residual standard deviation; ρ = true validity; SD_ρ = standard deviation of the true validity; % variance SE = percentage of variance due to sampling error; % variance acc. for = percentage of variance due to all corrected statistical artifacts; 90% CV = lower 90% credibility value.
[a] Criteria included narrow and broad criteria of disruptive behavior, such as actual theft, admitted theft, dismissals for actual theft, illegal activities, absenteeism, tardiness, and violence. [b] Included narrow criteria of admissions of theft, actual theft, and dismissals for actual theft. [c] Broad criteria included violence on the job, tardiness, absenteeism, and other disruptive behaviors not included in the narrow criteria.

McDaniel and Jones's findings and indicated that admissions criteria yielded a mean true validity estimate of .58, whereas for predicting external criteria the mean true validity estimate was .32 (SD_ρ = .40 and .22, respectively). Only 10% of the variance was accounted for by artifacts with admissions criteria, and 16% was accounted for with external criteria. The fairly large standard deviations of the true validities and relatively small percentages of variance accounted for indicated that validities of integrity tests may be affected by other moderators. However, the positive 90% credibility values indicated that the integrity test validities could be expected to be positive across situations for both the criteria of admissions of counterproductivity and of externally measured counterproductivity.

We next examined criterion breadth as a potential moderator of validity for counterproductive-behaviors criteria. As shown in Table 9, integrity test validities against theft criteria yielded an estimated mean operational validity of .52 and a 90% credibility value of .06, with 10% of the variance accounted for (SD_ρ = .39). Validities against broad criteria (general disruptive behaviors) had an estimated mean corrected validity of .45, with a 90% credibility value of .04 and 9% of the variance accounted for by the statistical artifacts. In this case, the standard deviation of the true validity was .36, again a fairly large value. The difference in operational validities for theft criteria (ρ = .52) versus other disruptive behaviors (ρ = .45) indicated that criterion breadth may have been a moderator of integrity test validities.

The fourth potential moderator we studied for the criterion of counterproductivity was the validation strategy used in the studies. To determine whether concurrent validities estimated predictive validities accurately in this noncognitive domain, we separately analyzed predictive and concurrent studies. Predictive validities had a mean of .36, whereas concurrent studies had a mean of .56. These results suggested that concurrent validities might overestimate predictive validities in this research domain. The utility of a selection test depends on its predictive validity; although concurrent validities can shed light on construct validity, the major purpose of concurrent validity in selection research is to estimate predictive validity. Thus, the present finding is potentially important. The percentage of variance accounted for with both concurrent and predictive validities was 10%. The standard deviation of the true validity 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 indicated that validity was likely to be greater than zero, regardless of the validation strategy used.

The next potential moderator we tested was the validation sample (applicant vs. employee). This analysis was not redundant with the analysis of predictive versus 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 was .44 in applicant samples and .54 in employee samples. Thus, employee samples appeared to yield larger validity estimates, a finding consistent with the results of the analysis of predictive versus concurrent studies. The standard deviations of the true validities for these two categories were .35 and .47, respectively. For both types of samples, the lower 90% credibility interval was positive, indicating that the validities were positive across all situations and settings.

A sixth potential moderator of integrity test validities in predicting counterproductive job behaviors was job complexity. As in the job-complexity analysis in Study 1, we used three job-complexity levels: high, medium, and low (as defined by Hunter et al., 1990). Three hundred studies reported too little information for us 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 that supplied information on the jobs studied, most were conducted on medium-complexity jobs. Of the 143 correlations indicating specific jobs used in validation, the percentages of studies at each complexity level were as follows: low, 31%; medium, 54%; and high, 15%. This distribution was reasonably similar to that for jobs in the U.S. economy as a whole (20%, 63%, and 17%, respectively; Hunter, 1980; Hunter & Hunter, 1984). The results indicated that for low-complexity jobs, the mean true validity of integrity tests across 14,301 people was .43, the standard deviation of the true validity was .25, and the artifacts that we corrected for explained 23% of the observed variation in integrity test validities. For medium-complexity jobs, the estimated mean true validity across 32,764 people was .40, the standard deviation of the true validity was .24, and statistical artifacts accounted for 24% of the variance. For high-complexity jobs, the mean true validity across 2,372 people was .68 (SD_ρ = .20). The percentage of variance accounted for by the statistical artifacts was 45%. Because our classification of the validities into the three categories 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 did emerge: There seemed to be some evidence that the mean validity of integrity tests for predicting counterproductive behaviors was highest for high-complexity jobs. This was an unexpected result, and a difference of this magnitude for the prediction of job performance was not observed (see Tables 5 and 6). One possible explanation for this finding 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, and further research is needed to provide more definitive answers.

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

Results indicated that the moderators of job complexity and validation sample (applicants vs. employees) were not highly correlated with the other moderators. Most other moderators seemed to be substantially correlated with each other. Predictor type (overt vs. personality based) correlated substantially with

Table 10
Intercorrelations Between Moderators of Integrity Tests in Predicting Counterproductive Behaviors

Moderator	1	2	3	4	5	6
1. Predictor type	—	.56	.44	.42	−.19	.16
N		442	443	433	409	143
2. Criterion measurement method		—	.38	.74	.22	.00
N			442	433	408	142
3. Criterion breadth			—	.24	.00	.05
N				433	409	143
4. Validation strategy				—	.10	.04
N					402	142
5. Validation sample					—	−.11
N						138
6. Job complexity						—

Note. *N* = number of studies used in calculating the correlations (i.e., sample size). All of the moderators were dummy coded as follows: for predictor type, overt = 1 and personality based = 2; for validation strategy, concurrent = 1 and predictive = 2; for validation sample, applicants = 1 and employees = 2; for job complexity, high = 1, 2, medium = 3, and low = 4, 5.

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 correlated very highly with validation strategy (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 were substantial, we conducted a fully hierarchical moderator analysis for all potential moderators except job complexity.

In a fully hierarchical moderator analysis, the data set of correlations is first broken down by one key potential moderator variable, and then, within each subgroup, subsequent moderator analyses are undertaken one by one in a hierarchical manner (Hunter & Schmidt, 1990b, p. 527). First, we divided the validities for counterproductive behaviors into two categories by predictor type (overt vs. personality based). We then sorted validities within each predictor subgroup into the external criteria or the admissions criteria. Next, we further grouped the validities in each subgroup by theft criteria versus broad criteria, predictive versus concurrent validation, and applicant versus employee sample. The fully hierarchical moderator analysis takes all of the moderators into consideration simultaneously: in this case, 5 moderators with 2 levels each resulted in $2^5 = 32$ combinations. The results of the fully hierarchical analysis are reported in Table 11. Because of a lack of information on some potential moderators in some studies, the breakdown of our database to 32 separate analysis categories, 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 was that many studies did not report whether the predictor data were collected from current employees or from applicants.

Overt tests. The results in the upper half of Table 11 indicate that validities for overt tests were, in general, lower for applicant samples than for employee samples. The respective true estimated validities were as follows: .13 versus .16 for predictive validation using external theft criteria, .32 versus .94 for concurrent validation using externally measured broad counterproductivity criteria, .42 versus .54 for concurrent validation using theft admissions criteria, and .46 versus .99 for concurrent validation using admissions of broad counterproductivity criteria. The exception to this trend was the higher predictive validity obtained for applicant samples (ρ = .39) than for employee samples (ρ = .09) when overt tests were 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 were quite small for overt tests.

The operational selection validity of a test can best be estimated by its predictive validity when it is 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 test is administered to applicants indicates that overt tests have substantial potential utility for selection. The criteria in this category were as follows: composite measures of counterproductive behaviors (two samples), termination for counterproductive behaviors (two samples), disciplinary actions for counterproductivity (one sample), mishandling of cash (three samples), records of tardiness (one sample), and records of absenteeism (one sample). However, when the criterion was the much narrower one of (externally measured) theft alone, the mean estimated validity from predictive studies conducted on applicants was considerably smaller at .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. We did not find any reliability estimates of the latter measures. It is possible that the reliability of external theft measures was lower on average than is the reliability of measures of all counterproductive behaviors. However, if external theft measures had a true average reliability of only .30, this would mean that the mean true validity estimate of .13 in Table 11 would rise to only .20. Thus, the relatively low validities for externally mea-

Table 11
Fully Hierarchical Moderator Analyses of the Validity of Integrity Tests for Predicting Counterproductive Behaviors

	External criteria								Admissions criteria							
	Theft[a]				Broad[b]				Theft[c]				Broad[d]			
	Predictive		Concurrent		Predictive		Concurrent		Predictive		Concurrent		Predictive		Concurrent	
Statistic	App	Ees	App	Ees	App	Ees	App	Ees	App	Ees	App	Ees	App	Ees	App	Ees
								Overt								
N	2,434	9,005			5,598	17,580	277	7,909			68,613	3,217			90,527	27,887
K	7	11			10	23	2	14			63	34			24	46
mean r	.09	.11			.27	.06	.22	.71			.30	.38			.32	.76
SD_r	.1152	.1049			.1218	.1192	.1597	.2336			.2235	.1644			.2336	.1346
SD_{res}	.0781	.0923			.0837	.1091	.1208	.2072			.2128	.1125			.2233	.0771
ρ	.13	.16			.39	.09	.32	.94			.42	.54			.46	.99
SD_ρ	.12	.15			.13	.17	.19	.29			.33	.17			.35	.11
% variance acc. for	54.0	22.6			52.7	16.2	42.7	21.3			9.3	53.2			8.6	67.2
90% CV	−.01	−.01			.23	−.11	.10	.59			.04	.34			.06	.86
								Personality based								
N					93,092	37,415	4,350	1,511								210
K					62	5	6	12								2
mean r					.20	.18	.57	.20								.16
SD_r					.0555	.0118	.0519	.1033								.1000
SD_{res}					.0115	.0000	.0000	.0339								.0000
ρ					.29	.26	.77	.29								.23
SD_ρ					.02	.00	.00	.06								.00
% variance acc. for					95.7	100	100	89.3								100
90% CV					.27	.26	.77	.23								.23

Note. This table represents the following moderators being taken into consideration simultaneously: predictor type, criterion measurement method, breadth of criteria, validation strategy, and validation sample. App = applicants; Ees = employees; K = number of correlations; mean r = mean observed correlation; SD_r = observed standard deviation; SD_{res} = residual standard deviation; ρ = true validity; SD_ρ = standard deviation of the true validity; % variance acc. for = percentage of variance due to all corrected statistical artifacts; 90% CV = lower 90% credibility value.
[a] External measures of actual theft and dismissals for theft. [b] External measures of violence on the job, tardiness, absenteeism, and other disruptive behaviors excluding theft. [c] Admissions of theft and self-reports of dismissals for theft. [d] Admissions of violence on the job, tardiness, absenteeism, and other disruptive behaviors excluding theft.

sured theft were 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 overestimated the predictive validity of overt integrity tests derived from applicant samples. The mean operational validity of .94 was 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 did not overestimate predictive validity, this figure is based on only 2 studies and a total sample size of only 277. For this reason, this validity estimate received little weight in the interpretation of the findings. In addition, as discussed next, concurrent validities conducted on applicants are very atypical validity studies.

The results for overt integrity tests in Table 11 indicated that no matter what the content of the criterion measure (theft or broadly defined disruptive behaviors), self-reported criteria tended to result in higher estimates of validities. 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). On the other hand, when admissions are used as criteria, the difference between reliability and validity becomes potentially tenuous. That is, when tests that include some questions that ask for admissions are validated against admissions, the predictor–criterion correlations indicate in part only that admissions predict other admissions collected at about the same time. (As shown in Table 11, all studies using admissions criteria were concurrent.) In any event, validities against admissions criteria can be taken as evidence of construct validity (Goldberg et al., 1991). The meta-analysis of overt test correlations with admissions criteria indicated that correlations were higher for employees than for applicants. For self-reports of theft, the true estimated mean correlation was .54 for the employee sample (N = 3,217) and .42 for the applicant sample (N = 68,613). In both cases the standard deviations of the true validities were large enough to indicate that additional moderators might have been operating. However, the positive lower credibility values meant that a positive correlation can be expected between honesty test scores and admissions of theft in concurrent studies for both employee and applicant samples regardless of the setting and situation. When the admissions criteria included 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 increased to .99 for employee samples (N = 27,887) and .46 for applicant samples (N = 90,527). In both of these cases, self-report criteria were collected concurrently with the predictor data. The pattern of mean correlations for both theft and broad counterproductive criteria suggested that employees are more willing to admit negative behaviors than are applicants hoping to obtain a job. Under this interpretation, the lower correlations for applicants may be due to response distortion (whether conscious or not) by applicants. Here the focus was on response distortion on the (self-report) criterion measure, but there may also have been response distortion on the predictor by applicants. A much larger portion of the variance in the observed correlations was accounted for by statistical artifacts when the sample comprised employees rather than applicants (67% of the variance in the employee sample and 9% in the applicant sample). In both cases the positive lower credibility value indicated that the concurrent correlations of overt integrity tests with self-reported broad counterproductivity criteria were positive. Taken together, the results for self-report criteria supported the construct validity of overt integrity tests.

Summarizing across both admissions criteria and externally measured criteria, we noted that overt tests predicted broad disruptive behaviors better than they predicted 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 that includes theft, among many other disruptive behaviors on the job. (Future research may show that these disruptive behaviors are not confined to the workplace setting but occur in other areas of life also.) We suspected that this broad construct was general conscientiousness.

Personality-based tests. For personality-based tests, the estimated true validities from applicant samples were equal to or higher than validities obtained using employee samples, when all other moderators were controlled for. The respective mean validities for applicant and employee samples for externally measured broad counterproductivity criteria were .29 versus .26 (predictive) and .77 versus .29 (concurrent). In contrast to overt tests, the standard deviation of the true validity for personality-based tests was .00 or negligibly small (i.e., .02 or .06). For personality-based tests virtually all of the variance in the observed validities was accounted for by statistical artifacts. The mean true validities obtained for personality-based tests did not appear to vary across organizations or situations. One odd category of analysis for personality-based integrity tests was concurrent studies done on applicants with external criteria (K = 6, N = 4,350). These studies used reference checks from previous employers, police reports obtained, interviewer evaluations, and, in one case, disruptive behaviors observed during a 1-day assessment center. This constellation of broad disruptive-behaviors criteria was different in important respects from the other broad counterproductivity criterion measures, which were limited to behavior on the job. This difference might have been responsible for the extraordinarily large true validity obtained for this category (.77). An interesting question is whether measures of this sort, being broader and extending into more areas of life, are more valid measures of both a general tendency toward disruptive behavior and the tendency toward disruptive behaviors on the job. In any event, for present purposes, these six studies (N = 4,350) 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 93,092 applicants using broad measures of counterproductive job behaviors that were assessed externally. The criteria included in this category were as follows: composites of general counterproductivity (4 samples), terminations for counterproductive behaviors (32 samples), disciplinary actions for counterproductivity (5 samples), records of tardiness (7 samples), records of absenteeism (11 samples), citations for negligence (2 samples), and causing accidents (1 sample). 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, we estimated the operational validity for overt tests to be .39. Given that there are some criterion differences between these two categories of tests (see our earlier discussion of overt tests), the difference between the mean validity estimates of .39 and .29 should not be overemphasized. It is possible that the addition of future studies to these analyses would result in closer validity estimates. However, one interpretation that did appear justified was that overt tests, despite their exclusive focus on stealing and theft, are at least as valid as personality-based integrity tests for the prediction of externally measured composites of a wide variety of disruptive behaviors on the job. This was an interesting finding not only from a practical point of view but also from a construct point of view. It suggested that attitudes toward stealing and theft are particularly excellent indicators of general disruptive tendencies and perhaps of general conscientiousness. Measurement of attitudes in this one area appears to be as effective as measurement in a variety of areas (as done by personality-based tests). This was an unexpected finding.

Summary of Major Findings

Job Performance

In selection settings, the best estimate of integrity test validities for predicting job performance would be based on (a) predictive studies that were (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 7,550 people, the best estimate of the mean true validity was .41 (SD_ρ = .00), and 100% of the variance was accounted for. These findings imply that the average validity that integrity tests may be expected to have in selection settings for supervisory ratings of overall job performance is .41 and that this value is constant across settings. The meta-analysis results presented in this research also showed that overt and personality-based tests produce fairly similar operational validities when the criterion of interest is supervisory ratings of overall job performance.

Counterproductive Behaviors

Generally, validities for integrity tests for predicting counterproductive behaviors on the job appear to be fairly substantial. However, we identified several methodological moderators 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 methodological moderators are controlled (see Table 11), the standard deviations of true validity for integrity tests appear to be no larger than those of ability tests in predicting job performance (e.g., Pearlman, Schmidt, & Hunter, 1980; Schmidt, Hunter, Pearlman, & Shane, 1979), thus indicating similar levels of generalizability and a similar lack of support for purely situational moderators. Some exceptions to this conclusion are concurrent studies of overt tests conducted on employees and using externally measured broad counterproductivity criteria (SD_p = .29, Table 11) and concurrent studies of overt tests conducted on applicants using both admissions of theft criteria and broad counterproductive-behaviors criteria (SD_p = .33 and .35, respectively, Table 11). All of these large standard deviation of true validity values are associated with overt integrity tests.

For the criterion of counterproductive behaviors, admissions produced much higher correlations than did external criteria, and concurrent studies often seemed to overestimate predictive validity. The utility of a selection test depends on its predictive validity; although concurrent validities are relevant to questions of construct validity, the major purpose of concurrent validity in selection research is to estimate predictive validity. Thus, it is potentially important that in this research domain, at least for overt tests, concurrent validity estimates overestimate predictive validity.

In selection research, the best estimate of operational selection validities of integrity tests for predicting theft would be based on predictive studies conducted on applicants. In addition, many would argue for reliance on external criteria in preference to admissions criteria. Considering externally measured theft as the criterion in predictive studies, we found the mean operational validity of overt integrity tests to be estimated at .13 (Table 11). For reasons explained earlier, this value may have been underestimated. 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 found that the mean operational validity of both types of integrity tests was positive across situations and was substantial (see Table 11). Theft appears to be less predictable than broad counterproductive behaviors, although we could make this comparison only for overt integrity tests.

In sum, integrity tests predict overall job performance with moderate and generalizable validity. They also predict such counterproductive behaviors as theft, absenteeism, tardiness, and disciplinary problems, but validity estimates seem to be affected by several simultaneously operating methodological 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.

Practical and Theoretical Implications of Findings

Implications for Incremental Validity

A key unanswered question in predicting overall job performance in personnel selection is the size of the increment in validity that results from adding integrity tests to general mental ability tests. Available studies have suggested that the correlations between integrity measures and ability measures are low and negligible. For example, Jones and Terris (1983) found that the correlations between an overt integrity test and a measure of general mental ability were −.02 for the theft admissions subscore and −.03 for the theft attitudes subscore. Gough (1972) reported that a vocabulary test correlated −.05 with the Personnel Reaction Blank. Finally, Werner, Jones, and Steffy (1989) reported that integrity test scores were unrelated to educational level (an arguable proxy for ability), and J. Hogan and Hogan (1989) reported correlations of .07 and −.09 between the Hogan Reliability Scale (a personality-based integrity test) and the quantitative and verbal portions of the Armed Services Vocational Aptitude Battery, respectively. Thus, assuming on the basis of these studies that the correlation between ability and integrity measures is zero, we were able to calculate the expected incremental validity of integrity tests in predicting supervisory ratings of overall job performance. Table 12 Shows the predictions of the incremental validity of integrity tests in predicting supervisory ratings of overall job performance for each of the five job-complexity levels used by Hunter (1980).

In Table 12, the second 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 was .65. This is an increase in validity of 27% in comparison with ability alone and an increase in validity of 59% in comparison with integrity alone. The third column of multiple correlations in Table 12 shows 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 approximately .30 in each of the job-complexity levels. The multiple correlation for predicting overall job performance was .64 for the lowest complexity jobs (Level 5), .67 for medium-complexity jobs (Level 3), and .71 for highest complexity jobs (Level 1). These preliminary results 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 to obtain more precise estimates of the magnitude of the incremental validity of integrity tests.

Implications for Adverse Impact

Hunter and Hunter (1984) indicated that it might be possible to identify other predictors that would add to the validity of

Table 12
Effect of Combining Integrity Tests With Measures of Ability in Predicting Ratings of Overall Job Performance

	Validity			Multiple correlation		
Job-complexity level[a]	GMA[b]	PA[b]	I[c]	GMA + PA[d]	GMA + I	GMA + PA + I
Level 1	.58	.21	.41	.58	.71	.71
Level 2	.56	.30	.41	.55	.69	.72
Level 3	.51	.32	.41	.53	.65	.67
Level 4	.40	.43	.41	.50	.57	.65
Level 5	.23	.48	.41	.49	.47	.64

Note. The multiple correlations reported in this table were computed assuming that both general mental ability and psychomotor ability correlate zero with integrity. GMA = general mental ability; PA = psychomotor ability; I = integrity.
[a] Job-complexity levels are those used by Hunter (1980) and are from highest to lowest. [b] Validities are from Hunter (1980). [c] Predictive validity of integrity tests for supervisory ratings of overall job performance was calculated using applicants (see Table 8). [d] From Hunter (1980).

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) concluded that "minority groups are not adversely affected by either overt integrity tests or personality oriented measures" (p. 499). Integrity test scores and race appear to be uncorrelated. From the ability-testing and personnel selection literatures, it is known that Blacks average about 1 standard deviation below Whites on tests of general mental ability. This difference between Blacks and Whites on general mental ability tests can also be expressed as a correlation between ability and race (r = .45). This correlation is obtained using the standard formula for converting effect sizes (effect size = 1.00 here) to correlations (r = .45). The evidence indicates that ability and integrity scores are uncorrelated and that race and integrity scores are uncorrelated. The correlation between race and an optimally weighted (using regression weights) composite of ability and integrity can therefore be computed. This correlation would be .363, which would convert to an effect size of .78. In other words, the mean difference between Blacks and Whites on an optimally weighted composite of ability test and integrity test scores and race would be .78 standard deviations. Thus, when ability and integrity test scores are optimally weighted, the Black–White difference in standard deviation units is reduced by 22% in comparison with 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 that all those above the White mean were selected (i.e., a selection ratio of .50 for Whites). In this case, assuming normality of the scores, the percentage of Blacks selected solely on the basis of ability, without an integrity test, would be 15.9%. However, if an integrity and an ability test were used together, with scores optimally weighted in a regression equation, the percentage of Blacks selected would increase to 21.8%. This would be an increase in the hiring rate of Blacks of 37.3%. This increase would be effected with no reduction in mean job performance of selectees; in fact, because validity increases, mean job performance increases. Equal weighting, rather than regression weighting, of the ability and integrity measures would reduce the mean Black–White difference somewhat more, to .67 standard deviations. With equal weighting, the percentage of Blacks selected in our example would be 25.1%, representing an increase of 58% in the hiring rate for Blacks. However, equal weighting would reduce validity and utility somewhat in comparison with regression weighting.

Even though the use of integrity tests alone should produce no adverse impact, it can be expected to result in a loss of utility of at least 37% in comparison with the use of ability tests and integrity tests in combination. Alternatively, using a composite of ability and integrity tests in selection can be expected to result in improved utility of at least 58% in comparison with using integrity tests alone. (These calculations were based on the figures in Table 12.) Hence, the implication is that employers seeking to maximize work-force output should use both integrity tests and measures of general mental ability in making hiring decisions. This combination has the potential for simultaneously reducing adverse impact and enhancing validity and utility, in comparison with selecting on ability alone. This analysis has focused on utility gains from increases in job performance. In addition to these increases, use of integrity tests can be expected to result in utility gains from reductions in counterproductive behaviors. Questions related to adverse impact and utility of integrity tests are explored in detail in Ones, Viswesvaran, and Schmidt (1992). In summary, one of the practical implications of the present study is that the most commonly used selection procedure could be a combination of general mental ability scores and an integrity test (or equivalent measure of conscientiousness).

Implications for Theory Development

Beyond the practical implications for enhanced validity and utility and reduced adverse impact, the findings of these meta-analyses have implications for the theoretical understanding of job performance. Specifically, these findings indicate that the broad construct measured by integrity tests, which we hypothesize is general conscientiousness, may be a major causal determinant of job performance (see Table 8), along with general mental ability and length of experience on the job (Schmidt, Hunter, & Outerbridge, 1986). These findings raise the question of whether general conscientiousness is actually the motivation variable that has been so elusive in personnel psychology (Schmidt, Ones, & Hunter, 1992); that is, conscientiousness may be the most important trait motivation variable. This question is discussed in more detail in Schmidt and Hunter (1992) in connection with causal models of the determinants of job performance. Considerably more research will be needed on this question in the future.

Discussion

One question we have repeatedly pondered since beginning our research on integrity tests has been the issue of potential response distortion by test takers, including the possibility of faking, responding in a socially desirable manner, or otherwise responding inaccurately. The conclusion we inferred from our meta-analytic results was that response distortion, to the extent that it exists, does not seem to destroy the criterion-related validities of these tests. (Similar findings were reported by Hough, Eaton, Dunnette, Kamp, & McCloy, 1990.) We found substantial validities for studies conducted on applicants. Applicants in these studies experienced all the usual inducements for response distortion, yet, nevertheless, we observed substantial estimated mean validities.

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 (U.S. OTA, 1990). This argument implicitly assumes that all applicants would be accepted if an integrity test were not used. Such an assumption is untenable in a selection setting. The failure to use any valid selection predictor will result in a higher false-positive rate than will 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 error 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, that part of the utility of integrity tests that results from reductions in theft and dishonesty does depend on the base rate of dishonesty in the applicant pool. The larger this base rate (up to 50%), the greater will be this 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. (However, the part of utility that results from prediction of overall job performance should not be affected by this consideration.)

Some limitations of the present study need to be pointed out. First, in some of the fully hierarchical moderator analyses, the number of existing studies was 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, 1990b, pp. 411–450). But even with this limitation, a meta-analytic review based on a reasonable conceptual or theoretical framework provides sounder conclusions than do other approaches to understanding the data, including the traditional narrative review.

A second limitation of this study is its inability to conclusively determine the validities of integrity tests as a function of job complexity. Nonetheless, our exploratory moderator analyses suggested that the mean validity of integrity tests for predicting job performance is about the same in high- and low-complexity jobs and the mean validity for predicting counterproductive behaviors is highest for high-complexity jobs. This latter result may imply increased opportunity to be dishonest in higher complexity jobs. Such opportunities 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 (about 10%) 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 research in this area become widely known, researchers will be more willing to submit more complete research reports for publication, and journals will be more likely to publish studies in this area.

This validity generalization effort is noteworthy in three respects. First, most of the studies reporting criterion-related validities for integrity tests came from service jobs (the largest sector of the U.S. economy), although some validities for manufacturing jobs were reported. Second, the meta-analysis of integrity tests was based on one of the largest databases in the literature (665 validity coefficients based on 576,460 data points). Even in the domain of mental abilities, few databases have been this large. Before beginning this research, we would not have estimated that the extant database for integrity tests was this large. Third, this study illustrated the usefulness of meta-analysis, given adequate data, for identifying moderators in heterogeneous research domains. The moderators identified were mostly methodological moderators (e.g., broad vs. narrow criteria, admissions vs. externally measured behaviors, predictive vs. concurrent study designs, and applicants vs. employees as subjects) rather than substantive moderators. Job complexity was the only substantive moderator that we identified. We did not identify any specifically situational moderators. Methodological moderators affect validity estimates but not operational (true) validities (Schmidt & Rothstein, in press). Nevertheless, it is important to identify methodological moderators and to quantify their effects. This study illustrated how meta-analysis can be used to do this. In particular, the study illustrated the power of fully hierarchical meta-analysis in identifying moderators.

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, or excessive 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 3,000 employers, 7 of the top 8 qualities were related to integrity, trustworthiness, conscientiousness, and related qualities. The other quality in the top 8 (ranked fifth) referred to general mental ability.

Additional research is needed on the construct validity of integrity tests. With the exception of Woolley and Hakstian (1992) and Collins and Schmidt (1993), there has been little research aimed at determining what constructs are measured by integrity tests. We currently have work underway to investigate 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); such research is critical if the theoretical meaning of the findings observed in this study is to be determined.

When we started our research on integrity tests, we, like many other industrial psychologists, were skeptical of integrity tests used in industry. Now, on the basis of analyses of a large database consisting of more than 600 validity coefficients, we conclude that integrity tests have 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 externally measured and self-reported counterproductive behaviors. Our meta-analyses confirmed 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. M., & 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 No. 22). Chicago: 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. In *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.

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). Orem, UT: CHC Forecast.

Collins, J. M., & Schmidt, F. L. (1993). Integrity, personality, and white-collar crime: A construct validity study. *Personnel Psychology, 46,* 295–311.

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.

Dictionary of occupational titles. (1991). Washington, DC: U.S. Government Printing Office.

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 factor structure. *Journal of Personality and Social Psychology, 59,* 1216–1229.

Goldberg, L. R., Grenier, J. R., Guion, R. M., Sechrest, L. B., & Wing, H. (1991). *Questionnaires used in the prediction of trustworthiness in preemployment selection decisions.* Washington, DC: American Psychological Association.

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. (No date). *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.

Hough, L. M., Eaton, N. K., Dunnette, M. D., Kamp, J. D., & McCloy, R. A. (1990). Criterion-related validities of personality constructs and the effect of response distortion on those validities. *Journal of Applied Psychology, 75,* 581–595.

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

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). Dichotomization of continuous variables: The implications for meta-analysis. *Journal of Applied Psychology, 75,* 334–349.

Hunter, J. E., & Schmidt, F. L. (1990b). *Methods of meta-analysis:*

Correcting error and bias in research findings. Newbury Park, CA: Sage.

Hunter, J. E., & Schmidt, F. L. (in press). The estimation of sampling error variance in the meta-analysis of correlations: The homogeneous case. *Journal of Applied Psychology.*

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.

Johnston, W. B., & Packer, A. E. (1987). *Workforce 2000: Work and workers for the twenty-first century.* Indianapolis, IN: Hudson Institute.

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, 3,* 85–94.

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

Law, K. S., Schmidt, F. L., & Hunter, J. E. (1992a). *Non-linearity of range corrections in meta-analysis: A test of an improved procedure.* Manuscript submitted for publication.

Law, K. S., Schmidt, F. L., & Hunter, J. E. (1992b). *A test of two refinements in meta-analytic procedures.* Manuscript submitted for publication.

London House Press. (1975). *Personnel Selection Inventory manual.* Park Ridge, IL: Author.

London House Press. (1980). *Personnel Selection Inventory manual.* Park Ridge, IL: Author.

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

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

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.* Ann Arbor: Author.

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

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

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.

Paajanen, G. (1985). *PDI Employment Inventory.* Minneapolis, MN: Personnel Decisions.

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.

Reid Psychological Systems. (1951). *Reid Report.* Chicago: 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., Burris, L. R., & Callahan, C. (1989). Integrity testing for personnel selection: An update. *Personnel Psychology, 42,* 491–529.

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.

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

Schmidt, F. L., & Hunter, J. E. (1992). Development of causal models of processes determining job performance. *Current Directions in Psychological Science, 1,* 89–92.

Schmidt, F. L., Hunter, J. E., & Outerbridge, A. N. (1986). Impact of job experience and ability on job knowledge, work sample performance, and supervisory ratings of job performance. *Journal of Applied Psychology, 71,* 432–439.

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.

Schmidt, F. L., & Rothstein, H. R. (in press). Application of validity generalization methods of meta-analysis to biographical data scales used in employment selection. In M. D. Mumford & G. S. Stokes (Eds.), *Advances in biodata research.* Palo Alto, CA: Consulting Psychologists Press.

Science Research Associates. (1983). *Personal Outlook Inventory.* Park Ridge, IL: Author.

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

Strand, A. L., & Strand, M. L. (1986). *Validity, reliability, compliance, and efficacy studies of the TA AIESDA and PASS surveys* (Tech. Rep.). Oak Brook, IL: Psychological Surveys Corporation.

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

Thorndike, R. L. (1949). *Personnel selection.* New York: Wiley.

U.S. Congressional Office of Technology Assessment. (1990). *The use of integrity tests for pre-employment screening* (Report No. OTA-SET-442). Washington, DC: U.S. Government Printing Office.

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, Quebec, 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. (1992). An examination of the construct validity of personality-based and overt measures of integrity. *Educational and Psychological Measurement, 52,* 475–489.

Appendix

Studies Coded for the Meta-Analysis of Integrity Tests

The following is a list of the studies that were coded for the meta-analyses in this article. In addition to using these studies, we personally contacted the following individuals many times in writing and by phone for additional data and clarification of ambiguous reporting in studies: Gerald Borofsky, Joan Brannick, Paul Brooks, David Cherrington, Brian Durbrow, Harrison Gough, William Harris, Joyce Hogan, Robert Hogan, Leatta Hough, Robin Inwald, John Jones, Dennis Joy, Scott Martin, John Miner, George Paajanen, Fred Rafilson, Alan Strand, and Mark Strand.

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 [Summary]. *Proceedings of the 78th Annual Convention of the American Psychological Association, 5,* 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 No. 3). Chicago: Reid Psychological Systems.

Ash, P. (1987). *Honesty test scores, biographical data, and delinquency indicators.* Paper presented at 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. In *Applied psychology: An international review.*

Berland, A. C. (1990). *Psychometric assessment of honesty testing for personnel selection.* Unpublished master's thesis, Iowa State University, Ames.

Berte, D. L., Moretti, D. M., Jusko, R., & Leonard, J. (1984). *An investigation of a combined withdrawal and counterproductive behavior decision model* (Tech. Rep. No. 13). Park Ridge, IL: London House Press. (Also presented as a paper at the 2nd Annual National Graduate Student Convention in Industrial and Organizational Psychology–Organizational Behavior, East Lansing, MI.)

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

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

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

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

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

Brooks, P., & Arnald, D. (1989). *Integrity attitudes as predictors of public safety organizational criteria* (Research Rep. No. 29). Chicago: Reid Psychological Systems.

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

Carpenter, B. N. (No date). *Reliability and motivational distortion in the Wilkerson pre-employment audit* (Tech. Rep.). Houston, TX: Team Building Systems.

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

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., Bernardin, 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 master's thesis, Johnson and Wales University, Providence, RI.

Cunningham, M. R. (1988). *Summary statistics for the Reid Drug Scale across five validation samples* (Research Memorandum No. 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 98th Annual Convention of the American Psychological Association, Boston.

Debro, J., & Murty, K. S. (No date). *Validity and reliability of test scores of measuring integrity and honesty.* Atlanta, GA: 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.

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

Durbrow, B. R. (No date). *Summary of study results.* Cincinnati, OH: Barbrisons Management Systems.

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,* 269–277.

Gerhardt, V. (No date). *Research and statistics: Pre-employment analysis questionnaire "H."* Dallas, TX: Psychometric Behavioral Group.

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

Gough, H. G. (1971). The assessment of wayward impulse by means of the Personnel Reaction Blank. *Personnel Psychology, 24,* 669–677.

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

Grimsley, D. L. (1986a). *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. (1986b). *Test–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, OK: University of Tulsa.

Harris, M. M. (1980). *A research statement on the Reid report.* Unpublished manuscript, University of Illinois at 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. (No date). *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* (Tech. Rep. No. 1409). Author.

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

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

Hogan, R. (1991). [Correlations between the HPI reliability scale (REL) and various scales]. Unpublished raw data.

Hogan, R. (No date). *Profiles interpretive manual.* Tulsa, OK: Hogan Assessment 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 edition). In J. Hogan & R. Hogan (Eds.), *Business and industry testing* (pp. 434–440). Austin, TX: Proed.

Inwald, R. E. (1988). Five-year follow-up study of departmental terminations as predicted by 16 pre-employment psychological indicators. *Journal of Applied Psychology, 73,* 703–710.

Inwald, R. E. (1990a). *IPI predictions of private security personnel performance* [Research Abstract]. New York: Hilson Research.

Inwald, R. E. (1990b). *Predicting the performance of private security personnel with the 153 & HPP/SO* [Research Abstract]. New York: Hilson Research.

Inwald, R. E., & Brockwell, A. L. (1989). *Predicting the performance of government security personnel with the IPI & MMPI.* New York: Hilson Research.

Inwald, R. E., & Knatz, H. F. (1988). *Seven-year follow-up of officer terminations predicted by psychological testing.* New York: Hilson Research.

Inwald, R. E., Knatz, H. F., & Shusman, E. J. (1982). *Inwald Personality Inventory technical manual.* New York: Hilson Research.

Inwald, R. E., & Patterson, T. (1989). *Use of psychological testing to predict performance of law enforcement trainees.* New York: Hilson Research.

Inwald, R. E., & Shusman, E. J. (1984). The IPI and MMPI as predictors of academy performance for police recruits. *Journal of Police Science and Administration, 12*(1), 1–11.

Jones, J. W. (1979). *Employee deviance: Attitudinal correlates of theft and on-the-job alcohol abuse* (Tech. Rep. No. 5). Park Ridge: IL: London House Press. (Also presented as a paper at the 5th Annual Convention of the Society of Police and Criminal Psychology, Chicago.)

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

Jones, J. W. (1980b). 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, August). *Personality profiles of endorsers of nuclear crime.* Paper presented at the 89th Annual Convention of the American Psychological Association, Los Angeles.

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

Jones, J. W. (1982b). *Dishonesty, staff burnout, and employee theft* (Tech. Rep. No. E2). Park Ridge, IL: London House Press. (Also presented as a paper at the 89th Annual Convention of the American Psychological Association, Los Angeles, 1981, August.)

Jones, J. W. (1982c). *Psychological predictors of employee theft* (Tech. Rep. No. 17). Park Ridge, IL: London House Press. (Also presented as a paper at the 90th Annual Convention of the American Psychological Association, Washington, DC, 1982, August.)

Jones, J. W. (1983a). *The Employee Attitude Inventory: A validity study on employees' theft in the auto parts industry* (Tech. Research Rep.). Park Ridge, IL: London House Press.

Jones, J. W. (1983b). *Validation of the Employee Attitude Inventory with retail employees* (Tech. Research Rep.). Park Ridge, IL: London House Press.

Jones, J. W. (1984). *Psychologically identifying current employee thieves* (Tech. Rep. No. 36). Park Ridge, IL: London House Press.

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 Press.

Jones, J. W. (Ed.). (1991). *Pre-employment honesty testing.* New York: Quorum Books.

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

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

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

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., 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., & Scruggs, D. P. (1981, April). *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, June). *Predictive validation of a dishonesty test that measures theft proneness.* Paper presented at the 18th Inter-American Congress of Psychology, Santo Domingo, Dominican Republic.

Jones, J. W., & Terris, W. (1982a, August). *Attitudes toward theft as predictors of employee theft among sales clerks.* Paper presented at the 90th Annual Convention of the American Psychological Association, Washington, DC.

Jones, J. W., & Terris, W. (1982b). *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. (1982c). *The use of the Personnel Selection Inventory to predict turnover in a group of security guards: A pilot study* (Tech. Rep. No. 21). Park Ridge, IL: London House Press.

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

Jones, J. W., & Terris, W. (1983a). Personality correlates of theft and drug abuse among job applicants. *Proceedings of the Third International Conference on the 16PF Test, 3,* 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 Attitude Inventory to assess dishonesty among hotel employees* (Tech. Research Rep.). Park Ridge, IL: London House Press.

Jones, J. W., & Terris, W. (1984). *Psychological correlates of employee theft in department stores* (Tech. Rep. No. 20). Park Ridge, IL: London House Press.

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

Jones, J. W., & Wuebker, L. J. (1988a). Accident prevention through personnel selection. *Journal of Business and Psychology, 34,* 187–198.

Jones, J. W., & Wuebker, L. J. (1988b). Pre-employment screening to control hospital crime and counterproductivity. *Readings in Health Care Risk Management, 35,* 93–96.

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.

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

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

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* (Tech. Research Rep. No. 55). Park Ridge, IL: 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* (Tech. Research Rep. No. 63). Park Ridge, IL: London House Press.

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

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

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

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

Klump, C. (1980). *The Stanton Survey manual: Description and validation.* Chicago: Stanton Corporation.

Knatz, H. F., Inwald, R. E., Brockwello, A. L., & Tran, L. N. (1989). *IPI and MMPI predictions of counterproductive job behaviors by group.* New York: Hilson Research.

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.

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

London House Press. (1980). *Personnel Selection Inventory manual.* Park Ridge, IL: Author.

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

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

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

Malin, S. Z., Luria, J., & Morgenbesser, L. I. (1987, August–September). *New York State Pre-Employment Psychological Screening Program: Longitudinal validation study.* Paper presented at the 95th Annual Convention of the American Psychological Association, New York.

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, University of Ohio, Athens.

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 for 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* (Tech. Rep. No. 34). Park Ridge, IL: London House Press.

Moretti, D. M. (1983). *A note on the validity of the Personnel Selection Inventory distortion scale* (Tech. Rep. No. 33). Park Ridge, IL: London House Press.

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.

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

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, Atlanta.

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.

Paajanen, G. (1990). *Validity of the PDI Employment Inventory for drug stores* (Tech. Rep.). Minneapolis, MN: Personnel Decisions.

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

Personnel Decisions Incorporated. (1986a). *Summary of findings—Retail industry.* Minneapolis, MN: Author.

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

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

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

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

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

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

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

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

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

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

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

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

Personnel Decisions Incorporated. (1988g). *Validation of the PDI Employment Inventory for food wholesalers.* Minneapolis, MN: Author.

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

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

Personnel Decisions Incorporated. (1988j). *Validation of the PDI Employment Inventory for national discount fashion retailers.* Minneapolis, MN: Author.

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

Personnel Decisions Incorporated. (1988l). *Validation of the PDI Employment Inventory for seafood restaurants.* Minneapolis, MN: Author.

Personnel Decisions Incorporated. (1988m). *Validation of the PDI Employment Inventory for supermarkets.* Minneapolis, MN: Author.

Personnel Decisions Incorporated. (1988n). *Validation of the PDI Employment Inventory for toy stores.* Minneapolis, MN: Author.

Personnel Decisions Incorporated. (1988o). *Validation of the PDI Employment Inventory for video stores.* Minneapolis, MN: Author.

Personnel Security Corporation. (No date). *Survey description manual: T. A. Survey.* Oak Brook, IL: Author.

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

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

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., & 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: Stanton Corporation.

Reid Psychological Systems. (1989). *Reid Report: Examiner's manual* (3rd ed.). Chicago: Author.

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

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

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

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

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

Shusman, E. J., & Inwald, R. E. (1984, August). *A longitudinal validation study of correction officer job performance as predicted by the IPI and MMPI.* Paper presented at the 92nd Annual Convention of the American Psychological Association, Toronto, Ontario, Canada.

Shusman, E. J., Inwald, R. E., & Knatz, H. 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, E. J., Inwald, E. E., & Landa, B. (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 master's thesis, Radford University, Radford, 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.), *Pre-employment honesty testing: Current research and future directions* (pp. 171–183). Westport, CT: Quorum Books.

Springborn, B. A. (No date). *Concurrent validity study of the Phase II Profile Addendum.* Lakeland, FL: Kelly/Springborn Associates.

Statistical Consulting Associates. (1986). *Preliminary validation of the drug and alcohol usage items from the Wilkerson Pre-Employment Audit.* (Available from Team Building Systems, 363 North Belt, Suite 990, Houston, Texas 77060.)

Strand, A. L., & Strand, M. L. (1986). *Validity, reliability, compliance, and efficacy studies of the TA, AI, ES, DA, and PASS Surveys* (Tech. Rep.). Oak Brook, IL: Psychological Surveys Corporation.

Terris, W. (1979, July). *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 Press.

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 Press.

Terris, W., & Jones, J. W. (1982b). *Psychological correlates of employee* *theft in department stores* (Tech. Rep. No. 20). Park Ridge, IL: London House Press.

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

U.S. Congressional Office of Technology Assessment. (1990). *The use of integrity tests for pre-employment screening* (Report No. OTA-SET-442). Washington, DC: U.S. Government Printing Office.

Weichman, D. J., & Bae, R. P. (1983). *The Personnel Selection Inventory: An empirical examination* (Tech. Rep. No. 14). Park Ridge, IL: London House Press.

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.

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

Received February 1, 1991
Revision received October 1, 1992
Accepted October 5, 1992 ■

Low Publication Prices for APA Members and Affiliates

Keeping You Up-to-Date: All APA members (Fellows; Members; Associates, and Student Affiliates) receive–as part of their annual dues–subscriptions to the *American Psychologist* and *APA Monitor.*

High School Teacher and International Affiliates receive subscriptions to the *APA Monitor,* and they can subscribe to the *American Psychologist* at a significantly reduced rate.

In addition, all members and affiliates are eligible for savings of up to 60% (plus a journal credit) on all other APA journals, as well as significant discounts on subscriptions from cooperating societies and publishers (e.g., the American Association for Counseling and Development, Academic Press, and Human Sciences Press).

Essential Resources: APA members and affiliates receive special rates for purchases of APA books, including the *Publication Manual of the APA,* the *Master Lectures,* and *Journals in Psychology: A Resource Listing for Authors.*

Other Benefits of Membership: Membership in APA also provides eligibility for low-cost insurance plans covering life, income protection, office overhead, accident protection, health care, hospital indemnity, professional liability, research/academic professional liability, student/school liability, and student health.

For more information, write to American Psychological Association, Membership Services, 750 First Street, NE, Washington, DC 20002-4242, USA