Validity of Integrity Tests for Predicting Drug and Alcohol Abuse: A Meta-Analysis

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

INTRODUCTION

The research described in this chapter used psychometric meta-analysis (Hunter and Schmidt 1990*b*) to examine the validity of integrity tests for predicting drug and alcohol abuse. Integrity tests have previously been found to predict other counterproductive workplace behaviors (e.g., absenteeism, property damage, and violence on the job) (Ones et al. 1993; Ones et al., unpublished observations). All studies located were concurrent in nature. For both drugs and alcohol, integrity tests correlated substantially (0.34 to 0.51) with admissions of abuse in student and employee samples. In samples of job applicants, however, the mean validity was lower (0.21) for drug abuse; validity for applicants was high for alcohol abuse, but only one study (N = 320) was found. All meta-analyses indicated that validity was generalizable. Based on these analyses, the authors conclude that the operational validity of integrity tests for predicting drug and alcohol abuse in the workplace is probably about 0.30. But further research is needed; predictive validity studies conducted on applicants would be particularly useful.

THE PROBLEM OF SUBSTANCE ABUSE

Substance abuse is a major societal problem. Numerous surveys (Johnston et al. 1994; Miller et al. 1983) have found that substance abuse, especially the abuse of alcohol and marijuana, is prevalent. Epidemiological surveys (Simpson et al. 1975) indicate that illicit drug abusers are predominantly young adults.

The relationships between substance abuse, job performance, and other job-related behaviors have been studied. In a large sample study of military personnel, McDaniel (1988) found that individuals who reported using drugs at earlier ages were more likely to be rated as unsuitable for service by their supervisors than a control group who indicated they did not use drugs when younger. In a sample of Navy recruiters, Blank and

Fenton (1989) found individuals testing positive for drugs had more behavioral and performance problems than individuals who tested negative for drugs.

Normand and colleagues (1990) found that postal employees who tested positive for substance abuse were more likely to be absent from work. Further, Winkler and Sheridan (1989) found that employees who entered employee assistance programs for drug addiction treatment were more likely to be absent, had twice the number of worker compensation claims, and used more than twice as many medical benefits as a matched control group. Crouch and colleagues (1989) found that drug use correlated with increased accident and absence rates.

Substance abuse has been found to be related not only to measures such as absenteeism, turnover, accidents, and productivity, but also to behaviors such as stealing on the job, violence, and effort expenditure (i.e., not daydreaming) on the job. In fact, Viswesvaran (1993) found that these various measures are positively correlated and a general factor exists across them, suggesting that the various measures of job performance may be influenced in part by the same underlying construct (presumably a personality dimension).

In addition to the above-mentioned studies that compare drug-using individuals to a matched set of controls on various job performance measures, laboratory studies have also found that substance abuse leads to impairment in performance of various experimental tasks (Herning et al. 1989; Jobs 1989; Streufert et al. 1991; Yesavage et al. 1985). Impairments in information-processing capabilities, decisionmaking, and quickness of reflexes have been found to result from drug or alcohol consumption.

With surveys indicating that abuse of alcohol and other drugs is prevalent in the general population and studies indicating a negative relationship between substance abuse and job performance, employers have tried different strategies to ensure a drug-free workplace. Coworkers, customers, and the general public also have a stake in ensuring a drug-free workplace. The growing concern of employers about drug abuse has resulted in increased testing of both current and prospective employees for drug abuse (Guthrie and Olian 1989).

A survey of the literature indicates that an employer's choice of strategies in drug testing is mainly based on four considerations: the validity and reliability of the techniques used to detect substance abuse; the legal viability of the techniques; the practicality and cost of

employing the techniques; and whether employees accept the use of a technique as justified.

Validity refers to whether the technique is measuring what it purports to measure. Reliability indicates whether the measurements are stable and replicable. Legal viability refers to the employers' concerns about whether the courts and arbitrators will accept the findings of the technique. In fact, studies have shown (Hill and Sinicropi 1987) that courts and arbitrators place considerable weight on the reliability and validity of the technique used in deciding cases involving substance abuse. Thus, the technique's validity and reliability have an indirect as well as a direct effect on the strategies used by the employers to combat substance abuse.

Employee perceptions of a drug testing program's acceptability have been widely researched. Negative employee reactions to drug testing, if ignored, may lead to lowered commitment and subsequent reduction in performance (Crouch et al. 1989). Knovsky and Cropanzano (1991) present data indicating that employee reactions to drug testing can be analyzed within an organizational justice framework (Adams 1965; Greenberg 1990). Specifically, Knovsky and Cropanzano (1991) found that perceptions of procedural justice affect reactions to drug testing. Two of the key elements in shaping perceptions of procedural justice are the validity, reliability, and psychometric properties of the testing procedures; and invasions of privacy concerns. Other elements include job characteristics, such as situations when impaired performance results in dangers to others (Stone and Vine 1989); the type of drug used (Murphy et al. 1990); the type of personnel action taken against employees testing positive (Gomez-Mejia and Balkin 1987; Stone and Kotch 1989); the role of explanations (Bies 1987; Bies and Shapiro 1987; Crant and Bateman 1989); the chance to appeal; the availability of advance notice; and whether random drug testing or testing with due cause is implemented. Employee objections could result in union contracts restricting the use of certain techniques for detecting substance abuse. Further, courts and arbitrators are likely to give some weight in their decisions to employee and applicant objections. Thus, employee acceptance has both direct and indirect effects (through legal acceptability) on the strategies used by an employer.

In short, the method's validity and reliability affect legal defensibility of the procedures and acceptability to test takers, as well as directly affecting the employer's choice of technique used. Further, validity and reliability affect employer strategies through an effect on legal defensibility and acceptability to test takers. Thus, it is of paramount interest to examine a

procedure's validity and psychometric properties to realize the benefits of drug testing without any loss of employee commitment.

Several approaches have been tried to detect drug abuse. Blood testing, breath analyzers, and urinalysis are some of the common approaches to drug testing and detection. One technique gaining prominence in employment settings is the use of paper-and-pencil preemployment integrity tests to assess a job applicant's predisposition to drug and alcohol abuse. Evidence available to date indicates that applicants do not object to such tests (Stone and Kotch 1989; Stone and Bommer 1990; Stecker and Rosse 1992). To the extent that selection methods can be used to eliminate drug abusers at the point of hire, drug testing programs for employees become less necessary.

INTEGRITY TESTS

Defining Integrity Tests

Integrity tests are designed to measure the predisposition of individuals to engage in counterproductive behaviors on the job. Integrity tests are paper-and-pencil tests, as opposed to other methods such as the polygraph (a physiological method), background investigations, interviews, and reference checks. These tests have been developed for use with applicants and employees (a normal population); hence instruments such as the Minnesota Multiphasic Personality Inventory (MMPI), which were designed for use with mentally ill populations, are not classified as integrity tests, even though some organizations claim to use them for screening out delinquent applicants. Most integrity tests have been initially designed to predict a variety of counterproductive behaviors; only later were they found to predict other criteria such as supervisory ratings of overall performance (Ones et al. 1993).

A Brief History of Integrity Tests

The first paper-and-pencil psychological test to assess the integrity of potential employees, the Personnel Reaction Blank, was developed in 1948 (Gough 1948). It was a derivative of what was then called the Delinquency Scale of the California Psychological Inventory. (This scale was later renamed the Socialization Scale.) In 1951 a second type of test, intended to assess honesty of job applicants, was developed. This test, the Reid Report, was a compilation of questions that seemed to distinguish honest and dishonest individuals during polygraph examinations. Since then several other instruments have been developed and used to select

applicants on the basis of integrity. A complete treatise on the history of integrity tests can be found in Ash (1989) and Woolley (1991).

There is relatively little information about which companies use paper-and-pencil integrity tests. According to Sackett and Harris (1985), as many as 5,000 companies may use preemployment integrity tests, assessing about 5 million applicants yearly. A variety of surveys of companies indicate that anywhere between 7 percent to 20 percent of all companies in the United States could be testing for integrity, at least for some jobs (American Society for Personnel Administration 1988; Blocklyn 1988; Bureau of National Affairs, Inc. 1988; O'Bannon et al. 1989). Even by the most conservative estimates, millions of people in the United States either have been or are being tested using integrity tests. There are at least 43 integrity tests in current use. Of these tests, about one-quarter seem to be small operations without much market share; 16 to 19 tests overall seem to serve most of the demand for integrity tests. However, this demand can be expected to increase, because in 1988 the Federal Polygraph Act effectively banned the use of the polygraph in most employment settings.

Over the last 15 years, scientific interest in 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 and Rieke 1991; Sackett et al. 1989; Sackett and Decker 1979; Sackett and Harris 1984). Recently Sackett and colleagues (1989) and O'Bannon and colleagues (1989) have provided extensive qualitative reviews and critical observations regarding integrity testing. In addition to these reviews, the U.S. Congressional Office of Technology Assessment (OTA 1990) and the American Psychological Association (APA) (Goldberg et al. 1991) have each released papers on integrity tests. The OTA paper (1990) was in part prompted by Congress' regulation of the polygraph. The OTA recommendations were based on a limited number of chosen studies and ignored most of the literature on integrity tests. Compared to the OTA paper, the APA report (Goldberg et al. 1991) was more thorough, objective, and insightful. It provided a generally favorable conclusion regarding the use of paper-and-pencil integrity tests in personnel selection.

Personality Constructs Underlying Integrity Tests

Sackett and colleagues (1989) classify honesty tests into two categories: overt integrity tests and personality-based tests. Overt integrity tests (also known as clear purpose tests) are designed to directly assess attitudes regarding dishonest behaviors. Some overt tests specifically ask about

past illegal and dishonest activities as well; for several tests, admissions are not a part of the instrument, but instead are used as the criterion. Overt integrity tests include the London House Personnel Selection Inventory (PSI) (London House, Inc. 1975), Employee Attitude Inventory (EAI) (London House, Inc. 1982), Stanton Survey (Klump 1964), Reid Report (Reid Psychological Systems 1951), Phase II Profile (Lousig-Nont 1987), Milby Profile (Miller and Bradley 1975), and Trustworthiness Attitude Survey (Cormack and Strand 1970). According to Sackett and colleagues (1989), "[T]he underpinnings of all these tests are very similar" (p. 493). Hence, high correlations may be predicted, and are found (Ones 1993), among overt integrity measures.

On the other hand, personality-based measures (also referred to as disguised purpose tests) aim to predict a broad range of counterproductive behaviors at work (e.g., violence on the job, absenteeism, tardiness, drug abuse, theft) via personality traits such as reliability, conscientiousness, adjustment, trustworthiness, and sociability. In other words, these measures have not been developed solely to predict theft or theft-related behaviors. Examples of personality-based measures used in integrity testing include the Personal Outlook Inventory (Science Research Associates 1983), the Personnel Reaction Blank (Gough 1954), the Employment Inventory (Paajanen 1985), and Hogan's Reliability Scale (Hogan 1981). 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, and energy level (O'Bannon et al. 1989). The similarity of integrity measures raises the question of whether they all measure primarily a single general construct. Detailed descriptions of all the above tests can be found elsewhere (Conoley and Kramer 1989), particularly in the extensive literature reviews (O'Bannon et al. 1989; Sackett et al. 1989; Sackett and Harris 1984).

Using both primary data (N = 1,365) and meta-analytic cumulation, Ones (1993) found that a general factor exists across different integrity tests. Ones (1993) found that the variance common to all integrity tests correlated highest with the personality dimension of conscientiousness, followed by emotional stability (neuroticism) and agreeableness. Based on these comprehensive analyses, researchers can conclude that integrity tests tap into the personality dimensions of conscientiousness, agreeableness, and emotional stability. This finding is significant; researchers now can focus on the theoretical construct underlying the different measures rather than investigating each measure separately as if it were unique. All theoretical propositions and causal explanations are stated in terms of constructs and not measures (Nunnally 1978).

Review of Causal Mechanisms: Why Personality Constructs Underlying Integrity Tests Should Predict Substance Abuse

In the literature, three causal mechanisms have been proposed that explain why personality constructs tapped into by integrity tests should predict substance abuse. First, Barrick and colleagues (1994) found evidence for the hypothesis that highly conscientious individuals set higher (or more difficult) goals for themselves and strive to accomplish them. Barrick and colleagues (1994) argued that individuals who set more difficult goals for themselves exhibit better job performance.

Further, Schmidt and Hunter (1992) noted that highly conscientious individuals can be expected to spend more time on task, which also contributes to better job performance. However, high-level job performance is usually incompatible with substance abuse (McDaniel 1988; Normand et al. 1990). Thus, integrity tests that seem to be assessing conscientiousness (Ones 1993) may also correlate with, and predict, substance abuse.

A second explanation lies in the social impulse control enunciated by Gough (1948). According to this explanation, substance abusers are likely to be individuals who have not learned the social skills necessary to function effectively in society and often have poor impulse control. From this perspective, it could be argued that scores on integrity tests found to correlate with measures of neuroticism (emotional stability) (Ones 1993) should also correlate with measures of substance abuse.

Finally, Zuckerman (1983) and others have posited that individuals differ in their proclivity to seek sensations. Individual differences in sensation seeking are reflected in differing personality measures of extroversion and agreeableness. Integrity tests are correlated with agreeableness (Ones 1993) and therefore may be related to substance abuse.

METHODS

A thorough search was conducted to locate all existing integrity test validities. All published empirical studies were obtained from published reviews of the literature (O'Bannon et al. 1989; Sackett et al. 1989; Sackett and Harris 1984), three other meta-analyses of integrity tests (Harris, undated; McDaniel and Jones 1986, 1988), and a computerized search to locate the most recent studies in psychological and management- related journals.

According to O'Bannon and colleagues (1989), there are 43 integrity tests in use in the United States. All the publishers and authors of the 43 tests were contacted by telephone or in writing requesting validity, reliability, and range restriction information on their tests. Of these, 36 responded with research reports. In addition, the authors identified other integrity tests overlooked by O'Bannon and colleagues (1989); their publishers 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 test publishers responded to the request for validity information on their test by sending computer printouts that had not been written up as technical reports. These were included in the database.

Still other integrity test publishers responded by sending raw data that had not been analyzed. In some instances, using the information supplied, the authors were able to calculate the phi correlation, and then correct it for dichotomization (Hunter and Schmidt 1990*a*). These corrected correlations were used in the meta-analysis. Sample sizes for these corrected correlations were adjusted to avoid underestimating the sampling error variance. First, the uncorrected correlation and the study sample size were used to estimate the sampling error variance for the observed correlation. This value was corrected for the effects of the dichotomization correction, and this corrected sampling error variance was then used with the uncorrected correlation in the standard sampling error formula to solve for the adjusted sample size, which was entered into the meta-analysis computer program. This process results in the correct estimate of the sampling error variance of the corrected correlation in the meta-analysis. The list of integrity tests contributing criterion-related validity coefficients, reliabilities, or range restriction information to this meta-analysis is presented in table 1.

TABLE 1. *Tests contributing data to the meta-analyses.*

Test Name
Accutrac Evaluation System[a]
Applicant Review[a]
Compuscan[a,c]
Employee Attitude Inventory (London House)[a]
Employee Reliability Inventory[a]
Employment Productivity Index[b]
Hogan Personnel Selection Series (Reliability Scale)[b]

Integrity Interview[a]
Inwald Personality Inventory[b]
Orion Survey[a,c]
P.E.O.P.L.E. Survey[a]
Personnel Decisions Inc. Employment Inventory[b]
Personal Outlook Inventory[b]
Personal Reaction Blank[b]
Personnel Selection Inventory (London House)[a]
Phase II Profile[a]
P.O.S. Preemployment Opinion Survey[a,c]
Preemployment Analysis Questionnaire[a]
Reid Report and Reid Survey[a]
Rely[a]
Safe-R[a,c]
Stanton Survey[a]
True Test[a]
Trustworthiness Attitude Survey; PSC Survey; Drug Attitudes/ Alienation Index[a]
Wilkerson Preemployment Audit[a,c]

NOTE: The list of publishers and authors of these tests are available in O'Bannon et al. 1989.

KEY: a = Overt integrity test; b = personality-based integrity test; c = no validity data were reported, but the test contributed to the statistical artifact distributions.

Some researchers have argued for the exclusion of unpublished studies in all meta-analyses based on misleading and erroneous arguments that such unpublished studies constitute poor quality data. The converse argument maintains that published studies have a positive bias that overstates the results. Taken together, these two arguments lead to scientific nihilism (Hunter and Schmidt 1990*b*). The hypothesis of methodological inadequacy of unpublished studies (in comparison to published studies) has not been established in any research area. In fact, evidence exists in many research areas indicating comparability of findings of published and unpublished studies (Hunter and Schmidt 1990*b*).

Hunter and Schmidt (1990*b*) present a hypothetical example that illustrates how differences between published and unpublished studies examining the effectiveness of psychotherapy could have been due to statistical artifacts. Ones and colleagues (unpublished observations) found that the correlation between the reported validity of integrity

tests and the dichotomous variable indicating published versus unpublished studies is negligible. In the literature on the validity of employment tests, impressive evidence has been accumulated indicating that published and unpublished studies do not differ in the validities reported (Hunter and Schmidt 1990*b*). For example, the data used by Pearlman and colleagues (1980) was found to be very similar to the Department of Labor's General Aptitude Test Battery (GATB) database used by Hunter (1983) and other large sample military data sets. The mean validities in Pearlman and colleagues' (1980) database are virtually identical to Ghiselli's (1966) reported medians. Further, the percentage of nonsignif-icant studies in Pearlman and colleagues' (1980) database perfectly matches the percent of nonsignificant published studies reported by Lent and colleagues (1971). Finally, the percentage of observed validities that were nonsignificant at the 0.05 level in Pearlman and colleagues' (1980) database (56.1 percent of the 2,795 observed validities) is consistent with the estimate obtained by Schmidt and colleagues (1976): The average criterion-related validation study has statistical power no greater than 0.50. If selectivity or bias in reporting were operating, many of the nonsignificant validities would have been omitted, and the percent significant should have been higher than 43.9 percent. On the other hand, if unpublished studies were of poorer quality, not meeting the standards of peer review, then there should have been more than 56 percent nonsignificant validities among the unpublished studies. Thus, there is ample evidence arguing for the equivalence of published and unpublished studies. The two databases are often comparable. Therefore, both published and unpublished reports are included.

Data Coded/Extracted From Primary Studies

An identification number was given to each study, and when more than one sample was reported in a study, a sample-within-study identification number was given to each sample within that study. Thus, each record contains a study identification number, a (within study) sample identifi-cation number, the validity coefficient, the sample size, the criterion used, whether the criterion measure was based on self-reports or external records, whether the sample was comprised of students or applicants for a job or current employees, and whether the validity coefficient was based on a predictive or a concurrent validation strategy. Wherever possible, the complexity levels of the jobs included in the analyses and other demographic characteristics were also coded.

Overall, 50 validation studies were located. Of these 50 studies, 24 had used employees as samples, 16 had used student samples, and the remaining 10 studies were based on applicant samples. All 50 studies employed the concurrent validation strategy. Forty-eight of the 50 studies had relied on admissions (self-reports) of substance abuse. There was one study conducted on a sample of 46 employees in a fire department that had used apprehension and conviction for substance abuse as the criterion. The observed validity coefficient in that study was 0.44. One study provided inadequate information as to whether admissions or external measures were employed. The observed validity coefficient in that study was 0.62, and it was based on a sample of 320 job applicants.

The admissions criterion was measured using self-report questionnaires. Measures of admissions of drug abuse included questions on number and type of illegal drugs used, number of times one had become high from drug use, and so forth. Measures of admissions of alcohol abuse included questions on frequency of alcohol intoxication, number of drinks consumed on the job, number of drinks on work breaks and during lunch on work-days, and number of alcohol-related problems. The final score was the sum (sometimes weighted) of such admissions.

Twenty of the 50 studies were conducted in the Midwest while 4 were conducted in the Northwestern region of the United States. Thirteen of the 50 studies were conducted in supermarket or grocery stores or convenience stores or on gas station employees. Seven of the 50 studies were done using security personnel as the sample. One study was conducted in a fire department while another was in a fast-food chain. Twenty studies focused on alcohol consumption while the remaining 30 used drug abuse as the criterion.

Given this set of validity coefficients, only two potential moderators could be tested: sample type (students, employees, and applicants) and criterion type (drug abuse versus alcohol abuse).

Intercoder agreement in summarizing or extracting information from the primary studies is a concern in meta-analyses. Haring and colleagues (1981) presented empirical data indicating that intercoder agreement in meta-analyses is a function of the judgmental nature of the items coded. Haring and colleagues' review of meta-analyses found that eight of the nine items lowest in coder agreement were judgments (e.g., the quality of the study) as opposed to calculation-based variables (e.g., effect sizes, number of subjects). Jackson

(1980) and Hattie and Hansford (1982, 1984) also provided data indicating that problems of intercoder agreement in meta-analyses are negligible for coding computation-based numerical variables. Finally, Whetzel and McDaniel (1988) found no evidence of any coder disagreements in validity generalization databases. The intercoder agreement in the present research was over 85 percent for all categories coded. Disagreements between the two coders were resolved through discussion.

Psychometric Meta-Analyses

Data from the sources described in the previous section were cumulated by the methods of psychometric meta-analyses (Hunter and Schmidt 1990*b*). Depending on the availability of information in the primary studies, the meta-analysis can either correct the observed correlations for the effects of statistical artifacts and cumulate the individually corrected correlations, use artifact distributions to correct the observed distribution of correlations, or use a combination of individual corrections and artifact distributions.

Because the degree of split for dichotomization was given in the research reports, it was possible to correct the correlations individually for the attenuating effects of dichotomization (Hunter and Schmidt 1990*a*). But to correct for the effects of artifacts such as unreliability and range restriction, where the available information was sporadic, recourse was made to the use of artifact distributions. That is, a mixed meta-analysis was employed. In the first step, the correlations were corrected individually for the effects of dichotomization. In the second step, the partially corrected distribution obtained from the first step was corrected for sampling error, unreliability, and range restriction using artifact distributions (Hunter and Schmidt 1990*b*).

In using artifact distributions for correcting two or more artifacts, one has the option to use either the interactive procedure (which corrects the observed correlations for the effects of the various statistical artifacts simultaneously), or the noninteractive procedure (which sequentially corrects the observed correlation for the effects of the statistical artifacts). Recent computer simulation studies (e.g., Law et al. 1994; Schmidt et al. 1993) have shown that among the methods of psychometric meta-analyses, the interactive procedure used with certain refinements (e.g., nonlinear range restriction and mean observed correlation in the sampling error formula) is the most accurate one.

The use of the mean observed correlation in the sampling error formula provides a more accurate estimate of the sampling error variance (Hunter and Schmidt 1994). The sampling error variance formula for the correlation requires knowledge of the population correlation. In individual studies, the observed correlation is taken as an estimate of the population value because nothing better is available. But meta-analysts can be more precise by using the mean observed correlation across studies. This value is a better estimate of the population correlation than the individual observed correlation, which is strongly affected by sampling error unless sample sizes are large.

The second refinement involves the use of a nonlinear range-restriction correction formula in estimating the standard deviation (SD) of true validities. In artifact distribution-based meta-analyses, the mean and SD of the residual distribution (the distribution of observed correlations expected when sample sizes are infinite and reliability and range-restriction values are held constant across studies at their mean values) are corrected for the mean value of the artifacts. This procedure is accurate when the artifact corrections are linear (e.g., reliability corrections) because the correction is the same for every value of the correlation in the residual distribution. But the correction for range restriction is not linear; it is smaller for large correlations and larger for smaller correlations. This results in an overestimation of the true SD when the linear approximation is used. Computer simulation studies have shown that a new, nonlinear correction procedure is more accurate (Law et al. 1994). That new procedure was used in this study. More details of these refinements can be found in Schmidt and colleagues (1993), where examples are also provided to illustrate application of the refinements.

In correcting for unreliability in the measures, the use of the correct form of reliability coefficient requires the specification of the nature of the error of measurement in the research domain of interest (Hunter and Schmidt 1990*b*). Several sets of artifact distributions were compiled: one distribution for the reliability of the integrity tests, one distribution for the reliability of the criterion variables, and one distribution of range restriction values. Descriptive information on the artifact distributions is provided in table 2.

TABLE 2. *Descriptive information on statistical artifact distributions used to correct validities.*

	N of	Mean	SD	Mean of	SD of

	values			the square roots of reliabilities	the square roots of reliabilities
Integrity test reliabilities	124	0.81	0.11	0.90	0.06
Criterion reliabilities	13	0.84	0.13	0.94	0.07
Range restriction values[c]	79	0.81	0.19	--	--

KEY: c = The ratio of the selected group standard deviation to the referent group standard deviation (s/S).

A total of 124 integrity test reliability values was obtained from the published literature and the test publishers. Of the 124, 68 were alpha coefficients (55 percent) and 47 were test-retest reliabilities over periods of time ranging from 1 to 1,825 days (mean = 111.4 days; SD = 379.7 days). The mean of the coefficient alphas was 0.81 (SD = 0.10) and the mean of the test-retest reliabilities was 0.85 (SD = 0.10). There were nine reliabilities reported with no statement of the type of reliability. The ideal estimate of test reliability for purposes of this meta-analysis is coefficient alpha or the equivalent. However, test-retest reliability estimates usually provide reasonably close approximations to alpha coefficients. In this case the means of the two reliability types were similar. The overall mean of the predictor reliability artifact distribution was 0.81 and the SD was 0.11. The mean of the square roots of predictor reliabilities was 0.90 with an SD of 0.06.

No correction for predictor unreliability was applied to the mean true validity because the interest was in estimating the operational validities of integrity tests for selection purposes. However, the observed variance of validities was corrected for variation in predictor unreliabilities in addition to variation in criterion unreliabilities, range restriction values, and sampling error. For comparison purposes, the authors provide the percent variance due to sampling error alone in the results.

To estimate the reliability of the criterion measure, the authors reviewed the literature on delinquency. Viswesvaran and colleagues (1992) meta-analyzed correlations between admissions and external measures of delinquency; the mean correlation was found to be 0.50. That study compiled a reliability distribution for questionnaires measuring admissions of delinquent acts. This distribution consisted of 13 values of

coefficient alpha. The average of the reliability distribution was 0.84 and the SD was 0.13. The average of the square roots of the reliability estimates was 0.94 and the SD was 0.07. This distribution was used in the present study for admissions of alcohol and drug abuse.

Because integrity tests are used to screen applicants, the validity calculated using an employee sample may be affected by restriction in range. A distribution of range restriction values was constructed from the studies contributing to the database. There were 75 studies which reported both the SD in the study sample and the applicant group SD. The range restriction ratio was calculated as the ratio of study to reference group standard deviations (s/S). In four studies, correlations were reported for both the applicant and the employee groups. From these four studies, range restriction ratios were calculated by taking the ratio of the two correlations reported and solving for the range restriction value using the standard range restriction formula (case II formula, Thorndike 1949). Overall there were 79 range restriction values included in the artifact distribution. The mean ratio of the restricted sample SD to the unrestricted sample SD was 0.81 and the SD was 0.19; these figures indicate that there is considerably less range restriction in this research domain than is the case for cognitive ability (Alexander et al. 1989). Thus, range restriction corrections were much smaller in present research than in meta-analyses in the abilities domain. No range restriction corrections were made for student samples.

The parameters of interest estimated from a meta-analysis are the true validity, the SD of the true validity, and the 90 percent credibility value. From the observed distribution of validities, the authors estimated the distribution of true validities. There are four substantive inferences of interest here. First, the authors want to know the average validity coefficient across situations. This is captured in the mean true validity. Second, the authors want to know whether the validity coefficient will be positive across situations. To answer this question, the authors examined the 90 percent credibility value. The 90 percent credibility value indicates that in 90 percent of the situations, the validity coefficient will be higher than this value. As such, if the 90 percent credibility value is positive, one can conclude that the instrument has a validity coefficient that is positive in over 90 percent of the situations. That is, validity generalizes across situations.

The third substantive question involves an examination of the SD of true validities to examine the extent to which the validity varies across situations. In a meta-analysis, if the 90 percent 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 (i.e., validity generalizes), although the actual magnitude may vary across settings. However, the remaining variability may also be due to uncorrected statistical artifacts as well as methodological differences between studies. A final possibility is truly situationally specific test validities and/or the operation of moderator variables. In sum, the 90 percent credibility value is used to judge whether the validities are positive across situations (i.e., validity generalizes), whereas the estimated SD of true score validities is used to assess whether the estimated true validity is constant across situations.

Finally, to test for the moderating influence of a hypothesized moderator, the validity coefficients are grouped into subsets based on the hypothesized moderator. Psychometric meta-analyses are then conducted on each subset. If the hypothesized moderator exists, it will be reflected in the following findings: the mean true validity computed for each subset will vary across the subsets, and will vary from the mean true validity computed with the entire set of validities across subsets; and the average SD of true validities in the subsets will be lower than the overall SD. The above two results are interrelated as the group means and variances in the analysis of variance (ANOVA) paradigm, and together they test the extent of the moderating influence of the hypothesized moderator.

TABLE 3. *Meta-analyses of the validity of integrity tests for predicting substance (alcohol and drug) abuse*.

Analyses categories	Total N	K	r_{mean}	SD_r	s_{res}	$	SD$	% Var. S.E.	% Var. Total	90% CV
All samples	25,594	50	0.20	0.1175	0.0984	0.26	0.14	13.1	29.9	0.10
Employee samples	1,131	24	0.28	0.1290	0.0000	0.36	0.00	100.0	100.0	0.36
Applicant samples	22,091	10	0.17	0.0710	0.0538	0.22	0.07	08.5	42.5	0.13
Student samples	2,372	16	0.45	0.1440	0.1266	0.48	0.14	20.8	28.0	0.31

KEY: K = number of correlations; r_{mean} = mean observed correlation; SD_r = observed standard deviation; s_{res} = residual standard deviation; **$** = true validity; SD$ = true score standard deviation; % Var. S.E. = % variance due to sampling error; % Var. Total = % variance due to all corrected statistical artifacts; 90% CV = lower 90% credibility value.

RESULTS AND DISCUSSION

The results of the psychometric meta-analyses of integrity test validities for predicting overall substance abuse (alcohol and drug together) are presented in table 3.

Based on all 50 samples, the mean true validity is 0.26. Further, the 90 percent credibility value of 0.10 implies that the true validity will be greater than 0.10 in more than 90 percent of the situations. These values are based on a total sample size of 25,594. The SD of the true score validities is low (0.14), which suggests that perhaps alcohol and drug abuse can be conceptualized as manifestations of the same phenomenon of substance or chemical abuse. That is, one might hypothesize that the same personality characteristics might underlie both alcohol and drug abuse.

The separate mean true validities for student, employee, and applicant populations are also provided in table 3. In a selection setting, the focal population of interest is the applicant population. Many researchers have argued (see Ones et al. 1993 for a summary) that conscious and/or unconscious response distortion will affect integrity test validities. In taking these tests, applicants have the greatest incentive for response distortion, followed by employees and students in that order. That is, to the extent integrity test validities are affected by response distortion, true validities based on applicant samples should be lower than true validities based on employee samples, which in turn should be lower than the true validities computed on student samples.

The results reported in table 3 confirm this expected gradient. Although response distortion seems to attenuate the validity of integrity tests, its effects do not destroy validity. Even in the applicant population the true validity was 0.22 and the 90 percent credibility value was 0.13. Although this level of validity is moderate, these values suggest that the use of integrity tests in employment selection will translate into reduced levels of substance abuse in the workplace.

It is of interest to note that most of the sample consisted of applicants (about 90 percent). This is significant because applicants to jobs are the focus of interest. However, it would have been better if the applicant validities had been predictive in nature. The reader will recall that all validities in this meta-analysis are concurrent. The

criterion for applicants was admissions of drug and/or alcohol abuse made at the time they were applicants. Use of this same criterion measure taken later (after participants had been on the job for some time) would have given a better indication of predictive validity. Since there may be less response distortion on the admissions criterion measure in predictive studies, predictive validity estimates might be higher than the 0.22 obtained here. (The authors return to this point later.)

Next, the authors analyzed the results of integrity tests for predicting alcohol abuse alone. The results are summarized in table 4.

The overall estimated true validity across 20 samples involving 1,402 individuals is 0.45 and the 90 percent credibility value is 0.29. The corresponding values in the employee samples were 0.34 and 0.34,

TABLE 4. *Meta-analyses of the validity of integrity tests for predicting alcohol abuse*.

Analyses categories	Total N	K	r_{mean}	SD_r	s_{res}	$	SD$	% Var. S.E.	% Var. Total	90% CV
All samples	1,402	20	0.35	0.1638	0.0966	0.45	0.14	41.2	63.0	0.29
Employee samples	644	16	0.27	0.1128	0	0.34	0	100.0	100.0	0.34
Applicant samples	320	1	0.62	--	--	--	--	--	--	--
Student samples	438	3	0.29	0.0125	0	0.31	0	100.0	100.0	0.31

NOTE: K = number of correlations; r_{mean} = mean observed correlation; SD_r = observed standard deviation; s_{res} = residual standard deviation; **$** = true validity; SD$ = true score standard deviation; % Var. S.E. = % variance due to sampling error; % Var. Total = % variance due to all corrected statistical artifacts; 90% CV = lower 90% credibility value.

respectively. All the observed variation in validities computed on employee samples was attributable to statistical and measurement artifacts. In the student samples, the mean true validity is 0.31 and the 90 percent credibility value is 0.31 (again, all the observed variation was
explained by variations in statistical artifacts across the samples). There was only one study that used an applicant sample; in that study the observed validity coefficient was 0.62. Studies using employee

samples and studies using student samples had similar levels of validity, implying that response distortion is not a serious problem in employee samples for the criterion of alcohol abuse. However, the key question is the extent to which there is response distortion among applicants; the data here are too thin to really answer this question.

The results for the integrity test validities for the criterion of drug abuse alone are summarized in table 5.

Across student, employee, and applicant populations there were 30 studies based on 24,192 individuals. Across these 30 studies, the overall true validity was 0.25 and the 90 percent credibility value was 0.10. The true validity was highest in student samples and lowest in applicant

TABLE 5. *Meta-analyses of the validity of integrity tests for predicting drug abuse.*

Analyses categories	Total N	K	r_{mean}	SD_r	s_{res}	$	SD$	% Var. S.E.	% Var. Total	90% CV
All samples	24,192	30	0.19	0.1075	0.0909	0.25	0.13	10.0	28.4	0.10
Employee samples	487	8	0.30	0.1468	0.0561	0.38	0.08	64.5	85.4	0.29
Applicant samples	21,771	9	0.16	0.0456	0.0097	0.21	0	18.9	95.5	0.29
Student samples	1,934	13	0.48	0.1444	0.1280	0.51	0.15	19.3	21.5	0.34

KEY: K = number of correlations; r_{mean} = mean observed correlation; SD_r = observed standard deviation; s_{res} = residual standard deviation; $ = true validity; SD$ = true score standard deviation; % Var. S.E. = % variance due to sampling error; % Var. Total = % variance due to all corrected statistical artifacts; 90% CV = lower 90% credibility value.

samples, indicating that response distortion may affect the operational validities of integrity tests for predicting the criterion of drug abuse. However, the same caveats apply here as in the case of alcohol abuse (table 4). Specifically, with admissions as the criterion measure, concurrent studies done on applicants may underestimate predictive validity computed on applicants. Concurrent studies done on applicants using admissions may strongly lend themselves to response distortion on the criterion measure, which in turn would bias validity estimates downward. Applicants for jobs have strong incentive to minimize

admissions of previous illegal drug use. Present employees already have jobs, and in addition are usually told their responses will be used for research purposes only. So present employees have much less incentive for response distortion on the criterion. In contrast, response distortion on the predictor (test) does not bias estimates of operational predictive validity, because it reflects the reality that will hold when the test is used in hiring applicants. That is, real applicants will display some response distortion.

Given this likely downward bias in the mean true validity derived from concurrent studies done on applicants, the actual operational validity of integrity tests for predicting drug abuse is probably somewhere between the value of 0.21 and the value of 0.38 obtained from concurrent studies of incumbent employees. For prediction of alcohol abuse, the value corresponding to this 0.38 is 0.34. (No meta-analytic estimate of the value for applicant concurrent validity was possible for the criterion of alcohol abuse.) Hence, the operational validity of integrity tests for predicting the two types of substance abuse may be very similar. The authors would speculate that in both cases operational validity is around 0.30, a value large enough to produce practically significant reductions in substance abuse on the job if integrity tests are used in hiring.

Some limitations of the present study need to be pointed out. First, a fully hierarchical moderator analysis (Hunter and Schmidt 1990*b*) was not possible. In fact, even the main effects of some moderators could not be tested. For example, the authors could not compare the results of predictive and concurrent studies because there were no predictive studies. Also, there was only one study that used a criterion measure other than admissions of drug and/or alcohol abuse. Second, the number of existing studies was small enough in certain analyses to raise concerns about the stability of the estimates. Third, the type of study most relevant to answering questions about operational validity—predictive studies conducted on applicants—was absent from this research literature.

Any meta-analysis of test validities is limited by the number and type of available validation studies with particular criterion-predictor combinations. This has implications for second-order sampling error in meta-analyses (Hunter and Schmidt 1990*b*). But even with this limitation, a meta-analytic review based on a sound theoretical framework provides a better basis for conclusions than other approaches to understanding research findings, including the traditional narrative review. However, in this area, more research is needed. Predictive validity studies conducted on applicants would be particularly useful.

REFERENCES

Adams, J.S. Inequity in social exchange. In: Berkowitz, L., ed. *Advances in Experimental and Social Psychology*. New York: Academic Press, 1965.

Alexander, R.A.; Carson, K.P.; Alliger, G.; and Cronshaw, S.F. Empirical distributions of range restricted SD_x in validity studies. *J Appl Psychol* 74:253-258, 1989.

American Society for Personnel Administration. Most employers test new job candidates, ASPA survey shows. *Resource*, no. 6, 1988.

Ash, P. *The Construct of Employee Theft Proneness*. Rosemont, IL: SRA/London House, Inc., 1989.

Barrick, M.R.; Mount, M.K.; and Strauss, J.P. Conscientiousness and performance of sales representatives: A test of the mediating effects of goal setting. *J Appl Psychol* 79:272-280, 1994.

Bies, R.J. The predicament of injustice: The management of moral outrage. In: Cummings, L.L., and Staws, B.M., eds. *Research in Organizational Behavior*. Vol. IX. Greenwich, CT: JAI Press, 1987. pp. 289-319.

Bies, R.J., and Shapiro, D.L. Interactional fairness judgments: The influence of causal accounts. *Soc Justice Res* 1:199-218, 1987.

Blank, D.L., and Fenton, J.W. Early employment testing for marijuana: Demographic and employee retention patterns. In: Gust, S.W., and Walsy, J.M., eds. *Drugs in the Workplace: Research and Evaluation Data*. National Institute on Drug Abuse Research Monograph 91. DHHS Pub. No. (ADM)89-1612. Washington, DC: Supt. of Docs., U.S. Govt. Print. Off., 1989. pp. 139-150.

Blocklyn, P.L. Consensus: Preemployment testing. *Personnel* 2:66-68, 1988.

Bureau of National Affairs, Inc. *Recruiting and Selection Procedures, Personnel Policy Forum Survey No. 146.* Washington, DC: 1988.

Conoley, J.C., and Kramer, J.J., eds. *The Tenth Mental Measurements Yearbook*. Lincoln, NE: Buros Institute of Mental Measurements, 1989.

Cormack, R.W., and Strand, A.L. *Trustworthiness Attitude Survey*. Oakbrook, IL: Personnel Systems Corporation, 1970.

Crant, J.M., and Bateman, T.S. A model of employee responses to drug testing programs. *Employees Responsibilities and Rights J* 2:173-190, 1989.

Crouch, D.J.; Webb, D.O.; Peterson, L.V.; Butler, P.F.; and Rollins, D.E. A critical evaluation of the Utah Power and Light Company's substance abuse program: Absenteeism, accidents and costs. In: Gust, S.W., and Walsh, J.M., eds. *Drugs in the Workplace: Research and Evaluation Data*. National Institute on Drug Abuse Research Monograph 91. DHHS Pub. No. (ADM)89-1612. Washington, DC: Supt. of Docs., U.S. Govt. Print. Off., 1989. pp. 169-193.

Ghiselli, E.E. *The Validity of Occupational Aptitude Tests*. New York: Wiley, 1966.

Goldberg, L.R.; Grenier, J.R.; Guion, R.M.; Sechrest, L.B.; and Wing, H. *Questionnaires Used in the Prediction of Trustworthiness in Pre- employment Selection Decisions*. An APA task force report. Washington, DC: American Psychological Association, 1991.

Gomez-Mejia, L.R., and Balkin, D.B. Dimensions and characteristics of personnel manager perceptions of effective drug-testing programs. *Personnel Psychol* 40:745-763, 1987.

Gough, H.G. A sociological theory of psychopathy. *Am J Sociol* 53:359-366, 1948.

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

Greenberg, J. Organizational justice: Yesterday, today, and tomorrow. *J Management* 16:399-432, 1990.

Guastello, S.J., and Rieke, M.L. A review and critique of honesty test research. *Behav Sci Law* 9:501-523, 1991.

Guthrie, J.P., and Olian, J.D. "Drug and Alcohol Testing Programs: The Influence of Organizational Context and Objectives." Paper presented at the Fourth Annual Conference of the Society of Industrial/ Organizational Psychology, Boston, April 1989.

Haring, M.J.; Okun, M.A.; Stock, W.A.; Miller, W.; Kinney, C.; and Ceurvorst, W.R. "Reliability Issues in Meta-Analyses." Paper presented at the annual meeting of the American Education Research Association, Los Angeles, April 1981.

Harris, W.G. *An Investigation of the Stanton Survey Using a Validity Generalization Model*. Charlotte, NC: Stanton Corporation, undated.

Hattie, J.A., and Hansford, B.C. Self-measures and achievement: Comparing a traditional review of literature with a meta-analysis. *Aust J Educ* 26:71-75, 1982.

Hattie, J.A., and Hansford, B.C. Meta-analysis: A reflection on problems. *Aust J Educ* 36:239-254, 1984.

Herning, R.L.; Glover, B.J.; Koeppel, B.S.; and Jaffe, J.H. "Cocaine and Workplace Performance: Inferences from Clinical Studies." Paper presented at the National Institute on Drug Abuse Conference on Drugs in the Workplace: Research and Evaluation Data, Bethesda, MD, September 1989.

Hill, M.F., Jr., and Sinicropi, A.V. *Evidence in Arbitration*. Washington, DC: BNA Books, 1987.

Hogan, R. *Hogan Personality Inventory*. Minneapolis: National Computer Systems Inc., 1981.

Hunter, J.E. *Test Validation for 12,000 Jobs: An Application of Job Classification and Validity Generalization Analysis to the General Aptitude Test Battery (GATB).* Test Research Report No. 45. Washington, DC: U.S. Employment Service, U.S. Department of Labor, 1983.

Hunter, J.E., and Schmidt, F.L. Dichotomization of continuous variables: The implications for meta-analysis. *J Appl Psychol* 75:334-349, 1990*a*.

Hunter, J.E., and Schmidt, F.L. *Methods of Meta-Analysis: Correcting Error and Bias in Research Findings*. Newbury Park, CA: Sage, 1990*b*.

Hunter, J.E., and Schmidt, F.L. The estimation of sampling error variance in the meta-analysis of correlations: Use of _r in the homogenous case. *J Appl Psychol* 78:171-177, 1994.

Jackson, G.B. Methods for integrative reviews. *Rev Educ Res* 50:438-460, 1980.

Jobs, S. "Impact of Moderate Alcohol Consumption on Business Decision Making." Paper presented at the National Institute on Drug Abuse Conference on Drugs in the Workplace: Research and Evaluation Data, Bethesda, MD, September, 1989.

Johnston, L.D.; O'Malley, P.M.; and Bachman, J.G. *National Survey Results on Drug Use from the Monitoring the Future Study, 1975-1993*. Vol. II. *College Students and Young Adults*. NIH Pub. No. 94-3810. Rockville, MD: National Institute on Drug Abuse, 1994.

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

Knovsky, M.A., and Cropanzano, R. Perceived fairness of employee drug testing as a predictor of employee attitudes and job performance. *J Appl Psychol* 76:698-707, 1991.

Law, K.S.; Schmidt, F.L.; and Hunter, J.E. Nonlinearity of range restriction corrections in meta-analysis: A test of an improved procedure. *J Appl Psychol* 79:425-438, 1994.

Lent, R.H.; Aurbach, H.A.; and Levin, L.S. Predictors, criteria, and significant results. *Personnel Psychol* 24:519-533, 1971.

London House, Inc. *Personnel Selection Inventory*. Park Ridge, IL: London House Press, 1975.
London House, Inc. *The Employee Attitude Inventory*. Park Ridge, IL: London House Press, 1982.
Lousig-Nont, G.M. *Phase II Profile*. Las Vegas, NV: Self-published, 1987.
McDaniel, M.A. Does pre-employment drug use predict on-the-job suitability? *Personnel Psychol* 41:717-729, 1988.
McDaniel, M.A., and Jones, J.W. A meta-analysis of the employee attitude inventory theft scales. *J Bus Psychol* 2:327-345, 1986.
McDaniel, M.A., and Jones, J.W. Predicting employee theft: A quantitative review of the validity of a standardized measure of dishonesty. *J Bus Psychol* 2:327-345, 1988.
Miller, J.F., and Bradley, P. *Milby Profile.* Minneapolis: Milby Systems Inc., 1975.
Miller, J.D.; Cisin, I.H.; Gardner-Keaton, H.; Harrell, A.V.; Wirtz, P.W.; Abelson, H.I.; and Fishburne, P.M. *National Surveys on Drug Abuse: Main Findings*. Rockville, MD: National Institute on Drug Abuse, 1983.
Murphy, K.R.; Thornton, G.C.; and Reynolds, D.H. College students' attitudes toward employee drug testing procedures. *Personnel Psychol* 43:615-631, 1990.
Normand, J.; Salyards, S.D.; and Mahony, J.J. An evaluation of preemployment drug testing. *J Appl Psychol* 75:629-639, 1990.
Nunnally, J.C. *Psychometric Theory*. New York: McGraw Hill, 1978.
O'Bannon, R.M.; Goldinger, L.A.; and Appleby, G.S. *Honesty and Integrity Testing*. Atlanta, GA: Applied Information Resources, 1989.
Ones, D.S. "The Construct Validity of Integrity Tests." Ph.D. diss., University of Iowa, 1993.
Ones, D.S.; Viswesvaran, C.; and Schmidt, F.L. Meta-analysis of integrity test validities: Findings and implications for personnel selection and theories of job performance. *J Appl Psychol* 78:4, 1993.
Paajanen, G. *PDI Employment Inventory*. Minneapolis: Personnel Decisions, Inc., 1985.
Paajanen, G.E. "The Prediction of Counterproductive Behavior by Individual and Organizational Variables." Ph.D. diss., University of Minnesota, 1987.
Pearlman, K.; Schmidt, F.L.; and Hunter, J.E. Validity generalization results for tests used to predict job proficiency and training success in clerical occupations. *J Appl Psychol* 65:373-406, 1980.

Reid Psychological Systems. *Reid Report*. Chicago, IL: Reid Psychological Systems, 1951.

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

Sackett, P.R., and Harris, M.M. Honesty testing for personnel selection: A review and critique. *Personnel Psychol* 37:221-246, 1984.

Sackett, P.R., and Harris, M.M. Honesty testing for personnel selection: A review and critique. In: Bernardin, H.J., and Bownas, D.A., eds. *Personality Assessment in Organizations*. New York: Praeger, 1985.

Sackett, P.R.; Brusis, L.R.; and Callahan, C. Integrity testing for personnel selection: An update. *Personnel Psychol* 42:491-529, 1989.

Schmidt, F.L., and Hunter, J.E. Development of a causal model of processes determining job performance. *Curr Dir Psychol Sci* 1:89-92, 1992.

Schmidt, F.L.; Hunter, J.E.; and Urry, J.E. Statistical power in criterion-related validation studies. *J Appl Psychol* 61:473-485, 1976.

Schmidt, F.L.; Law, K.; Hunter, J.E.; Rothstein, H.R.; Pearlman, K.; and McDaniel, M. Refinements in validation generalization methods: Implications for the situational specificity hypothesis. *J Appl Psychol* 78:3-12, 1993.

Science Research Associates. *Personal Outlook Inventory*. Parkridge, IL: Science Research Associates, 1983.

Simpson, D.D.; Curtis, B.; and Butler, M.C. Description of drug users in treatment: 1971-1972 DARP admissions. *Am J Drug Alcohol Abuse* 2(1):15-28, 1975.

Stecker, M., and Rosse, J. "Attitudes Toward Random Drug Testing in the Aviation Industry." Paper presented at the annual meeting of the Society for Industrial and Organizational Psychology, Montreal, April 1992.

Stone, D.L., and Bommer, W. "Effects of Drug Testing Selection Method and Justification Provided for the Test on Reactions to Drug Testing." Paper presented at the meeting of the Academy of Management, San Francisco, August 1990.

Stone, D.L., and Kotch, D.A. Individuals' attitudes toward organizational drug testing policies and practices. *J Appl Psychol* 74:518-521, 1989.

Stone, D.L., and Vine, P.L. "Some Procedural Determinants of Attitudes Toward Drug Testing." Paper presented at the Fourth Annual Conference of the Society of Industrial/Organizational Psychology, Boston, April 1989.

Streufert, S.; Pogash, R.; Campbell, L.; Kantner, A.; Gingrich, D.; and Jacques, C.H. "Effects of Alcohol Consumption upon Performance of White Collar Tasks." Paper presented at the National Institute on Drug Abuse Conference on Drug Abuse Research and Practice: An Alliance for the 21st Century, Washington, DC, January 1991.

Thorndike, R.L. *Personnel Selection*. New York: Wiley, 1949.

U.S. Congressional Office of Technology Assessment. *The Use of Integrity Testing for Pre-employment Screening*. (OTA-SET-442). Washington, DC: Supt. of Docs., U.S. Govt. Print. Off., 1990.

Viswesvaran, C. "Modeling Job Performance: Is There a General Factor?" Ph.D. diss., University of Iowa, 1993.

Viswesvaran, C.; Ones, D.S.; and Schmidt, F.L. Appropriateness of self-reported criteria for integrity test validation: Evidence from beyond. In: Schmidt, F.L., ed. *Criterion-related Validity, Construct Validity, and Legality of Integrity Tests.* Proceedings of the symposium conducted at the seventh annual meeting of the Society of Industrial and Organizational Psychologists, Montreal, Canada, May 1992.

Whetzel, D.L., and McDaniel, M.A. Reliability of validity generalization data bases. *Psychol Rep* 63:131-134, 1988.

Winkler, H., and Sheridan, J. "An Examination of Behavior Related to Drug Use in Georgia Power Company." Paper presented at the National Institute on Drug Abuse Conference on Drugs in the Workplace: Research and Evaluation Data, Bethesda, MD, September 1989.

Woolley, R. "An Examination of the Construct and Criterion-related Validity of Overt and Personality Oriented Predictors of Counter-productivity." Masters thesis, University of British Columbia, Vancouver, 1991.

Yesavage, J.A.; Leirer, V.O.; Denari, M.; and Holister, L.E. Carry-over effects of marijuana intoxication on aircraft pilot performance: A preliminary report. *Am J Psychiatry* 142:1325-1329, 1985.

Zuckerman, M., ed. *Biological Basis of Sensation-Seeking, Impulsivity, and Anxiety*. Hillsdale, NJ: Erlbaum Press, 1983.

AUTHORS

Frank L. Schmidt, Ph.D.
Professor

Vish Viswesvaran, Ph.D.
Professor

Deniz S. Ones, Ph.D.
Professor

College of Business Administration
University of Iowa
Iowa City, IA 52242

Click here to go to page 96