The Use of Integrity Tests for Pre-Employment Screening

September 1990

OTA-SET-442
NTIS order #PB91-107011



The use of integrity tests for pre-employment screening



CONGRESS OF THE UNITED STATES
OFFICE OF TECHNOLOGY ASSESSMENT

Recommended Citation:

U.S. Congress, Office of Technology Assessment, *The Use Of Integrity Tests for Pre-Employment Screening*, OTA-SET-442 (Washington, DC: U.S. Government Printing Office, September 1990).

For sale by the Superintendent of Documents
U.S. Government Printing Office, Washington, DC 20402-9325
(order form can be found in the back of this report)

USE OF INTEGRITY TESTS FOR PRE-EMPLOYMENT SCREENING

OTA PROJECT STAFF

John Andelin, *Assistant Director, OTA*
Science, Information, and Natural Resources Division

Nancy Carson, *Program Manager*
Science, Education, and Transportation

Michael J. Feuer, *Senior Analyst*

Denise Dougherty, *Senior Analyst*

Priscilla Reagan, *Senior Analyst**

Marsha Fenn, Office *Administrator*

Gay Jackson, *Administrative Secretary*

CONTRACTORS

Anand Desai, Ph.D.
Assistant Professor of Qualitative Methods
School of Public Policy and Management
Ohio State University

Judy Iwens Eidelson, Ph.D.
Clinical Psychologist
Clinical Supervisor
Department of Psychology
University of Pennsylvania

Robert M. Guion, Ph.D.
Consulting Research Psychologist
Bowling Green, Ohio

Mark G. Keiman, J.D.
Professor of Law
Stanford University Law School

James L. Outtz, Ph.D.
Industrial Psychologist
Washington, DC

* **Through August 1989.**

Paper-and-Pencil Integrity Tests Workshop, June 12, 1989

Ms. Patricia Ambrose
Hogan & Hartson
Washington, DC

Mr. James Bassett
J.W. Bassett Company
Cincinnati, OH

Mr. Roger Borgeson
Alexander Inc.
New York, NY

Dr. Wayne Camara
American Psychological Association
Washington, DC

Robert Fitzpatrick, Esq.
Fitzpatrick, Verstegan and Cashdan
Washington, DC

Ms. Linda Goldinger
Human Resources Consultant
Atlanta, GA

Dr. Robert A. Gordon
The Johns Hopkins University
Baltimore, MD

Dr. John Jones
London House
Parkridge, IL

Dr. Joseph D. Matarazzo
Oregon Health Sciences University
Portland, OR

Ms. Marilyn Mulhall
Advance Security, Inc.
Atlanta, GA

Dr. George Paajanen
Personnel Decisions, Inc.
Minneapolis, MN

Mr. Christopher Pyle
Mount Holyoke College
South Hadley, MA

Dr. Paul Sackett
University of Minnesota
Minneapolis, MN

Dr. Frank Schmidt
University of Iowa
Iowa City, Iowa

Mr. Mike Tiner
United Food & Commercial Workers' Union
Washington, DC

NOTE: OTA appreciates and is grateful for the valuable assistance and critiques provided by the workshop participants. OTA assumes full responsibility for the report and the accuracy of its contents. Those who assisted OTA do not necessarily approve, disapprove, or endorse this report.

Other Reviewers and Contributor

Dr. Gerald L. Borofsky
Massachusetts General Hospital
Boston, MA

Dr. Michael R. Cunningham
University of Louisville
Louisville, KY

Dr. Robert M. Gordon
Institute for Advanced
Psychological Training
Allentown, PA

Dr. William Harris
Stanton Corporation
Charlotte, NC

Dr. Robin Inwald
Kew Gardens, NY

Dr. Richard Reilly
Stevens Institute of Technology
Hoboken, NJ

Dr. Michael J. Saks
University of lowa
lowa City, 1A

Dr. Leonard Saxe
Brandeis University
Waltham, MA

Dr. Neil Schmitt
Michigan State University

Dr. Andrea Solarz*
Detailee to OTA from Carnegie
Corporation of New York
New York, NY

Mr. Jack Strayer
National Association of
Convenience Stores
Alexandria, VA

Mr. William Terris
Association of Personnel
Test Publishers
Washington, DC

Dr. Christopher Webster
Clark Institute of Psychology
Toronto, Ontario

Ms. Alexandra Wigdor
National Research Council
National Academy of Science
Washington, DC

*Through August 1990.

NOTE: OTA appreciates and is grateful for the valuable assistance and critiques provided by the other reviewers and contributors. OTA assumes full responsibility for the report and the accuracy of its contents. Those who assisted OTA do not necessarily approve, disapprove, or endorse this report.

Contents

SUMMARY 1
What are Integrity Tests 1
Why Do Businesses Use Integrity Tests? 3
Why Is Use of Integrity Tests Controversial? 4
Effectiveness 4
Consequences 5
The Scope of This Repott 6
OTA Methodology 7
Meanings of Validity 8
Findings 8
Credibility 8
Quality of Research 9
Test Fallibility 10
Potential Discriminatory Effects 14
Privacy Issues 15
Policy Directions -- A Framework of Questions 15

Chapter I- INTEGRITY TESTING FOR PRE-EMPLOYMENT SCREENING AND SELECTION: BACKGROUND 19
Screening for Productive and Honest Employees 19
Measuring Theft and Counterproductivity 20
WorkplaceTheft and Counterproductivity: Explanations and Remedies 25
Legal Incentives for Pre-employment Screening 30
What are Integrity Tests? 31
Overt Questions 31
Veiled-Purpose Questions 32
Traits, Attitudes, and Behavior: Some Basic Concepts 33
Situations and Behavior 35
How are Integrity Tests Different From Other Personality Tests? 36
Caveats to Comparisons of integrity Tests and Personality Tests: Additional Considerations 38
Issues of Test Scoring and Use 40

Chapter 2- INTEGRITY TEST VALIDITY: CONCEPTS AND EVIDENCE 4
Test Validity: General Issues 44
Empirical Validation of Integrity Tests 47
General Remarks 47
Method of OTA's Review 49
Predictive Validation Using External Criteria 52
General Remarks 61

Chapter 3 - POLICY ISSUES CONCERNING THE USE OF INTEGRITY TESTS 63
Test Fallibility: Issues for Consideration 64
Self Control and the Presumption of Innocence 64
What Happens to the Misclassified? 65
What Happens If Pre-Employment Integrity Tests are Not Used? 67
Potential Discriminatory Effects: Issues and Evidence 68
Privacy 70
Confidentiality 74
Informed Consent 76
Further Thoughts 77

Box

1. The Power of Test Results 42

Tables

1. Percentage of Employees Involved in Property Deviance: Retail Sector 22
2. Percentage of Employees Involved in Property Deviance: Hospital Sector 23
3. Percentage of Employees Involved in Property Deviance: Manufacturing Sector 24
4. Percentage of Employees Involved in Production Deviance: Retail Sector 26
5. Percentage of Employees Involved in Production Deviance: Hospital Sector 27
6. Percentage of Employees Involved in Production Deviance: Manufacturing Sector 28
7. Predictive Validity Studies of Overt Integrity Tests Using Detected Theft or Close Proxy as Criterion 55
8. Forecasting Efficiency of Integrity Tests 56
9. Classification and Misclassification in Five Predictive Validity Studies Using Detected Theft or Close Proxy as Criteria 58

SUMMARY

SUMMARY

An estimated 5,000 to 6,000 business establishments in the United States use honesty and integrity tests in the process of screening and selecting job applicants for employment. Analysts familiar with the issue believe the tests are principally used to screen applicants for nonmanagerial, less-skilled jobs, such as convenience store employees and retail clerks. OTA has defined honesty and integrity tests as written tests designed to identify individuals applying for work in such jobs who have relatively high propensities to steal money or property on the job, or who are likely to engage in behavior of a more generally "counterproductive" nature. Counterproductivity in this context often includes types of "time theft," e.g., tardiness, sick leave abuse, and absenteeism.

This definition does not necessarily resolve ambiguities over the universe of tests that should be considered integrity tests. Controversy surrounds the meanings of integrity and honesty in the workplace; there is disagreement over whether integrity tests differ from other personnel tests in design or in the kinds of inferences they support; and there is little relative information on how integrity and honesty tests are actually used in hiring decisions. The debate is made more difficult because some tests that appear on their face to be at least partially relevant to measuring integrity are not considered by their publishers to be integrity tests, and because the tests are evolving in content and scope.

WHAT ARE INTEGRITY TESTS?

Integrity tests are almost all paper-and-pencil instruments, administered to job applicants at some stage of the screening and selection process. Some tests, which are called "overt integrity tests," are clearly designed to query applicants about their attitudes toward specific manifestations of dishonesty -- theft in particular -- and about their past involvement in such behavior.[2] Examples of

1. As one group of researchers has noted: "This is a rapidly changing business." P. Sackett, L. Burris, and C. Callahan, "Integrity Testing for Personnel Selection: An Update, " Personnel Psychology, vol. 42, 1989, pp. 491-528.

2. According to Sackett et al. (ibid.), these tests include the Personnel Selection Inventory (London House), the Trustworthiness Attitude Survey (Psychology Surveys Corp.), Pre-employment

overt test questions include "how honest are you?," "how prompt are you?," and "do you think it is stealing to take small items home from work?"

'Personality-based measures" or "veiled purpose tests" may not contain obvious references to theft or other specific counterproductive behaviors, but are purported to be based on meaningful underlying constructs and to yield results that are meaningful to psychologists and psychometricians.[3] Examples of these questions are "how often do you blush?", "do You make Your bed?", and "how often are you embarrassed?"[4] True-false questions include "you are more sensible than adventurous," "you work hard and steady at whatever you undertake," "you love to take chances," and "you would never talk back to a boss or a teacher."[5]

It is important to note that publishers gauge the effectiveness of both types of tests in terms of similar outcome criteria: reduction of workplace theft and/or reductions in other counterproductive behaviors. Publishers of integrity tests (and many employers) increasingly argue that honesty and integrity in the workplace should be defined broadly, to include various types of counterproductive behavior as well as outright theft of money, property, or merchandise.

Moreover, some items on integrity tests, and the constructs they purport to measure, bear some similarity to items and constructs found in other psychological personality tests that are not typically considered integrity tests by their publishers or by independent reviewers. There is disagreement in the field regarding the criteria by which to distinguish honesty and integrity tests from the broader family of personality tests.

Opinion Survey (P.O.S. Corp.), the Reid Report (Reid Psychological Systems), the Stanton Survey (Stanton Corp.), TrueTest (Intergram, Inc.), and the Phase II Profile.

3. Sackett et al. (ibid.) include in the category of "personality-based" tests the Employment Productivity Index (London House), the Hogan Personnel Selection Series (National Computer Systems), the PDI Employment Inventory (Personnel Decisions, Inc.), and the Personnel Reaction Blank (Consulting Psychologists Press).

4. These questions are taken from the Administrator's Guide to a leading integrity test (name withheld for confidentiality).

5. These items are cited as examples by Sackett et al., op. cit., footnote 1, p. 493.

WHY DO BUSINESSES USE INTEGRITY TESTS?

Integrity tests are used for several reasons. First, test publishers, some employers, and some researchers believe that the use of integrity tests can stem employee theft and counterproductive behavior. According to some estimates, losses from such actions may be quite high in some business settings. It is very difficult to estimate employee dishonesty accurately, in part because of the lack of agreement on what dishonesty means: some definitions are limited to stealing money and/or property, while others include various other forms of "workplace deviance,"[M6] especially lateness, abuse of sick leave, participating in strikes, and absenteeism (which are referred to as "time theft"). One industry-based estimate of annual losses to U.S. businesses from 11 nonviolent crimes, including but not limited to employee theft, vandalism, and bribery, was $40 billion per year.[7]

Second, there has recently been increased concern over so-called "negligent hiring" lawsuits, in which plaintiffs seek damages for losses attributed to employers' hiring of dangerous or incompetent employees. While integrity test publishers do not necessarily claim that their instruments can detect potentially violent or hazardous behaviors, they do suggest that firms can point to the use of integrity tests as evidence of a broad strategy of conscientious pre-employment screening.[8]

Third, if machine-scorable paper-and-pencil tests are accurate and reliable, they can be cost-effective tools for employee screening.[9]

---

6. Richard Hollinger, Dishonesty in the Workplace: A Manager's Guide to Preventing Employee Theft (Park Ridge, IL: London House Press, 1989).

7. See ibid., and Richard C. Hollinger, Crimes Against Business: Background, Findings, and Recommendations (New York, NY: American Management Association, 1977). See also National Institute of Law Enforcement and Criminal Justice, Law Enforcement Assistance Administration, "Summary Overview of the State of the Art Regarding Information Gathering Techniques and Level of Knowledge in Three Areas Concerning Crimes Against Business," draft report, March 1977.

8. However, there is no published court case to date in which integrity testing was entered as a defense against a negligent hiring claim; and the fact that the tests are not intended as predictors of violent or hazardous behavior makes their usefulness in negligent hiring cases questionable.

9. OTA is not aware of research that formally addresses the cost-effectiveness of integrity tests. Such research could shed light on important issues, such as comparisons of the direct and indirect costs of integrity tests and other pre-employment selection tools necessary to achieve a defined level of theft reduction or decrease in counterproductivity. Similarly, research comparing the costs of integrity testing to nontest-based methods, including interviews, credit checks, and ex post detection of counterproductive behavior, could be informative. For general discussion of cost-effectiveness models (although not with specific reference to personnel testing) see Henry Levin, Cost Effectiveness: A Primer (Beverly Hills, CA: Sage Publications, 1983).

Fourth, a boost to the development and marketing of integrity tests came from the 1988 Federal ban on polygraph testing in most private establishments. It is widely believed that this prohibition led to a renewed interest in the use of paper-and-pencil instruments, which existed as early as the 1920s (but were seldom used until several decades later) .[10]

WHY IS USE OF INTEGRITY TESTS CONTROVERSIAL?

Honesty and integrity tests are controversial: concerns have been raised about both their effectiveness and the consequences of their use.

Effectiveness

There is a strong incentive for businesses to use pre-employment screening and selection tools that have been demonstrated to reduce the proportion of new employees who are likely to commit theft or other acts of counterproductivity at the workplace. Were integrity tests established as effective, they could be beneficial to many businesses (assuming they could be shown to achieve the stated objectives at lower cost than alternative methods).

Test publishers and some employers and researchers argue that integrity tests are effective, i.e., that they can be useful in reducing the proportion of new employees who are likely to commit theft or counterproductivity. Others argue that they work poorly, if at all. While most researchers agree that the individual studies conducted to date could be much improved, there is disagreement over the implications of the existing body of research taken as a whole. The debate is fueled further by critics who challenge the underlying concept that integrity tests are purported to measure, and who are therefore not convinced by findings of validity studies based on those constructs. '

10. Various surveys conducted between 1982 and 1989 provide only sketchy data on trends in the use of integrity tests by firms. The number of available tests, however, does seem to have grown. See R. Michael O'Bannon, Linda A. Goldinger, and Gavin S. Appleby, Honesty and Integrity Testing: A Practical Guide (Atlanta, GA: Applied Information Resources, 1989), pp. 2-8. On the growth of integrity test use in the wake of the ban on polygraph testing, see also Sackett et al, op. cit., footnote 1, pp. 491,492,496-498.

11. One social psychologist argues that the real problem is that ". . . the construct actually measured [by integrity tests] is either attitudes toward theft or self-reported illicit activities [and that it requires] a substantial leap of faith to label such responses as probative of their future honesty or

For the most scholarly reviews of the evidence on effectiveness of integrity tests, readers may wish to read in full two documents frequently referenced in this text and fully cited in footnotes. They are Sackett, Burris, and Callahan, *Integrity Testing for Personnel Selection: An Update* **and O'Bannon, Goldinger, and Appleby,** *Honesty and Integnly Testing: A Practical Guide.*

Consequences

Integrity tests, like all tests, are imperfect, and can result in erroneous inferences about individual test takers. Integrity test publishers argue that error of some kind is always a problem with imperfect selection procedures, and that compared to other screening and selection devices (such as interviews or credit checks) their tests result in relatively fewer errors. Critics, on the other hand, point to the lack of sufficient research data upon which to make credible comparisons of the errors resulting from the use of various hiring procedures. In addition, they argue that erroneous test inferences could result in the denial of employment to large numbers of honest persons, an outcome that could violate social and ethical mores as well as certain !egal principles. '2

A related source of controversy turns on the argument over whether dishonesty or propensity to counterproductivity are labels that carry more negative weight than the labels derived from other personality and cognitive ability tests. Integrity test publishers tend to minimize the importance of the potential social stigma resulting from the use of their instruments, on the grounds that test takers are usually not informed of their test results and that information provided to employers is kept from public disclosure. Critics worry about the effects of these labels, which can result from imperfect test instruments: if individuals learn their scores it could affect their morale and subsequent behavior; and even if scores are revealed only to employers, and not to test takers, they could influence employers' attitudes (and behavior) toward certain employees in ways that could undermine rather than enhance individual and organizational productivity.

---

dishonesty." Leonard Saxe, "The Social Significance of Lying," paper presented to the American Psychological Association, Boston, MA, August 1990.

12. A distinction can be drawn between prediction and measurement error in tests, which is largely a psychometric problem, and errors in classification and hiring of job applicants, which is a problem in the way test inferences are translated into personnel decisions. These issues are discussed in greater detail in the Findings section of this chapter as well as in subsequent chapters of this report.

The question of how integrity tests could affect members of minority groups is another source of controversy. The test publishers rely on their research to argue that the tests do not result in "adverse impact." Critics challenge both the quality of the research and the technical definition of adverse impact.

Another point of contention concerns the scoring of tests and reporting of results. Integrity testing critics are concerned that test results are usually presented in terms of simple dichotomous breakdowns such as "recommend/not acceptable," and that the tests are marketed in large part to companies lacking the psychological and statistical training necessary to interpret more sophisticated results. Although the test publishers warn against reliance on test results as the sole basis for hiring decisions, critics question whether these admonitions are followed in practice, especially in the light of publishers' marketing literature stating that their tests can reduce workplace theft and other counterproductive activity.

Finally, critics charge that tests may violate legal and ethical standards of privacy, especially because the tests often ask personal questions not obviously related to job performance, and because there are no protections against possible misuse of test data. Testing proponents argue that privacy is largely a subjective matter, and that available evidence suggests most job applicants do not mind taking integrity tests. More survey research could be useful in informing this issue. Moreover, some proponents argue that improvements in the employer's ability to reduce dishonest behavior serve the goals of business efficiency and national productivity, and thus justify potential intrusions of privacy.

Both sides can marshal quantitative and qualitative data, and there is no obvious or easy reconciliation of the opposing arguments.

THE SCOPE OF THIS REPORT

In response to a request from the House Committee on Education and Labor, OTA examined available evidence on integrity tests, with emphasis on two basic questions:

1. **Has the research on integrity tests produced data that clearly supports or dismisses the assertion that these tests can predict dishonest behavior'?**

2. **What public policy issues are raised by the use of integrity tests for pre-employment screening and selection?**

OTA METHODOLOGY

1. **OTA studied the two most current reviews of the integrity testing literature,[13] as well as reviews of individual tests published in major test review compendiums. '4**

2. **OTA reviewed copies of tests provided by leading publishers.**

3. **OTA reviewed studies (conducted by major integrity test companies) using detected theft and counterproductivity as criteria. OTA was asked not to cite any studies not published in journals.**

4. **OTA conducted interviews with a number of experts on various aspects of testing. Some of these experts are intimately familiar with integrity testing, others specialize in related testing issues.**

5. **As in any OTA Report, comments were solicited from a wide range of reviewers on various aspects of the study, and on various drafts of the document.**

13. Sackett et al., op. cit., footnote 1; and O'Bannon, et al., op. cit., footnote 10.

14. J. Mitchell (ed.), The Ninth Mental Measurements Yearbook (Lincoln, NE: The Buros Institute of Mental Measurements, University of Nebraska-Lincoln, 1985); J. Conoley and J. Kramer (eds.), The Tenth Mental Measurements Yearbook (Lincoln, NE: The Buros Institute of Mental Measurements, University of Nebraska-Lincoln, 1989); and J. Keyser, and R. Sweetland (eds.), Test Critiques (Kansas City, MO: Test Corporation of America, 1987).

MEANINGS OF VALIDITY

This Report does not contain a simple conclusion on the "validity" of integrity tests. To the general reader, validity usually has a straightfo~ard meaning: something that **is** valid works -- **it can be fully relied upon. Scientists use the term validity differently. A test may be defined by some as valid if research demonstrates that the test can predict an outcome somewhat more accurately than a random procedure. However, most scientists also consider many other factors in evaluating validity -- test structure, research design, and consequences of use. OTA determined that characterizing integrity tests as either "valid" or "invalid" is likely to be misleading to many readers given the varying definitions of this term. Chapter 2 of this Report explores in detail the many components and aspects of evaluating validity.**

FINDINGS

1. The research on integrity tests has not yet produced data that clearly supports or dismisses the assertion that these tests can predict dishonest behavior.

Credibility

Most research on integrity tests has been conducted by investigators working for integrity test publishers, and has not been replicated by independent researchers. Situations such as these, with stakeholders controlling performance and dissemination of research, necessarily raise caution flags. Some independent research projects have recently been initiated. If these projects are carefully done, the credibility problem of currently available data and analyses could be ameliorated.

Quality of Research

The two teams of scholars who have examined in depth the research studies on integrity tests are cautiously optimistic about the quality of the research. One of these teams notes that the scope and quality of validity studies has improved over the past 5 years;[15] both make clear that certain basic methodological difficulties have not been surmounted.

Other researchers take stronger positions, on both sides. One prominent personnel psychologist believes that while integrity tests are far from perfect, they are better than any available alternative for screening and selecting of honest job applicants.[16] Another expert in personnel selection and validation research reaches a fundamentally different conclusion: "The central methodological flaw in these [predictive validity] studies is the failure to establish the construct validity of the criterion measures."[17]

Integrity test publishers, too, advocate more and better research. But they believe the existing research to be adequate as a basis upon which to stake their claim for the usefulness of their products; they prominently display this in marketing literature and users' guides, and in presentations to interested parties.

OTA did not evaluate the progress of the research over time, but did identify numerous methodological difficulties. Some of these difficulties pertain to integrity test research specifically, others are more general problems in personnel research. While difficulties in conducting tightly controlled experiments in workplace *settings* have always beset industrial psychological research, these are exacerbated in the case of integrity tests by problems in defining the behaviors of interest and the criteria by which to confirm them. First, there are many definitions of theft, and not all acts of theft are equally pernicious. Second, it is difficult to detect theft, which complicates the evaluation of links between predictors (test scores) and criteria (theft). Third, studies focusing on broader definitions of counterproductive behavior, such as absenteeism, lateness, terminations, *or* supervisors' ratings of productivity, ought to be methodologically less vulnerable to definitional and

15. Sackett et al., op. cit., footnote 1.
16. Dr. Robert Guion, personal communication, August 1990.
17. James L. Outtz, "The Validity and Reliability of Integrity Tests," OTA contractor report, Nov. 27, 1989. This report contains proprietary information made available by test publishers on the condition it not be made public; the report is therefore not available.

detection problems. But there is room for substantial improvement in the design and conduct of these kinds of studies as well. External criteria such as supervisory ratings of performance and turnover data have been questionable, and the effects of specific situational variables need to be accounted for more rigorously in research designs.

Given the paucity of independent confirmation of research results, problems identified in published reviews and in OTA's review of a sample of validity studies, and unresolved problems relating to the definition and measurement of the underlying psychological constructs, OTA finds that the existing research is insufficient as a basis for supporting the assertion that these tests can reliably predict dishonest behavior in the workplace.

II. *Errors in test results, potential discriminatory impact, and potential violations of privacy raise important public policy issues pertaining to the use of integrity tests.*

Test Fallibility

Integrity tests, like all tests, are imperfect. Honest persons can "fail," i.e., they can score below some cutoff level or relatively low in a continuous ranking; and dishonest persons can "pass." Erroneous inferences from tests do not necessarily translate directly into erroneous classification and selection decisions; but it is common in the literature of testing and selection to refer to such errors as "misclassification" or "imperfect classification,"[18] especially when the tests are marketed as tools to aid in personnel decisionmaking.

Despite misgivings about the quality of the research, OTA analyzed existing studies in order to determine the potential of integrity tests for predicting honest and dishonest behavior in the workplace.

Theft Studies

Predictive validity studies using theft as a criterion (and in which all test-takers were hired) report that from less than 1 percent to 6 percent of those passing the tests (i.e., identified as honest)

---

18. See, for example, The National Commission on Testing and Public Policy, From Gatekeeper to Gateway: Transforming Testing in America (Chestnut Hill, MA: 1990).

were later found to have stolen from their employers, meaning that upwards of 94 percent of those identified by the test as honest were not subsequently detected committing theft. However, these studies also reported that from 73 percent to 97 percent of those failing the tests (i.e., identified as potentially dishonest) apparently did not steal from their employers either and were incorrectly identified by the tests. The overall misclassification rate -- defined as the number incorrectly identified as honest or dishonest as a percentage of the total sample -- was in the range from 18 to 63 percent in the studies OTA examined (see chapter 2 of this Report, especially table 9).

Counterproductivity Studies

Test publishers argue that theft in the workplace is extremely difficult to detect,[19] and that among the large proportion of apparently honest individuals -- who the studies suggest are misidentified by the tests -- there may in fact be unknown numbers of truly dishonest persons who steal from their employers.[20] Moreover, the test publishers point out that losses from various types of counterproductive behavior that do not necessarily involve overt theft of cash or property can be significant.

For these reasons integrity test publishers have expanded their research agenda to include studies using a range of more common counterproductive behaviors as criteria. The statistical results of these studies have been reported in two ways: one, in terms of correlation coefficients that serve as measures of association between integrity test scores and one or more indicators of counterproductive behavior; and two, in terms of percentages of honest and dishonest individuals who are correctly and incorrectly identified by the test.

Correlational studies[21] reported correlation coefficients in the range from 0.16 to 0.62, with all but two falling below 0.35.[22] From studies reporting correlation coefficients alone, however, it is not possible to ascertain the proportions of honest and dishonest individuals correctly and incorrectly

19. The studies OTA reviewed found that from 2 to 10 percent of employees hired were later found to commit theft. See ch. 2.

20. Estimates of the numbers of persons misidentified vary depending on the "base rate" of theft, i.e., the true prevalence of theft.

21. Predictive studies only (and not concurrent validity studies), as reported by Sackett et al., op. cit., footnote 1. OTA was provided with numerous unpublished studies using a broad range of counterproductive behaviors as criteria, but was asked not to report the results of any specific studies. Therefore, OTA used the reports provided to analyze the methodology used by test publishers to conduct such studies, and relied on Sackett's published article for specific results. See ch. 2 of this

identified by the tests. Three studies in which the necessary data were reported found that from 18 to 29 percent of counterproductive individuals (i.e., those terminated for cause) had been incorrectly identified by the test; two of these studies found that 22 percent and 29 percent of individuals not found to be counterproductive had failed the test.

Implications of Test Fallibility

As noted above, these results are based on flawed studies, and OTA believes the results to be inconclusive. One very important datum -- the overall "failure" rate of individuals taking integrity tests also varies widely according to the available research: the proportion of individuals who take the test. and fall below the cut score[23] ranges from 30 to 60 percent.[24] This result has obvious implications for an organization's human resources policy: ". . . in order to use an integrity test, an organization must be in a position to turn away a large proportion of applicants,"[25] many of whom are very likely to be honest.

This leads to the question of why misclassification of honest individuals is particularly onerous. First, honesty and integrity are highly value-laden concepts that cut to the core of basic concepts of morality. Identifying an individual as "at high risk to commit dishonest acts" almost certainly carries a greater stigma than does the classification of an individual in other terms, e.g., relatively low cognitive abilities: the latter may channel the individual toward certain kinds of jobs not requiring those specific cognitive skills, but there are virtually no jobs for which dishonesty would be either required or desired.

A second problem of classification error from integrity tests has to do with the question of whether honesty exists as a trait, and whether, if it exists, it is immutable. There is disagreement among psychologists about the extent to which honesty is an individual trait and the extent to which it

---

Report for a fuller accounting of these studies.

22. In a properly specified multiple regression model, these results would translate to a range of 3 to 38 percent of the observed variance in counterproductive behaviors predictable from the test scores.

23. Cut scores can vary considerably depending on the test under consideration, characteristics of the workplace and desires of the client, the size of the applicant pool at any particular time, and other factors. For discussion of problems related to setting of cut scores in general, see, e.g., National Commission on Testing and Public Policy, op. cit., footnote 18.

24. Sackett et al., op. cit., footnote 1, p. 522.

25. Ibid.

is situationally determined; and there is uncertainty over its persistence. Comparison with cognitive tests can be illustrative. First, the construct measured by a test of verbal or mathematics ability, for example, is ". . . much better understood because it is supported by an enormous research base which over many years has woven links between cognitive traits and the performance of interest. . . ."[26] Second, people who demonstrate low verbal or math ability on a test presumably can benefit from remedial work and increased study -- they can try to improve their skills in the domain of interest.[27] Similarly, individuals who perform poorly on honesty tests could, Presumably) seek professional counseling or somehow change their thinking. But the question is whether genuine changes in underlying character would be reflected in subsequent tests: for example, the answer to a question like "did you ever steal" would be the same despite an individual's successful transformation into an honest person.[28] On a math test, however, an individual who has mastered a skill since failing the first test would, presumably, answer the relevant questions more successfully on subsequent attempts.

A third issue concerns the likelihood of systematic misclassification. If integrity tests are reliable (in the sense that individuals' scores do not vary significantly over time), as the test publishers claim, then their use could create a population of persons who are repeatedly misclassified, and systematically denied employment without cause. Alternative methods to screen out dishonest job applicants, such as subjective interviews or letters of reference, are also imperfect instruments. They are, however, less likely to be as consistently wrong for specific individuals. Assuming even a modest error rate, widespread use of integrity tests could deny opportunity to many individuals.[29]

---

26. Alexandra Wigdor, personal communication, July 1990,

27. In this context it is useful to consider the controversy over the use of IQ tests, which turns in part on the degree to which general intelligence is assumed to be innate. See, for example, M. Snyderman and S. Rothman, The IQ Controversy: The Media and Public Policy (New Brunswick, NJ: Transaction Books, 1988); H. Gardner, Frames of Mind (New York, NY: Basic Books, 1983); and R. Sternberg, The Triarchic Mind (New York, NY: Viking, 1988).

28. In fact, if on the first test the individual lied about prior theft, then his or her repentance could conceivably lead to truthful disclosure on the second test -- and to a lower score.

29. This outcome depends on the extent to which a single test is used for classification and/or the degree of correlation among different tests. The absence of comparative studies to determine how well different tests perform is a major deficiency of the research literature. Dr. Robert Guion, personal communication, August 1990.

Finally, integrity tests carry a scientific imprimatur -- they are marketed with literature proclaiming their "experimental validation" -- therefore substantially intensifying an individual's burden of proving that misclassification has occurred. Thus, while a virtue of the tests is their attempt to reduce the prevalence of subjective biases that might contaminate other screening and selection processes, the result can be more severe for individuals who are misclassified.

One response of test publishers to concerns over misclassification of honest people is to claim that even though employers using the tests may reject large numbers of honest people, they will still benefit from a reduction in employee dishonesty. This conclusion assumes that the available data are correct. As noted above, OTA has found that available data are insufficient to ensure such claims.

Potential Discriminatory Effects

An important concern about the effects of integrity tests is whether members of various ethnic, racial, or gender groups could suffer from discrimination in hiring as a result of test results. This is particularly important with respect to protected groups in society, and much of the research that has been conducted on discrimination has focused on so-called "adverse impact" considerations. A widely used convention in determining the presence of adverse impact is the "4/5th rule," which stipulates that a hiring rate for a minority group that is less than 80 percent of the rate for the majority will be regarded as evidence of adverse impact of the hiring system.

According to the available research, integrity tests do not violate this standard, although there appear to be differences in the mean scores of various groups. However, there is debate over the appropriateness of the 4/5th criterion in making judgments of discrimination, and the courts may be shifting their stance toward more stringent statistical criteria to use in ruling out adverse impact. If that were to occur, more research would be necessary to resolve the question of discriminatory impact of integrity tests, including substantial reanalysis of existing data.

Other questions complicate this issue. First, it is not clear whether adverse impact can refer to tests, or whether there must be evidence of test scores leading to differential selection rates. If discrimination refers to selection, evidence would be required on the precise role of test scores in

employers' hiring decisions; such evidence does not yet exist in the aggregate, and there has not been a court case in which the effects of an integrity test, per se, were adjudicated.

Because of the existence of some confusion over the appropriate standards by which to assess discrimination, as noted above, it is important to point out that even if discrimination were defined as differences in test scores without necessarily being linked to selection, there would remain the question of which standards to apply in deciding whether observed differences in group performance constitute adverse impact. In the light of these uncertainties over legal interpretations, and because the available data -- which come from test publishers' studies -- are ambiguous on how members of different ethnic, racial, and gender groups perform on integrity tests, OTA concludes that additional research is required in order to inform policy deliberations concerning discrimination and adverse impact.

Privacy Issues

Integrity tests require job applicants to disclose information about themselves that is of a personal nature, that may not be related to honesty or to the jobs for which they are applying, and that they might not choose to disclose in other settings.

Privacy is a fundamental value in American society. But it is difficult to define and conceptualize. Recurring ethical issues related to privacy appear in the debate over integrity testing: boundaries between individuals and others, the responsibility of individuals and organizations in respecting privacy, and definitions of so-called "invasive" questions are difficult issues to resolve.

At present there is no apparent protection to prevent the sharing or dissemination of this information.

POLICY DIRECTIONS - A FRAMEWORK OF QUESTIONS

Policy considerations for integrity testing are complex and difficult. At present, integrity testing is an entrepreneurial activity, lacking any regulation or formal oversight. Standards issued by the American Psychological Association and the American Test Publishers Association can serve only

as a guide to practice. Employers seek both freedom to choose employee selection methods, within the bounds of employment law, and assurance that screening practices are effective and acceptable. Available information generated by scholarly reviewers can assist sophisticated readers; marketing materials and general articles in magazines and newspapers can present a confusing picture to the general consumer.

In addition, Congress is faced with a situation in which little data exist on the actual extent and nature of use of these tests. There is no agreement on the amount of loss that business absorbs each year from employee dishonesty, and no agreement on the proportion of the population likely to engage in "dishonest behavior" under various circumstances. As pointed out in this report, there is disagreement among personnel test publishers as to which of their tests are integrity tests, and it is not clear that a simple definition could be constructed to fairly identify these instruments.

The crux of the policy problem confronting Congress is to weigh:

- **the potential gains to business of an effective pre-employment screening and selection instrument, and therefore gains to society;**

- **the potential harm to individuals, to business, and to society of instruments that do not correctly identify individuals; and**

- **the disagreement within various research and stakeholder communities over the existing research data.**

These statements make clear that Congress is faced with difficult value judgments in determining whether to take any action on this issue, and if so, *what* actions to take. The words of a leading testing and measurement expert are fitting:

The point is that in evaluating test use in selection and classification, one should not focus on one value basis -- even the value perspective of the decisionmaker -- to the exclusion of all others. To do so engenders too narrow a validation inquiry and reduces our sensitivity to side effects that are likely to be seen as adverse by other value positions. . . .[30]

OTA suggests that policymakers consider at least the following questions in their deliberations on integrity tests:

1. **Are the potentially harmful effects of the use of integrity tests justified by evidence of sufficiently high net gains in business efficiency and productivity growth?**

2. **If tests are to be used, are standards of evidence needed to approve or certify specific tests? Upon whom should the burden of proof for effectiveness fall?**

3. **What type of evaluation criteria and experimental conditions would be needed for research that more fully resolves the technical controversy over these tests?**

4. **Is there a role for the Federal Government in fostering incentives for independent research? Is there a Federal role in securing greater access to existing test industry data, either for independent researchers or for a regulatory body?**

5. **What are the rights and obligations of test publishers, employers, and test-takers regarding information generated by these tests? How secure should individual test scores be? Do these tests require full disclosure of intent to test-takers?**

6. **Do the privacy questions raised by these tests justify any particular examination by Congress?**

---

30. **S. Messick, "Validity," Educational Measurement, R. Linn (cd.), 3rd ed. (New York, NY: Macmillan, 1989), p. 87. Messick's essay addresses the importance of values in testing generally -- it does not focus on integrity testing.**

7. **Does Congress wish to obtain more information on actual test use and application? Would this include the role of test scores in the job selection process, or only aggregated test results?**

8. **If regulation is needed, who should regulate? Integrity tests are similar in some ways to a number of other tests now in use. Are all employment screening tests to be regulated, or only integrity tests? Can integrity tests be identified adequately to be regulated?**

9. **What kinds of evidence are needed for Congress (or the courts) to be assured that there is no adverse impact stemming from the use of integrity tests? Need the research providing these data be conducted by other than integrity test publishers?**

Chapter 1

INTEGRITY TESTING FOR PRE-EMPLOYMENT SCREENING AND SELECTION: BACKGROUND

Chapter 1

INTEGRITY TESTING FOR PRE-EMPLOYMENT SCREENING AND SELECTION: BACKGROUND

SCREENING FOR PRODUCTIVE AND HONEST EMPLOYEES

Hiring new workers is always risky. Applicants who are selected may turn out to be less productive than expected, while those rejected might have proven productive if given the chance. Although the costs to employers of the first type of error are more readily observable, both types can undercut the productive efficiency of a firm. Firms have an incentive to minimize the costs caused by hiring unproductive workers as well as the costs of denying employment to workers who would have become productive.

Since the early 20th century, a number of psychological tests have been developed to assist employers in making personnel decisions. For example, following the development of intelligence tests at the turn of the century, and their application by the military to recruit and assign soldiers during both World Wars, the use of personality and cognitive ability tests in industry became widespread. [1] Pressures on organizations to select and place employees more carefully have increased with specialization of job categories, high rates of employee turnover, concerns about worker productivity, workplace theft, increased liability and insurance costs, and drug and alcohol use on the job;[2] and the impetus to use more effective screening techniques has grown with the

1. For the history of personality testing in industry, see Robert Hogan, Bruce N. Carpenter, Stephen R. Briggs, and Robert O. Hansson, "Personality Assessment and Personnel Selection," Personality Assessment in Organization, John Bernardin and David A. Bownas (eds.) (New York, NY: Praeger, 1985); and Robert Perloff, James A. Craft, and Evelyn Perloff, "Testing and Industrial Application," Handbook of Psychological Assessment, Gerald Goldstein and Michel Hersen (eds.) (New York, NY: Pergamon Press, 1984). For a modern critique of assessment based on traditional tests of intelligence, see Robert Sternberg, The Triarchic Mind (New York, NY: Viking Penguin, 1988). A thorough analysis of the use of the General Aptitude Test Battery in employment can be found in John Hartigan and Alexandra Wigdor (eds.), Fairness in Employment Testing (Washington, DC: National Academy Press, 1989).

2. Anne Anastasi, "The Use of Personal Assessment in Industry: Methodological and Interpretive Problems," Personality Assessment in Organizations, H. John Bernardin and David A. Bownas (eds.) (New York, NY: Praeger, 1985), p. 2.

development and marketing of new written, physiological and chemical tests designed for use in personnel screening.[3]

Measuring Theft and Counterproductivity

With a growing awareness of the prevalence of workplace theft and counterproductivity, many employers are interested in prospective employees' honesty, indebtedness, prior convictions, drug and alcohol use, health, and dependability, in addition to their prior education and specific job skills. These hiring concerns were always a high priority for employers, and have spurred the search for innovative and effective ways to deal with employee dishonesty; and the possibility that reducing theft and counterproductivity could play a role in restoring the Nation's aggregate economic performance has gained credibility.[4]

It is important to distinguish attempts to measure the prevalence and incidence of theft from attempts to explain its origins and/or cures.[5] Neither issue is easily answered. The measurement problem is plagued by the fact that ". . . try as they might, businesses do not have any reliable statistics on the amount of employee theft and other forms of workplace crime [and] we are forced to make educated guesses regarding the scope of the problem."[6] Nevertheless, some research efforts are often cited. The American Management Association (AMA), in a frequently cited study conducted in 1977 at the request of the U.S. Department of Justice, estimated annual losses to U.S. businesses

3. For a review and discussion of biological tests designed to screen job applicants with relatively high vulnerabilities to various workplace health hazards, see Dorothy Nelkin and Laurence Tancredi, Dangerous Diagnostics: The Social Power of Biological Information (New York, NY: Basic Books, Inc., 1989).

4. The U.S. Chamber of Commerce, for example, estimates that ". . . approximately 30 percent of the business failures that occur each year are precipitated or related in some way to employee dishonesty." Chamber of Commerce of the United States, A Handbook of White Collar Crime (Washington, DC: 1974). OTA did not independently analyze these or other data regarding theft and counterproductivity levels in U.S. firms, and cannot assess the claim that the U.S. productivity growth slowdown might be arrested or reversed through the implementation of strategies designed to reduce theft in the workplace. For a recent analysis of productivity growth and comparison with other countries see, for example, William J. Baumol, Sue Anne Batey Blackman, and Edward N. Wolff, Productivity and American Leadership: The Long View (Cambridge, MA: The MIT Press, 1989).

5. Prevalence refers to the proportion of employees involved in theft or other behaviors, while incidence refers to the number of times theft occurs in a given time period. These are the definitions adopted by Hollinger. See Richard C. Hollinger, Dishonesty in the Workplace: A Manager's Guide to Preventing Employee Theft (Park Ridge, IL: London House Press, 1989), p. 13.

6. Ibid., p. 6.

from 11 nonviolent crimes (including employee theft, vandalism, and bribery) in the range of $40 billion.[7] Of the nonviolent crimes studied, AMA estimated that employee pilferage accounted for between $5 and $10 billion.[8] These estimates were used by the Bureau of National Affairs in a 1988 study showing a dramatic increase in losses from theft over the 1975 data: ". . . annual economic losses to U.S. business from employee theft ranges from $15 billion to $25 billion per year."[9]

A comprehensive analysis of workplace theft was funded by the National Institute of Justice.[10] Based on a survey of over 9,000 employees in the retail, hospital, and manufacturing sectors, including 47 corporations in three metropolitan areas (Minneapolis-St. Paul, Cleveland, and Dallas-Ft. Worth), this study found that 35 percent of employees responding in the retail sector reported some involvement in some type of theft (see table 1), 33 percent in the hospital sector (table 2), and 28 percent in the manufacturing sector (table 3). " Reported figures for involvement in production

7. See National Institute of Law Enforcement and Criminal Justice, Law Enforcement Assistance Administration, "Summary Overview of the State of the Art Regarding Information Gathering Techniques and Level of Knowledge in Three Areas Concerning Crimes Against Business," draft report, March 1977; and American Management Association, Crimes Against Business: Background, Findings, and Recommendations (New York, NY: 1977).

8. This figure could be an underestimate, however, if dishonest employee behavior contributes to other crimes such as kickbacks/bribery ($3.5 to $10 billion), embezzlement ($4 billion), and securities theft/fraud ($5 billion); but it could also be too large, if some so-called employee theft is the result of clerical or billing errors, shoplifting, or incorrect inventory control.

9. I. Shepard and R. Duston, Thieves at Work: An Employer's Guide to Combating Workplace Dishonesty (Washington, DC: Bureau of National Affairs, 1988). Cited in Hollinger, op. cit., footnote 5, p. 7.

10. Richard C. Hollinger and John P. Clark, Theft by Employees (Lexington, MA: Lexington Books/D.C. Heath and Co., 1983). Originally published as Richard C. Hollinger and John P. Clark, Theft by Employees in Work Organizations, (Washington, DC: U.S. Department of Justice, National Institute of Justice, September 1983). An update with commentary has been authored by Hollinger (workbook author) and John Jones (series editor) as Dishonesty in the Workplace: A Manager's Guide to Preventing Employee Theft (Park Ridge, IL: London House Press, 1989). For other studies of theft, see, for example Terry L. Baumer and Dennis P. Rosenbaum, Combating Retail Theft: Programs and Strategies (Boston, MA: Bitterworth Publishers, 1984), ch. 2, which concluded that losses in retail stores due to internal sources were over $8 billion for 1980, while losses attributable to customer shoplifting were about $3.7 billion. A recent study of small businesses in New York City concluded that crime cost them more than $1 billion a year, and that the most common types of crimes are break-ins, vandalism, auto and truck theft and break-ins, and shoplifting; the study, "Small Business, Big Problems," was done by Interface, a public policy research organization, as reported in Dennis Hevesi, "Crime is Costing Small Businesses $1 Billion a Year, a Study Shows," New York Times, May 15, 1989.

11. Other recent studies tend to confirm the significance of workplace theft in specific industries: for example, it was found that internal theft, and not shoplifting, was the leading cause of retail losses, and FBI statistics on bank losses point to the significant impacts of employee dishonesty. See W. Zierden, "Statement of the Chamber of Commerce of the United States on the Polygraph in the Workplace," testimony before the Senate Committee on Labor and Human Resources, June 19, 1987; and American Bankers Association, 1988 ABA Bank Insurance Survey (Washington, DC: 1988).

Table 1-- Percentage of Employees Involved in Property Deviance
Retail Sector (N = 3,567)

	Involvement				
Items	Almost daily	Once a week	4-12 times/year	1-3 times/year	Total
Misuse the discount privilege	0.6	2.4	11.0	14.9	28.9
Take store merchandise	0.2	0.5	1.3	4.6	6.6
Get paid for more hours than were worked	0.2	0.4	1.2	4.0	5.8
Purposely under-ring a purchase	0.1	0.3	1.1	1.7	3.2
Borrow or take money from employer without approval	0.1	0.1	0.5	2.0	2.7
Be reimbursed for more money than spent on business expenses	0.1	0.2	0.5	1.3	2.1
Damage merchandise to buy it on discount	---	0.1	0.2	1.0	1.3
Percentage of employees involved in one or more of the above					35.1

SOURCE: Richard D. Hollinger and John P. Clark, Theft by Employees (Lexington, MA: D.C. Heath and Company, 1983), p.42; also cited in Richard C. Hollinger, Dishonesty in the Workplace: A Manager's Guide to Preventing Employee Theft (Park Ridge, IL: London House Press, 198

Table 2- Percentage of Employees Involved in Property Deviance
Hospital Sector (N = 4,111)

	I evolvement				
Items	Almost daily	Once a week	4-12 times/year	1-3 times/year	Total
Take hospital items (e.g., linens)	0.2	0.8	8.4	1 7.9	27.3
Take or use medication intended for patients	0.1	0.3	1.9	5.5	7.8
Get paid for more hours than were worked	0.2	0.5	1.6	3.8	6.1
Take hospital equipment or tools	0.1	0.1	0.4	4.1	4.7
Be reimbursed for more money than spent on business expenses	0.1		0.2	0.8	1.1
Percentage of employees involved in one or more of the above					33.3

SOURCE: Richard D. Hollinger and John P. Clark, Theft by Employees (Lexington, MA: D.C. Heath and Company, 1983), p.42; also cited in Richard C. Hollinger, Dishonesty in the Workplace: A Manager's Guide to Preventing Employee Theft (Park Ridge, IL: London House Press, 198

Table 3- Percentage of Employees Involved in Property Deviance
Manufacturing Sector (N = 1,497)

	Involvement				
Items	Almost daily	Once a week	4-12 times/year	1-3 times/year	Total
Take raw materials used in production	0.1	0.3	3.5	10.4	14.3
Get paid for more hours than were worked	0.2	0.5	2.9	5.6	9.2
Take company tools or equipment	---	0.1	1.1	7.5	8.7
Be reimbursed for more money than spent on business expenses	0.1	0.6	1.4	5.6	7.7
Take finished products	---	---	0.4	2.7	3.1
Take precious metals (e.g., silver, platinum and gold)	0.1	0.1	0.5	1.1	1.8
Percentage of employees involved in one or more of the above					28.4

SOURCE: Richard D. Hollinger and John P. Clark, Theft by Employees (Lexington, MA: D.C. Heath and Company, 1983), p.42; also cited in Richard C. Hollinger, Dishonesty in the Workplace: A Manager's Guide to Preventing Employee Theft (Park Ridge, IL: London House Press, 198

deviance, which included taking long lunch breaks and misusing sick leave, were even higher: 64 percent in the retail sector, 69 percent in the hospital sector, and 82 percent in the manufacturing sector (see tables 4, 5, and 6).

Workplace Theft and Counterproductivity: Explanations and Remedies

As compelling as these statistics appear, they may obscure certain fundamental questions about the nature of theft and other forms of workplace deviance -- their origins and causes -- which could play an important role in devising appropriate management and public policy responses. The strategy inherent in integrity testing is to identify individuals with relatively high propensities to commit theft or other counterproductive acts. This reflects a view that some people are inherently more honest (or dishonest) than others.

However, other experts emphasize the organizational and situational influences on behavior. In addition, the question is made complicated by differences in the definition of dishonest behavior at the workplace. For example, some experts regard theft on a continuum of production and property deviance: the former includes acts such as participating in strikes, coming to work late, and abusing sick leave, and the latter refers to pilferage, embezzlement, sabotage, and stealing of property. '2

Second, there are many factors that can stimulate these acts. Some researchers argue that "**. . . most incidents of [theft] are unrelated to an employee's particular economic situation. . . ,**"[13] although there is still extensive debate on this subject. Another factor, job dissatisfaction, seems to be more important: "**. . . employees who felt that their employers were dishonest, unfair, and uncaring about their workers were significantly more involved in theft and other forms of workplace deviance.**"[14]

A very important question about workplace deviance, then, is the relative effects of individual propensities, on the one hand, and characteristics of the work environment or situation, on the other. Although this is a specific instance of the debate between "traits and states" that continues to occupy psychological researchers,[15] there appears to be widespread agreement **that it is useful to discuss**

12. Hollinger, op. cit., footnote 5, p. 34.
13. Ibid., p. 21.
14. Ibid., p. 23.
15. This issue is discussed in greater detail below.

Table 4 – Percentage of Employees Involved in Production Deviance
Retail Sector (N = 3,567)

ems	A mo da	On e a week	4 2 mes ea	3 mes ea	Tota
Take a long lunch or break without approval	6.9	13.3	15.5	20.3	56.0
Come to work late or leave early	0.9	3.4	10.8	17.2	32.3
Use sick leave when not sick	0.1	0.1	3.5	13.4	17.1
Do slow or sloppy work	0.3	1.5	4.1	9.8	15.7
Work under the influence of alcohol or drugs	0.5	0.8	1.6	4.6	7.5
Percentage of employees involved in one or more of the above					**65.4**

SOURCE: Richard D. Hollinger and John P. Clark, Theft by Employees (Lexington, MA: D.C. Heath and Co., 1983) p. 45.

Table 5 -- Percentage of Employees Involved in Production Deviance
Hospital Sector (N = 4,111)

ems	Almost da	Once a week	4-12 mes ea	1-3 mes ea	ota
			Involvement		
Take a long lunch or break without approval	8.5	13.5	17.4	17.8	57.2
Come to work late or leave early	1.0	3.5	9.6	14.9	29.0
Use sick leave when not sick	---	0.2	5.7	26.9	32.8
Do slow or sloppy work	0.2	0.8	4.1	5.9	11.0
Work under the influence of alcohol or drugs	0.1	0.3	0.6	2.2	3.2
Percentage of employees involved in one or more of the above					**69.2**

SOURCE: Richard D. Hollinger and John P. Clark, Theft by Employees (Lexington, MA: D.C. Heath and Co., 1983), p. 45.

Table 6 -- Percentage of Employees Involved in Production Deviance
Manufacturing Sector (N = 1,497)

Items	Involvement: Almost da	Once a week	4-12 mes ea	1-3 mes vea	Tota
Take a long lunch or break without approval	8.0	23.5	22.0	8.5	72.0
Come to work late or leave early	1.9	9.0	19.4	3.8	44.1
Use sick leave when not sick	---	0.2	9.6	28.6	38.4
Do slow or sloppy work	0.5	1.3	5.7	5.0	2.5
Work under the influence of alcohol or drugs	1.1	.3	3.1	7.3	2.8
Percentage of employees involved in one or more of the above					**82.2**

SOURCE: Richard D. Hollinger and John P. Clark, Theft by Employees (Lexington, MA: D.C. Heath and Co. 1983), p. 45.

theft and workplace deviance with reference to situational as well as individual variables. A good example of accounting explicitly for the effects of management decisions is found in a discussion of sick leave abuse: ". . . if management should institute strict controls over sick leave abuse, we may discover that people simply leave early or come in late or have friends 'clock them out' without their physically being at work. Or, if management tries to increase productivity without a corresponding increase in wages, we might expect to find employees compensating themselves informally through theft and pilferage."[16] Thus, while there are strong incentives to screen out job applicants with a " . . . predisposition to excusing or rationalizing theft behavior . . . ,"[17] the importance of supervisory personnel creating an atmosphere conducive to honesty and productivity seems at least as important.[18] According to this view of theft and other deviant acts, ". . . the 'crime in the workplace' perception of employee theft is usually incorrect. Employee theft is a management problem, not a crime problem."[19]

There are other sociological factors that can enrich the discussion of workplace deviance. For example, some scholars have pointed to the effects of work group norms on theft levels. One study found that ". . . the men who loaded and unloaded ships 'taxed' cargo in transit by stealing a percentage of the ship's contents . . . ," and concluded that ". . . this informal system of worker rewards is so pervasive that it constitutes a substantial 'hidden economy' found in every society around the world."[20] A special case has been documented in which management actually encouraged certain forms of employee theft: "A number of researchers have observed instances of

16. Hollinger, op. cit., footnote 5, p. 35.
17. Ibid., p. 41.
18. According to Hollinger (ibid.), pre-employment integrity testing is". . . perhaps the single most important step that an organization can take . . ." (p. 41), but the author also notes that ". . . perhaps the single most effective tool in reducing employee theft is for supervisory personnel to set a good example. . . ." (p. 42). OTA did not assess the comparative advantages of these basic approaches, but notes that if they were found to be equally effective, then a comparison of their social and private costs would be an important criterion in deciding whether to implement them. An equally important question is whether the use of tests enhances or detracts from management's efforts at creating a productive environment.
19. Ibid., p. 33.
20. Ibid., p. 24.

supervisors allowing productive employees to take from the organization as an unauthorized 'perk' rewarding high productivity."[21]

Legal Incentives for Pre-employment Screening

Employers' incentives for improved screening go beyond their desire for productive and honest workers, and may be driven also by the need to protect themselves from a variety of legal actions. For example, under "negligent hiring" doctrine, employers may be liable for the wrongful actions of their employees, even if the action occurred outside the scope of employment, if employers do not exercise reasonable care in selecting and retaining competent and safe employees.[23] While standards for reasonable care are still being developed, some employers believe that use of integrity tests might bolster their case in a negligent hiring suit, and some integrity test publishers concur with this strategy. However, whether courts will accept this defense remains unclear. To date there has not been a published negligent hiring case in which an employer's defense rested on the use of paper-and-pencil integrity tests, few integrity tests claim to predict violence, and since most negligent hiring suits involve violent behavior by employees, it is not clear that tests to screen thieves (or people who miss work or get to work late) would ever be germane.

21. Richard C. Hollinger and John P. Clark, "Deterrence in the Workplace: Perceived Severity, Perceived Certainty, and Employee Theft," Social Forces, vol. 62, 1983, pp. 398-418. This is illustrated by an example from the shoe department of a major department store: "The leading salesman of the group was allowed to freely steal from the store as a reward for making his supervisor's sales statistics look good. . . . Poor salesmen were turned in to store security personnel for relatively minor instances of pilferage." See Hollinger, op. cit., footnote 5, p. 35. For an enlightening discussion of the potentially positive effects on workplace morale and organizational productivity of certain behaviors that appear deviant, see M. Dalton, "The Interlocking of Official and Unofficial Reward," The Sociology of Organizations, O. Grusky and G. Miller (eds.) (New York, NY: The Free Press, 1981), pp. 324-346.

22. Michael Silver, "Negligent Hiring Claims Take Off," ABA Journal, vol. 73, May 1, 1987, pp. 72-78.

23. For a review of these issues see Mark Minuti, "Employer Liability Under the Doctrine of Negligent Hiring: Suggested Methods for Avoiding the Hiring of Dangerous Employees," Delaware Journal of Corporate Law, vol. 13, No. 2 , spring 1988, pp. 501-532; and Sue Walker, "Negligent Hiring: Employer's Liability for Acts of an Employee," The American Journal of Trial Advocacy, vol. 7, No. 3, summer 1984, pp. 603-610.

WHAT ARE INTEGRITY TESTS?

Integrity tests are viewed by employers as one tool in the armamentarium of personnel screening techniques, which can also include other tests of personality and/or cognitive ability, background checks into criminal history and credit records,[24] reference checks, blood or urine tests,[25] handwriting analysis, and personal interviews. These tests, almost always paper-and-pencil instruments, contain, either in whole or in part, questions about an individual's attitudes toward theft and other deviant or illegal acts, and questions about an individual's prior involvement in such behavior. Answers to these queries lead to inferences about the test-taker's propensity to commit workplace theft and/or other counterproductive acts.

Some tests, which are called "overt integrity tests," are clearly designed to query applicants about their attitudes towards specific manifestations of dishonesty -- theft in particular -- and about their past involvement in such behavior. [26]

To better understand the nature of questions that typically appear on integrity tests, consider the following examples:[27]

Overt Questions

- "How often do you tell the truth?"
- "Do you think that you are too honest to take something that is not yours?"
- "How much do you dislike doing what someone tells you to do?"

---

24. For a discussion of issues regarding access to criminal history records, see U.S. Congress, Office of Technology Assessment, An Assessment of Alternatives for a National Computerized Criminal History System, OTA-CIT-161 (Washington, DC: U.S. Government Printing Office, October 1982).

25. For discussion, see "Employment Testing," in U.S. Congress, Office of Technology Assessment, Medical Testing and Health Insurance, OTA-H-384 (Washington, DC: U.S. Government Printing Office, August 1988), ch. 3.

26. According to Sackett et al. (P. Sackett, L. Burris, and C. Callahan, "Integrity Testing for Personnel Selection: An Update," Personnel Psychology, vol. 42, 1989), these tests include the Personnel Selection Inventory (London House), the Trustworthiness Attitude Survey (Psychology Surveys Corp.), Pre-employment Opinion Survey (P.O.S. Corp.), the Reid Report (Reid Psychological Systems), the Stanton Survey (Stanton Corp.), TrueTest (Intergram, Inc.), and the Phase II Profile.

27. These questions are based on existing test questions found in a variety of integrity tests examined by OTA. The questions have been changed slightly to avoid proprietary disclosures.

- **"Do you feel guilty when you do something you should not do?"**
- **"Do you think it is stealing to take small items home from work?"**
- **"Do you believe that taking paper or pens without permission from a place where you work is stealing?"**
- **'What percentage of the people you know are so honest they wouldn't steal at all?"**
- **"How many people have cheated the government on their income tax returns?"**
- **"How easy is it to get away with stealing?"**
- **"in any of your other jobs, was it possible for a dishonest person to take merchandise if a dishonest person had your job?"**
- **"Do you believe most employers take advantage of the people who work for them?"**
- **"Do you think company bosses get away with more illegal things than their employees?"**

Veiled-Purpose Questions

- **True or False: Eating right is important to my health."**
- **"True or False: 1 like to create excitement."**
- **"True or False: I like to take chances."**
- **"On the average, How often during the week do you go to parties?"**
- **"True or False: lam usually confident about myself."**
- **"True or False: A lot of times I went against my parents' wishes."**
- **"1 feel lonely even when I am with other people {all of the time, most of the time, sometimes, almost never, never}."**
- **"How often do you blush?"**
- **"How often do you make your bed (everyday, never, etc.)?"**
- **"How many people don't you like?"**
- **"Are you an optimist?"**

Whether or not tests question applicants overtly about behavior and attitudes related to honesty, they now almost all rely on a broad range of behaviors as measures of their effectiveness. The distinction between overt and veiled-purpose integrity tests appears to be disappearing. A review

of the marketing information from publishers of the more overt tests indicates that the constructs these tests are said to measure are becoming less precise; in many cases, "theft" is broadening to include "theft and other forms of counterproductive behavior." In addition, publishers of the original "honesty" tests appear to be expanding their portfolios to include tests intended to measure a range of attitudes and predict a range of behaviors.

Traits, Attitudes, and Behavior: Some Basic Concepts[28]

The debate over integrity testing revolves around interlocking issues of test design, use, and effects. One focal point of the debate is the question of whether dishonesty is a personality trait. If a test is designed to measure the degree to which an individual possesses this trait, however, there remains the question of how the trait is linked to specific behaviors of interest. It is at least theoretically possible for individuals to be identified as possessing a trait called dishonesty without their necessarily committing theft or other counterproductive acts in the workplace. Indeed, this has led some psychologists to question the very basis of integrity tests: "It is a substantial leap of faith to label [individuals'] responses [to questions on integrity tests] as probative of their future honesty or dishonesty. . . ."[29]

It can be argued, however, that integrity tests are designed strictly to help employers weed out job applicants who are relatively likely to commit certain undesirable behaviors, including but not limited to stealing, and that the existence of definable personality traits is irrelevant. This might be called a more purely predictive model, in which test questions that work well in predicting behavior under experimental conditions are kept and those that do not contribute useful information are discarded. Pure predictive empiricists would claim that they are only mapping answers to behaviors, and not measuring any particular traits. While such tests inevitably contain at least some questions that appear to suggest personality types, they are not necessarily based on any particular

28. This section is based in part on Mark Kelman, "A General Framework for Evaluating Classification Errors, With Special Reference to Integrity Testing," OTA contractor report, June 26, 1990.

29. Leonard Saxe, "The Social Significance of Lying," paper presented to the American Psychological Association, Boston, MA, August 1990, p. 14.

psychological theory of personality.

Under either of these conceptualizations of what is being tested, an important question arises as to the relative efficacy of attitudinal and behavioral questions in predicting future behavior. On the one hand, there is empirical and theoretical support for the notion that intention is a strong predictor of behavior.[30] Assuming that individuals answer questions about their feelings regarding certain types of action with candor, and assuming further that these answers can be interpreted as reflecting intent, it may be possible to draw inferences about the likelihood of certain behaviors being committed in the future.[31]

By and large, however, prior acts are generally assumed to be better than beliefs or intentions as predictors of future acts. Test questions based on prior behavior are therefore based on a different empirical model, one which assumes that people tend more or less to keep acting the same way they have been acting. For example, persons who have stolen before are, probabilistically at least, more likely to steal in the future than those who have never stolen before (which is perhaps why detectives typically start their search for suspects by considering evidence linking a crime to known -- rather than new -- criminals). The validity of integrity tests based on these questions, then, depends in large part on whether admissions of past acts are a reasonable surrogate for actual past acts. It is difficult to assess the accuracy of self-report data in the absence of objective benchmarks.[32] Moreover, if admissions-based data are accurate, then people who confess to prior acts are reporting honestly. They might be probabilistically more likely to commit the undesirable behavior of interest in the future, however, and this makes the interpretation of such tests particularly complicated. (It is important to keep in mind that integrity tests do not usually rely on questions about prior behavior alone.)

---

30. See, for example, 1. Azjen, "Attitudes, Traits, and Actions: Dispositional Prediction of Behavior in Personality and Social Psychology, " Advances in Experimental Social Psvchology, vol. 20, 1987, pp. 1-63. This article reflects a stronger position on the role of attitude and intention from that taken in an earlier work. See 1. Azjen and M. Fishbein, "Attitude-Behavior Relations: A Theoretical Analysis and Review of Empirical Research, Psychological Bulletin, vol. 84, 1977, pp. 888-918.

31. OTA did not assess the extent to which attitude questions on integrity tests would be considered as questions of "intent," and found no research that addressed this issue specifically.

32. On the reliability of admissions data, see Sackett et al., op. cit., footnote 26, pp. 517-519. In this context it is worth noting that although some predictors of criminality might be more accurate than personality-based ones -- see, for example, Herbett Packer, The Limits of the Criminal Sanction (Stanford, CA: Stanford University Press, 1968) -- they are not necessarily admissible as criteria for selection: in other words, predictive validity is not the sole criterion for determining the uses of screening instruments. See, e.g., Kelman, op. cit., footnote 31; or Nelkin and Tancredi, op. cit.,

Purely predictive tests, as well as those aimed at identifying theoretical psychological traits, can consist of both attitudinal and behavioral questions. Attitudinal questions probe beliefs and feelings about dishonesty, counterproductivity, and/or other even seemingly unrelated attitudes. Behavioral questions seek to correlate prior acts -- overtly related to honesty -- with future ones.

Situations and Behavior

An important point regarding the predictive ability of integrity tests concerns the relative importance of individual personality variables and environments in explaining behavior. Despite efforts to declare the debate over,[33] psychologists continue to disagree on their relative importance. And this general debate has extended into the more specific area of honesty.[34] In any event, it 's unclear to what extent integrity test publishers take seriously the effects of situations on personal behavior. One spokesman for the integrity test industry claims that ". . . integrity test publishers typically assume that dishonesty is a relatively stable personality trait, but that counterproductive behavior can be influenced by a variety of situational factors. "[35] There have been no studies of integrity tests in which organizational level variables have been fully integrated. These variables are difficult to define, and integrity test publishers are not alone in encountering this methodological problem.

---

footnote 3.

33. See, for example, R. Carson, "Personality," Annual Review of Psychology, vol. 40, 1989, pp. 227-248. Carson cites W. Mischel, who is one of the psychologists most clearly identified with a situationist perspective, as having adopted over the years a more interactionist perspective. In this view, traits are taken as ". . . conditional probabilities that a particular action will be evoked by a particular environmental state." See Wright and Mischel, 1987, cited in Carson. See also W. Mischel, Personality and Measurement (New York, NY: John Wiley, 1968); and his more recent article, "Toward a Cognitive Social Learning Reconceptualization of Personality," Psychological Review, vol. 80, 1973, pp. 252-283. For different viewpoints that have appeared in the literature, see also D. Bem and D. Funder, "Predicting More of the People More of the Time: Assessing the Personality of Situations," Psychological Review, vol. 85, No. 6, November 1978, pp. 485-501; there the authors write that ". . . the recent debate [over the interaction of trait and situation] appears now to have evolved into a consensus that it is the interaction between the person and the situation that supplies most of the psychologically interesting variance in behavior . . ." (pp. 485-486).

34. The work of H. Hartshorne and M. May, Studies in the Nature of Character (New York, NY: Macmillan, 1928) was seminal. More recently P. Ekman, Why Kids Lie: How Parents Can Encourage Truthfulness (New York, NY: Charles Scribner's Sons, 1989) suggests that personality matters more than previously believed. Saxe however, notes that Ekman's critique of the Hartshorne and May work, on the grounds that they "made it too easy to cheat," serves to reinforce the situational explanation (op. cit., footnote 28, p. 9).

35. John Jones, London House, personal communication, July 1990.

HOW ARE INTEGRITY TESTS DIFFERENT FROM OTHER PERSONALITY TESTS?

While there are still some integrity tests that purport to predict theft alone, as noted above, the majority appear to be marketed as instruments designed to assess a wider range of personality traits and to predict a wider range of behaviors. Publishers of integrity tests (and many employers) now increasingly argue that honesty and integrity in the workplace should be defined broadly, to include various types of counterproductive behavior as well as outright theft of money, property, or merchandise. Moreover, some items on integrity tests, and the constructs they purport to measure, bear some similarity to items and constructs found in other psychological personality tests.

Thus, with respect to criteria (i.e., outcomes of interest) and predictors (test constructs) it is sometimes difficult to distinguish honesty and integrity tests from the broader family of personality tests: in fact, one integrity test publisher has argued that "there is no fundamental conceptual difference between integrity tests and other personality tests," such as the Sixteen Personality Factors Test (16PF), the Minnesota Multiphasic Personality Inventory (MMPI), and the California Psychological Inventory (CPI).[36] Nonetheless, there are differences among these latter tests and between any one of them and an integrity test.

It is commonly agreed that integrity tests are tests of personality, as they claim to measure an individual's propensity to behave in certain ways. But the professional and academic literature on integrity tests is ambiguous on the question of whether integrity tests are somehow special within this broader family. The leading academic and professional reviewers[37] note that most tests now include more than just honesty scales, which, at least until very recently, clearly distinguished them from other tests [38] But these reviewers also imply that integrity tests are different from other Personality tests and

36. Ibid.

37. R. Michael O'Bannon, Linda A. Goldinger, and Gavin S. Appleby, Honesty and Integrity Testing: A Practical Guide (Atlanta, GA: Applied Information Resources, 1989); and Sackett et al., op. cit., footnote 26.

38. See, for example, H. Heneman, D. Schwab, J. Fossum, and L. Dyer, Personnel/Human Resource Management (Homewood, IL: Irwin, 1989), p. 338, where honesty tests are singled out as a separate category from "work sample tests," "personality and interest tests," etc.

that they ought to be considered in a class by themselves. For example, a comprehensive directory of available integrity tests omits several widely used personality tests, despite certain similarities in question content and scope.[39] One is therefore left with the impression that experts continue to sense important differences between integrity tests and other personality tests, but that the differences are difficult to pinpoint.

This issue of deciding which tests are integrity tests and which are not seems to ignite considerable debate and acrimony. Some tests include items or scales seemingly related to honesty generally (if not in the workplace); but the publishers of these tests assert -- often quite vehemently -- that they are not integrity tests. For example, one test designed and used for screening law enforcement applicants includes the item: "1 have to admit it, I once took money from an employer," and a scale called "Trouble with the Law and Society." In validation research on this test, criteria such as turnover, absences, lateness, and disciplinary actions have been used.[40]

Nevertheless, the developer of this test does not consider it an integrity test, primarily because it has never been validated using theft as a criterion and because it is not intended for predicting theft or screening out potential thieves per se.[41] similarly, the Army's ABLE test, which contains measures designed to predict turnover, is not considered an integrity test by its developers, the claims of some integrity test publishers notwithstanding.

This point of contention has more than just academic interest. Policymakers contemplating possible regulatory action must keep in mind the formidable barriers to defining precisely what tests would be included and under which criteria.

---

39. O'Bannon et al., op. cit., footnote 38. See also Sackett et al., op. cit., footnote 26, whose review of research on validity omits studies based on several personality tests known to be used in personnel selection.

40. See B. Bolton, review of the Inwald Personality Inventory, in J. Mitchell (ed.), The Ninth Mental Measurements Yearbook (Lincoln, NE: Buros Institute of Mental Measurements, University of Nebraska-Lincoln 1985), pp. 711-714.

41. Robin Inwald, personal communication, 1990. Note also that this test was not included in the review by Sackett et al. (possibly because its use is limited to a single employment setting, namely enforcement and security), nor in the review by O'Bannon et al. (op. cit., footnote 38). It is included as an "interpersonal skills and attitudes" test under the broader category of "Business and Industry" tests in Test Critiques. J. Keyser and R. Sweetland (eds.), Test Critiques (Kansas City, MO: Test Corporation of America, 1987).

Even if honesty tests resemble personality tests because they share some common items or scales, they are somewhat distinguishable by the scope of their questions and by the nature of their intended uses. Thus ". . . personality and interest tests seek to measure motivation . . . ,"[42] and " . . . with few exceptions [these tests] have not been developed for use as employee selection techniques. Personality tests are typically intended . . . to identify broad personality dimensions or mental disorders . . . [while] interest tests are used to provide people with information about their preferences for various activities, and, in turn, such information can be of assistance in making personnel choices."[43] When personality tests are used, they can provide information on such matters as individual interests, which presumably can be helpful in assigning people to appropriate jobs. While in practice they are also sometimes used for personnel selection (i. e., for hire/no hire decisions), that use is considered controversial. Honesty tests are specifically designed and marketed for selection of applicants and not for their assignment to particular jobs.[44]

Caveats to Comparisons of Integrity Tests and Personality Tests: Additional Considerations

As stated earlier, some integrity test publishers tend to compare their products with several widely used personality tests, and claim they are identical in most important aspects. OTA believes this claim to be weak. Consider, for example, the Minnesota Multiphasic Personality Inventory (MMPI), a widely used and validated personality test, originally designed for use in identifying clinically significant levels of psychopathology. The test was validated on a clinical sample of psychiatric inpatients, and while it has been applied to "normal" populations, these applications have raised a controversy within the psychological community. The recently added subscales intended specifically for use in employment screening have not been validated independently and have been controversial as well.[45]

---

42. Anne Anastasi, Psychological Testing, 6th ed. (New York, NY: Macmillan, 1988), pp. 523.
43. Heneman et al., op. cit., footnote 39, p. 336.
44. For honesty tests to be useful in assigning individuals to jobs, one would have to assume that some jobs (or some employers) have less need for honest workers. To illustrate this point, suppose one's interest in classical music suggested a poor match for work in a video arcade; it would not necessarily rule out work in a concert hall. But one's "high risk of committing theft" would likely be undesirable in any job, which suggests why misclassification from honesty tests may be especially troubling. See ch. 3 for discussion.
45. One source of controversy is the face validity of these scales, which closely resembles

Even the user's manual for the MMPI-2 reflects the controversy over using the test for normal populations, and cautions that ". . . preemployment screening [is justified] for positions for which *clinical personality assessment is recommended, namely, positions involving public safety and trust,* and those in which personality factors affect the performance of hazardous jobs. . . ."[46]

Moreover, even in high-risk environments where the MMPI can be deemed acceptable, it is not to be used as the sole instrument of selection. Individuals applying for sensitive jobs who test negative on the MMPI are usually subjected to additional screening.[47] It is not clear to what extent business establishments rely exclusively or principally on integrity test scores.

The proprietary nature of integrity tests is another distinguishing characteristic that raises problems. Unlike the MMPI, for example, for which item banks and scoring keys (the templates used to interpret raw scores) have generally been available to independent researchers, the content and scoring algorithms of integrity tests have been more closely held.[48] critics argue that as long as integrity tests remain proprietary, it is unlikely that the research base will improve substantially, either in quantity or in credibility. It is important to keep in mind, however, that test publishers believe that the effectiveness of their instruments could be jeopardized if the contents were made public; this, they would argue, could cut revenues and constrain their resources available for research and test innovations.[49]

---

honesty tests. According to Dr. Robert M. Gordon (Institute for Advanced Psychological Training, Allentown, PA), an expert on MMPI research and practice, when people apply for jobs they are particularly defensive about their faults, and have a strong incentive to fake their answers. While truthfully reporting an occasional immoral act would result in a low (i.e., good) score on the "L" (lie) scale, it would show up on the "ASP" (antisocial practices) scale as an indicator of immorality.

46. S.R. Hathaway, J.C. McKinley, and James Butcher, MMPI-2: Minnesota Multiphasic Personality Inventory-2 -- User's Guide (Minneapolis, MN: National Computer Systems, Inc., 1989), p. 1 (italics added). This caveat is mentioned in a section entitled "Clinical Testing in Personnel Selection, An Invasion of Privacy?" and points to one of the public policy problems of concern in this report (see ch. 3). Note that the invasion of privacy issue seems to persist despite extensive revisions of the original MMPI, and deletion of particularly offensive questions.

47. Although the MMPI demonstrates extremely low rates of false positive error (less than 1 percent of persons who are diagnosed by the test as ill are in fact free of the psychopathology(ies) of interest), it continues to suffer from what might be an unacceptably high rate of false negatives. This represents an inversion of the kind of public policy problem that arises because of false positive errors of integrity tests: false negative errors with MMPI can lead to the mistaken hiring of individuals who are not psychologically suited for dangerous or sensitive jobs. (OTA is grateful to Dr. Robert M. Gordon, for bringing these matters up in a personal communication to OTA.)

48. The IPI and 16-PF are copyrighted, but are not as closely held as integrity tests.

49. The original developers of the MMPI may have made little money, but spawned a generation of high-quality psychological research; indeed, much of the research was devoted to the detection

Issues of Test Scoring and Use

Another aspect of integrity tests that may help distinguish them from other personality tests is the manner in which they are scored and their results presented to clients. Honesty and integrity tests on the market today can be scored by the test publisher or the employer. Although no statistics have been gathered on scoring procedures for the available tests, the impression is that slightly more than one-third offer both on-site and publisher scoring, and less than one-third offer only one of either option.[50]

An important issue concerning the scoring of tests and reporting of results arises in light of the fact that ". . . integrity tests are marketed in large part to nonpsychologists . . . ,"[51] who may be inadequately equipped to interpret the results. For example, most publishers who score the tests themselves provide an interpretation of test performance in terms of "recommend/not acceptable." Although individual propensities to act dishonestly or counterproductively are often classified in more than two dichotomous risk categories, the intent is that employers -- with professional guidance from test vendors -- use these classifications in making hiring decisions. While psychometrician and some test publishers recognize that continuous score distributions are superior to such classifications, the latter are easier to interpret and are therefore more compelling to employers.[52]

---

and weeding out of flaws in the test instrument. Some long-time users and advocates of the MMPI have expressed their dismay over the recent commercialization of the revised version of that test, and its direct marketing as a personnel selection device. Integrity test publishers, on the other hand, believe that innovations and improvements in their tests are furthered, not impeded, by the potential for financial reward; and that ultimately both the producers and consumers of integrity tests are better off in this arrangement, because publishers assume responsibility for the psychological interpretation of their tests (based on an interview with John W. Jones, London House, June 26, 1990). OTA did not analyze these issues in full. The American Psychological Association has recently created a task force to look into various aspects of commercialization in psychological research, and its findings will undoubtedly be useful to policymakers.

50. Linda Goldinger, personal communication, 1990.

51. Sackett et al., op. cit., footnote 26, p. 523.

52. For continuous score distributions to yield meaningful inferences about differences in risk level among applicants with different scores, additional information is required (the standard error of measurement). But this information is rarely given and even more rarely understood by nonpsychologists or psychologists without adequate training in statistics.

It is important to point out that test inaccuracy in itself is not a measure of actual misclassification of job applicants, but rather provides a measure of the potential harm that could result if test results were the dominant or sole criterion for selection. Few experts would argue with the publishers' warning that test scores should not be the sole basis for hiring decisions.[53] Whether these admonitions are followed in practice, however, is questionable, and seem to be confounded by claims in publishers' marketing literature "proving" that tests -- without reference to other elements of hiring decisions -- can reduce workplace theft and other counterproductive activity. The role of tests in reaching hiring decisions remains largely unknown. Although similar arguments could be advanced regarding any test for which discrete classes of performance, rather than continuous scores, are repotted, the categories often provided for integrity tests -- "at risk to commit theft," e.g. -- may be particularly influential in hiring decisions (see box 1).

53. See Association of Personnel Test Publishers, Model Guidelines for Preemployment Integrity Testing Programs, 1st ed. (Washington, DC: 1990).

Box 1

The Power of Test Results

One question that warrants careful empirical investigation is how employers use different kinds of test-based information. For example, some observers argue that the seductive nature of quantitative data, generated from "scientifically validated" studies, could induce employers to base their hiring decisions solely or primarily on test scores. To illustrate, consider a hypothetical test, which provides information on a range of 18 personality traits from "interpersonal style" and "caring" to "natural v. logical problem solving style." The intent of this instrument is to provide clues to rather complex psychological traits, not all of which are clearly defined or necessarily consistent with one another. While the test might supply some useful information to employers, particularly for meeting certain job needs, it is not intended as the basis for a dichotomous "hire/no hire" decision.

But now suppose there were a "19th" factor added to the list, called "dishonesty" or "proclivity to violence." Given that no employer wants dishonest or violent workers -- regardless of other cognitive or interpersonal attributes -- this variable could very well dominate the list. The scientific imprimatur associated with scoring this factor would be likely to make any employer reluctant to hire someone with this ranking. To a large extent this is why many psychologists who advocate the cautious use of any personality test in selection are concerned over the apparent ease with which integrity tests can be misused.

End of Box

Chapter 2

INTEGRITY TEST VALIDITY: CONCEPTS AND EVIDENCE

Chapter 2

INTEGRITY TEST VALIDITY: CONCEPTS AND EVIDENCE

Perhaps the first and most often asked question about integrity tests is whether they are valid. This question is not easily or intuitively answered. At its simplest, the question means "do the tests work?" -- or, "do people who do well on the tests actually tend to act 'honestly' more than those who do poorly?'

Beneath these rather obvious questions are layers of subtle problems that have challenged generations of psychologists and other social scientists: Is dishonesty a character trait? If so, is it permanent and does it manifest similarly in all workplace settings? Can written tests effectively and reliably expose the presence of the trait (if it exists) and/or an individual's propensity to commit certain behaviors of interest? Why probe attitudes, intentions, or feelings if evidence of past behavior is available and is considered a powerful predictor of future behavior?

Because the answers to these and related questions can influence decisions affecting many people, they raise a set of formidable public policy concerns (see also chapter 3). And even if one wished to concentrate on the purely empirical question -- how well do the tests do what they are purported to do? -- the research challenge is impressive. Gathering evidence to compare the behavior of individuals with different test scores, drawing statistically valid inferences (predictions) from those scores about individual propensities to act in certain ways, and establishing reasonable levels of confidence in those predictions require a mobilization of sophisticated analytical methods.

This chapter discusses these issues and reviews empirical research on the validation of integrity tests. Discussed first are general issues in validity: What is meant by validity and what are the important issues in test validation? Construct validity, content validity, predictive validity, test reliability, and the internal validity (research design) of studies designed to demonstrate test validity are described, as well as the relatively new concept of consequential validity. In the next section, studies designed to evaluate the construct and predictive validity of integrity tests are described and discussed. Particular attention is paid to issues of the quality of the research that has been conducted.

TEST VALIDITY: GENERAL ISSUES

Although intuitively appealing, the implied definition of validity in the opening sentences of this chapter is not, technically speaking, correct. For it is not a test, per se, which is valid or invalid; rather it is the set of inferences drawn from a test: "Validity is an overall evaluative judgment, founded on empirical evidence and theoretical rationales, of the adequacy and appropriateness of inferences and actions based on test scores.'" In common parlance it is customary to refer to a test's validity in either-or terms: either the test is valid or it is not. But measurement theorists now recognize that validity is a form of evaluation, of a number of issues, and that the result of the evaluative process is a sense of the relative strength or weakness of the inferences drawn from test scores. These varieties of evaluative information, which measurement scientists have attempted to group under various headings such as "content" or "construct" or "criterion-related" validity, come together in an "... argument [that] must link concepts, evidence, social and personal consequences, and values."[2] In a word, then, the best that can be said about any test is that attempts to validate it yield persuasive and acceptable inferences.[3]

Test theorists have identified several components of validation, and while". . . the 30-year old idea of three types of validity. . . is an idea whose time has gone. . . ,"[4] the ideas underlying "content," "construct," and "criterion-related" validity are still very much part of the psychometrician's arsenal.[5]

---

1. S. Messick, "The Once and Future Issues of Validity: Assessing the Meaning and Consequences of Measurement," Test Validity, H. Wainer and H. Braun, (eds.) (Hillsdale, NJ: Lawrence Erlbaum Associates, 1988), p. 33. See also American Psychological Association, Standards for Educational and Psychological Testing (Washington, DC: 1985).

2. L. Cronbach, "Five Perspectives on Validity Argument," in Wainer and Braun, op. cit., footnote 1, p. 4. Cronbach reminds test validators of the importance of what Messick calls "consequential" validity: "Tests that impinge on the rights and life chances of individuals are inherently disputable . . . the bottom line is that validators have an obligation to review whether a practice has appropriate consequences for individuals and institutions, and especially to guard against adverse consequences." See also S. Messick, "Test Validation and the Ethics of Assessment," American Psychologist, vol. 35, 1980, pp. 1012-1027.

3. "It might also be pointed out that the use of any given test may have as many validities as there are inferences to be drawn from the scores. An integrity test may or may not have much validity for inferences about how generally honest a person is, and it may or may not have much validity for inferences about future counterproductive behavior on a specific job, but these are not interchangeable." Dr. Robert Guion, personal communication, August 1990.

4. Cronbach, op. cit., footnote 2, p. 4.

The construct validity of an instrument is the extent to which one can be sure it represents the construct which it seeks to measure. "A test with good construct validity can be considered a substitute for actually observing a person displaying a skill or attitude in everyday life."[6] Content validity refers to the "representativeness" of the sample of questions on a test, i.e., the extent to which they cover the construct or constructs being measured. "High content validity means that the test 'maps onto' the collection of possible questions by sampling representatively from its various manifestations. . . ."[7]

Both of these aspects of test validity are internal criteria, i.e., they relate to the construction of the test. To determine whether a test measures what it claims to measure, it should also satisfy external criteria: for example, how well the test mimics scores on established and reputable tests that are used to measure similar constructs would be one indication of its ability to measure what it claims to measure. But that would not be sufficient. It is more important to". . . find out whether it correlates with other things implied by [what the test claims to measure] and whether it is uncorrelated with things irrelevant to that claim."[8]

When a test is intended for selection, the most compelling aspect of its validity is the extent to which test scores correlate with later behavior. "Predictive validity," therefore, occupies a central place in discussions of personnel testing in general and of integrity testing in particular. A variant on predictive validity is the so-called "concurrent validity" approach, in which predictors and behaviors are measured at the same time. "Typically, concurrent validity data were taken as evidence that a newly proposed test, or a brief version of an existing test, was measuring a given trait if it correlated strongly with another test already acknowledged to be a measure of that trait . . . concurrent validity was, and still is, held to be useful for predictive purposes if it could be demonstrated, or argued convincingly, that scores on the test would not change systematically during the period between the

---

5. Classification of the various approaches to validity has evolved. During the 1940s and 1950s predictive and concurrent validity were considered separate aspects (or types) of validity, and were later combined under the single heading of "criterion-related" validity. For discussion, see W. Angoff, "Validity: An Evolving Concept," in Wainer and Braun, op. cit., footnote 1, pp. 19-32.
6. Marlene E. Henerson, Lynn L. Morris, and Carol T. Fitz-Gibbon, How to Measure Attitudes (Beverly Hills, CA: Sage Publications, 1978), p. 135.
7. Ibid., p. 140.
8. Dr. Robert Guion, personal communication, August 1990.

time when the test *might* have been given as an actual predictor and the time when criterion data would normally become available."[9]

In addition to these aspects of validity, which pertain to the usefulness of a test as a decisionmaking aid, researchers have begun to incorporate the notion of "consequential" validity in their studies. As argued by one prominent measurement theoretician, ". . . judging whether a test does the job it is employed to do . . . requires evaluation of the intended and unintended social consequences of test interpretation and use."[10] Note, however, the link between consequential validity and other aspects of validity: if adverse consequences can be ascribed to some aspect of score distributions (such as ethnic differences), ". . . which would directly reflect on the functional worth of the selection testing, . . . [the question becomes] whether the adverse impact is attributable to construct-relevant or construct-irrelevant test variance or to criterion-related or criterion-unrelated test variance. . . ."

Another important feature of a test instrument is its so-called "reliability," which reflects ". . . the extent to which measurement results are free of unpredictable kinds of error.'"2 For instance, repeated administrations of a test to the same sample of subjects should yield similar scores. Note that while a valid measure is always reliable, the opposite is not necessarily true: reliability does not necessarily imply validity.[13] Underlying the concept of reliability is the notion of a "true score," i.e., the score that an individual would obtain on a test as a reflection of his or her propensities or abilities. However, when the test is administered, the score falls within some range around this "true" score, and measures of reliability are generally based on estimates of the variability in the observed score around the true score.[14]

---

9. Angoff, op. cit., footnote 5, p. 21.
10. Messick, in Wainer and Braun, op. cit., footnote 1, p. 39.
11. Ibid., p. 40.
12. Henerson et al, op. cit., footnote 6, p. 146.
13. A broken watch is very reliable -- it always tells the same time. But because it provides no information about the real time, it is not valid.
14. For discussion of the technical issues in measurement of reliability, see, e.g., A. Anastasi, Psychological Testing, 3rd ed. (New York, NY: Macmillan, 1968); or Leonard S. Feldt and Robert Brennan, "Reliability," Educational Measurement, R. Linn (ed.) (New York, NY: American Council on Education/Macmillan Publishing, 1989), pp. 105-146.

A related issue is the sensitivity of a test: a test should yield results that can identify differences between two individuals, but it should also not give wildly divergent scores for two fairly similar individuals. A test that is not sensitive to differences is not useful in discriminating between individuals; but an overly sensitive test can lose some of its reliability.

An important consideration in understanding all efforts at test validation is the quality of the research conducted. A valid study design contributes to the confidence that can be placed in a study's results. issues of the quality of a research design are generally known as internal validity. The level of internal validity is the extent to which the relationships detected in a study are not spurious, that is, due to factors not accounted for in the study. Among the factors that may undermine internal validity are: poor sample selection, the occurrence of events during the course of a study that affect the outcome variable in unanticipated ways, nonindependence of observations, and unintended effects on a research subject of being measured.[15] Appropriate use of statistics is another important aspect of study design.

Finally, a critical consideration in determining the quality of any research is the quality and depth of the research report. Because science is a systematic process for creating and disseminating new knowledge, research reports should provide sufficient detail to enable independent scientists to evaluate the credibility of the reported results.

EMPIRICAL VALIDATION OF INTEGRITY TESTS

General Remarks

It should be clear from the foregoing discussion that validation of any test or treatment is a complex process, requiring a balance of subjective judgment and scientific evidence. The determination of construct validity often relies heavily on the opinion of experts who must define the theoretical constructs to be measured in order to identify the presence or absence of specific human traits. Content validation involves the assessment of how well test questions correspond to these

15. See, e.g., T. Cook and D. Campbell, Quasi-Experimentation: Design and Analysis Issues for Field Settings (Chicago, IL: Rand McNally, 1979); and L. Saxe and M. Fine, Social Experiments (Beverly Hills, CA: Sage Publications, 1981).

constructs. Criterion-related validation requires the implementation of the test and the subsequent determination of how the test compares with other measures of the same constructs (to assess "concurrent validity") or how well the test predicts behaviors or actions it is supposed to forecast ("predictive validity"). Concurrent and predictive validity studies require the identification of one or more "criteria," i.e., variables that serve as indicators of the types of behavior under study. Finally, all steps in the process should reflect generally accepted principles of valid research design and should be repotted in enough depth so that the research process is clear to readers.

For integrity testing, validity is especially problematic because integrity and honesty are extremely difficult constructs to define with sufficient precision to enable empirical measurement. On the one hand, the temptation to stick to easily defined acts of dishonesty, such as theft, is stymied by the relatively low frequency of detected theft and, therefore, its limited use as an external criterion. Extending the definition, however, to encompass a wider range of behaviors can result in greater ambiguity about the value of a test as a predictor of the kinds of dishonest acts of greatest interest to employers; "wayward impulse," for example, a construct included in one popular integrity test, maybe a meaningful psychological or characterological trait indicative of a propensity toward certain behaviors, but its usefulness as a predictor of an individual's future commitment of dishonest deeds is tenuous.

Other factors affecting the feasibility or accuracy of empirically validating integrity tests include: the multiple and often unobservable determinants of trends in aggregate measures of organizational productivity, which could confound time series studies of shrinkage; incentives for respondents to answer high-stakes tests strategically, rather than with complete candor, and the possibility that over time job applicants will learn how to answer the tests even more skillfully; and potential biases in criterion measures. Even the reviewers whose analyses end on a relatively optimistic note agree that research in this field faces formidable methodological problems. '6

16. For example, R. Michael O'Bannon, Linda A. Goldinger, and Gavin S. Appleby, Honesty and Integrity Testing: A Practical Guide (Atlanta, GA: Applied Information Resources, 1989): "Unlike much of the earlier research, studies are beginning to appear occasionally in the open literature after review by other professionals. . . . Honesty test publishers will need to become more supportive of independent efforts if a satisfactory body of research and knowledge is to evolve" (pp. 116-117). The 1989 review article by Sackett et al. (P. Sackett, L. Burris, and C. Callahan, "Integrity Testing for Personnel Selection: An Update," Personnel Psychology, vol. 42, 1989) is also cautiously sympathetic

Nevertheless, the amount of research on integrity test validity has increased considerably in recent years, and according to some reviewers the quality of this body of research has improved. For example, one group of reviewers notes that". . . there has been a substantial increase in the number of studies using an external criterion . . . and significant correlations with absence, turnover, behavioral indicators such as grievances and commendations, and supervisory ratings are being reported."[17] These authors were able to report on 24 studies using external, nonpolygraph criteria in their 1989 review (see below for a discussion of problems in studies using polygraph results as criteria) whereas in 1984, they found only 7 such studies.

Aside from methodological problems, a serious issue concerns the proprietary nature of the tests and the fact that ". . . nearly all research is being conducted by investigators associated with honesty test publishers."[18] While this does not necessarily impugn its quality, it does undermine its credibility. The reasons commonly cited for this state of affairs in integrity test research offer little consolation: the proprietary nature of scoring keys, the difficulty in gaining cooperation from some publishers, and the fact that it is not a traditional area for academic research". . . may help explain the lack of independent research, [but] without independent research there is no compelling response to the speculation that only successes are publicized."[19]

Method of OTA's Review

To conduct its review of the research literature on integrity testing, OTA reviewed the two most current reviews of the integrity testing literature,[20] as well as reviews of specific tests published in test review compendiums.[21] OTA also reviewed copies of tests provided by leading publishers, and

---

albeit more favorable in tone than the earlier work by P. Sackett and M. Harris, "Honesty Testing for Personnel Selection: A Review and Critique," Personnel Psycholoay, vol. 37, 1984 pp. 221-245.

17. Sackett et al., op. cit., footnote 16, p. 507.
18. O'Bannon et al., op. cit., footnote 16, p. 117.
19. Sackett et al., op. cit., footnote 16, p. 521.
20. Ibid.; and O'Bannon et al., op. cit., footnote 16.
21. J. Mitchell (cd.), The Ninth Mental Measurements Yearbook (Lincoln, NE: The Buros Institute of Mental Measurements, University of Nebraska-Lincoln, 1985); J. Conoley and J. Kramer (eds.), The Tenth Mental Measurements Yearbook (Lincoln, NE: The Buros Institute of Mental Measurements, University of Nebraska-Lincoln, 1989); and J. Keyser and R. Sweetland (eds.), Test Critiques (Kansas City, MO: Test Corporation of America, 1987). Note that these reviews are written by single individuals, and are not subject to outside review.

reviewed studies conducted by major test publishers. Many studies using counterproductivity as a criterion were supplied by publishers. These studies are not cited, however, in response to the test publishers' request that only studies published in journals be referenced. The studies provided were used to analyze the methodology used by test publishers to conduct such studies. OTA also conducted interviews with a number of experts on various aspects of testing. Some of these experts are intimately familiar with integrity testing; others specialize in related testing issues.

Concurrent Validation Research

One strategy of concurrent validation research is to compare test results with other accepted measures of a particular behavior.[22] There have been numerous attempts to use polygraph 'cores '" this context, some of which have yielded particularly high validity scores.[23] But reviewers have highlighted numerous problems with some studies of this sort, which

> . . . use only the theft attitudes section as the predictor, while others include . . . theft admissions; some use only admissions made during the polygraph as the criterion, while others use polygrapher judgment about the suitability of the candidate for employment; the time interval between the integrity test and the polygraph is often not specified; [it is not] always clear whether or not candidates expected that a polygraph exam would follow the integrity test [in which case individuals would perhaps decide not to conceal, on the integrity test, history of wrongdoing]; some studies preselect equal numbers of individuals passing and failing the polygraph for inclusion in the study, thus maximizing variance in the criterion and increasing the resulting correlation between test and criterion. .. .24

But perhaps the most obvious reason to be wary of concurrent validation studies using polygraph is that polygraph itself has never been demonstrated to be sufficiently valid when used in personnel selection.[25] In one of the two reviews of integrity test validity research, the authors excluded research that ". . . used polygrapher judgments as a criterion . . . because of controversy surrounding the reliability and validity of polygrapher ratings."[26]

---

22. In the physical sciences, for example, a new instrument designed to measure length would obviously need to be validated against previously accepted instrumentation (e.g., the standard meter, wavelength of light, etc.).
23. Product moment correlations were in the range of 0.29 to 0.86 in 14 studies reviewed by Sackett and Harris, op. cit., footnote 16, pp. 221-245.
24. Sackett et al., op. cit., footnote 16, p. 500.
25. See, for example, U.S. Congress, Office of Technology Assessment, Scientific Validity of Polygraph Testing: A Research Review and Evaluation, Technical Memorandum (Washington, DC: U.S. Government Printing Office, 1983), p. 100.
26. O'Bannon et al., op. cit., footnote 16, p. 70,

Although concurrent validation studies are not considered an adequate substitute for predictive validity,[27] these efforts show promise for measuring similarities between the constructs measured by integrity tests and those measured by other personality and cognitive tests.[28]

Validation Research Using "Contrasted-Groups" Method

The basic principle in this approach to construct validation is that". . . if the honesty test is indeed a good measure of integrity, large differences should be found [between the scores of two groups of people who are known a priori to differ in honesty]."[29] There have been less than a dozen such studies, most of which compare honesty test scores of convicted felons and job applicants. The results have generally shown statistically significant differences (as large as two standard deviations) between average test scores of the two groups.[30] Unfortunately, the underlying assumption that convicted felons have attitudes and lifestyles similar (in construct) to those of normal job applicants or employees "who pilfer small amounts of merchandise atwork" cannot be substantiated.

Admissions of Prior Wrongdoing

A common method of validating honesty tests is to compare a test's predictions based on attitudes to an individual's own confessions of wrongdoing, provided contemporaneously. In other words, for a given definition of dishonesty, admissions of prior acts are compared to how closely responses on the test would have been able to predict the propensity to commit those acts. These tests vary in their definitions of honesty; i.e., what kinds of acts to include in confessions, in the methods used to obtain admissions, and in the ways in which scores and admissions data are associated.

While it is believed that admissions provide more data than detected thefts, researchers recognize the inherent limitations to admissions data as criteria: incentives to withhold information, coupled with the bounds on precision of the definition of the acts to be included in admissions, make the admissions criteria very imperfect. A fundamental logical conundrum is that the admission of a past wrongdoing is itself an act of honesty.

---

27. See, for example, Robert Guion, Personnel Testing (New York, NY: McGraw Hill, 1965), p. 371.
28. Sackett et al., op. cit., footnote 16, p. 515.
29. O'Bannon et al., op. cit., footnote 16, p. 70.
30. Sackett et al., footnote 16, p. 512.

The basic conclusion of various reviews is that there is a positive relationship between honesty test scores and confessions, but that ". . . admissions studies are limited to demonstrating a relationship between two types of self-description. . . ."[31]

The use of admissions data as validity criteria also raises a conceptual puzzle. If these data are assumed to be reliable, i.e., if job applicants included in a validity study sample are assumed to confess prior wrongdoing with candor, then why would this assumption not extend to all job applicants? In a word, why not simply ask job applicants about their prior behavior, rather than use tests designed with (imperfect) surrogates for evidence of prior dishonesty?[32] On the other hand, if the answer is that job applicants will have incentives to conceal some information, or to exaggerate other information, then the question becomes whether that type of information can be admissible as criteria in a validation study.

Predictive Validation Using External Criteria

The most compelling line of research on integrity tests is based on the predictive-validity model, which addresses the following basic question: if an integrity test is used in the process of selecting job applicants in order to screen out individuals most likely to commit certain kinds of behavior, to what extent does the test actually predict the relevant behavior? Thus, most industrial psychologists would agree with the statement that ". . . when the objective is to forecast behavior on the basis of scores on a predictor measure, there is simply no substitute for [predictive validity]."=

There have been two basic approaches to validation research using external criteria in which the unit of analysis is the individual: studies using detected theft as the criterion and studies using other external criteria, such as absenteeism, turnover, and supervisors ratings. The trade-off in the

31. Ibid. These reviewers add that ". . . high correlations are found when correlating the attitude and admission sections of various tests; lower correlations are found when single-item measures (admission of arrests, admission of being fired from a previous job) rather than many composites across many illegal activities are used" (p. 508).

32. O'Bannon et al. (op. cit., footnote 16) raise the same question.

33. W. Cascio, Applied Psychology in Personnel Management (Reston, VA: Reston Publishing Co., 1982), p. 150.

value of these studies can be summarized thus: the former address a principal concern, namely theft at the workplace, but are hindered by the difficulty in detecting theft; the latter are more feasible to conduct, but raise concerns about appropriate measures of outcome criteria. A third approach, in which the unit of analysis is the organization, is discussed below under "Time Series Designs." These studies can use either theft or counterproductivity as external criteria.

Theft Studies

A point frequently raised in this report is that workplace theft is a particularly difficult behavior to use as a criterion -- for evaluating any instrument -- if the assumption that a large fraction of workplace theft goes undetected is true.[34] This problem continues to undermine the credibility of predictive validity studies. Because few researchers believe that detected theft is an accurate measure of true theft, the correlations from their studies are probably inaccurate. To clarify this point, suppose that it is known with certainty that some thieves are caught and some are not. Then the correlation found to exist between test score (predictor) and detected theft (criterion) would be lower than the true correlation, as long as those thieves not detected are assumed to score the same as those who are detected. If, however, detection and test performance are not independent, e.g., if the high scorers are the thieves who are best at evading detection, then the observed correlation could be lower, higher, or the same as the true correlation.

in addition to the basic problem of undetected theft, which may not be able to be remedied by improvements in research design and reporting, independent reviewers -- including OTA -- have identified other design flaws in the available studies attempting to use theft as a criterion. For example, there are problems in criterion definition. In one study, mishandling of cash is equated with stealing, when some of the employees so identified may have been careless rather than dishonest.[35] Another study of Salvation Army bellringers had a similar problem; it did not adequately establish that the monetary differences among volunteers' collections resulted from theft, as the researchers concluded; the volunteers could have been in more or less generous locations.[36]

---

34. This assumption does not necessarily mean that there is a very high rate of theft, but rather than whatever the true rate of theft is, much of it is difficult to detect. The question of detection, then, can be distinguished from the question of incidence.
35. O'Bannon et al., op. cit., footnote 16.
36. Alternatively, the volunteers could have spent less time at their posts, an indicator of

In some studies it is difficult to interpret either methods or results for one or more reasons: several scales developed by the same company are used to screen employees, thus preventing an unequivocal assessment of the honesty scale; numbers in subgroups are not reported; test results for these not terminated for theft are not reported; and statistical tests of significance are not presented.[37]

OTA identified five predictive validity studies in which the criterion measure was either detected theft or a reasonably close proxy. The characteristics of these studies, chosen because their research design involved predictive validity, are summarized in table 7. Two of these studies involved applicants for jobs in the grocery industry; two of the studies involved department stores; and one study was of a national convenience store chain. All the studies were conducted by the publisher of the integrity test analyzed in the studies.

Table 8 presents the raw frequency counts as reported in the respective studies. The top row in these tables gives the number of employees not caught committing theft, and the bottom row gives the number detected; these figures are cross-tabulated by test performance as marked in the studies. Note that because some theft undoubtedly is not detected, the bottom row in each table potentially underestimates the true amount of theft. To illustrate the meaning of these tables, consider Study # 2: a total of 3,790 employees were given the test and hired regardless of their test performance. Subsequent investigations by management revealed that 91 employees had committed some type of theft. Among these 91, 75 had failed the integrity test and 16 had passed. Among the 3,699 for whom the investigation did not reveal any theft, 2,145 had failed the test and 1,554 passed. Thus, 75 of those taking the test (2 percent of the total 3,790) are known to have been characterized correctly by the test, and 16 are known to have been characterized incorrectly. But what about the rest? If those 3,699 not detected as thieves are assumed to be honest, then 2,145 (58 percent) were misclassified; if a substantial number of them were indeed thieves, the observed correlation between the test and the outcome measure could be higher, lower, or equal to the actual correlation.

A central concern for public policy is the potential for classification errors, especially of honest

---

counterproductive behavior, though not outright theft (0' Bannon et al., op. cit., footnote 16).

37. See O'Bannon et al., op. cit., footnote 16; and Sackett et al., op. cit., footnote 16.

Table 7- Predictive Validity Studies of Overt Integrity Tests Using Detected Theft or Close Proxy as Criterion

Study	Sample size	Criterion	Test performance* Number passed (percent of total)	Number failed (percent of total)	Number of persons detected committing theft or other dishonest act (percent of total)
1	479	"Thefts detected by admissions and/or signed statements of employees."	241 (50%)	238 (50%)	(3.5%)
2	3,790	'Terminated for reasons of dishonesty."	1,570 (41.4)	2,220 (58.6)	91 (2.4)
3	527	"Discharged for theft or some related offense."	173 (32.8)	354 (67.2)	33 (6.3)
4	61	"Caught stealing cash/ merchandise or disciplined for mishandling company cash/ merchandise."	50 (82.0)	(I8)	6 (9.8)
5	801	"Caught stealing."	472 (58.9)	329 (41.1)	(2.6)

* "Passed" or "failed" in these studies reflect cut scores defined for research purposes. These cut scores mayor may not be the cut scores used by any given employer.

SOURCE: Office of Technology Assessment.

Table 8

Forecasting Efficiency of Integrity Tests (2x2 Contingency Tables for Validation Studies Using Detected Theft or Close Proxy for Criterion)

Study 1

	Failed test	Passed test	Total
Not detected	222	240	462
Detected	16	1	17
TOTAL	238	241	479

Study 2

	Failed test	Passed test	Total
Not detected	2,145	1,554	3,699
Detected	75	16	91
TOTAL	2,220	1,570	3,790

Study 3

	Failed test	Passed.test	Total
Not detected	326	168	494
Detected	28	5	33
TOTAL	354	173	527

Study 4

	Failed test	Passed test	Total
Not detected	8	47	55
Detected	3	3	6
TOTAL	11	50	61

Study 5

	Failed test	Passed test	Total
Not detected	318	462	780
Detected	11	10	21
TOTAL	329	472	801

SOURCE: Office of Technology Assessment.

persons incorrectly identified as dishonest. Table 9 shows that the overall level of misclassification in these studies ranged from 18 percent (in a study with small sample size) to over 60 percent. From less than 1 percent to 6 percent of those passing the tests (that is, identified by the tests as honest) were later found to have stolen from their employers, meaning that upwards of 94 percent of those identified by the tests as thieves were correctly identified.[38] Such reported results are no doubt compelling to employers. But of concern to potential employees, the data in the fourth column of the table suggests why the predictive validity research, even if it is found to be valid, provokes public controversy: of those classified as dishonest on the basis of an integrity test, the proportion who are not detected committing theft ranges from 73 to 97 percent. These data are useful to illustrate the divergence between possible consequences that is at the core of the public policy dilemma.

Counterproductivity-Based Studies

In contrast with the limited amount of research relying on detected thefts for criterion measures, there have been many studies using a variety of counterproductivity-based outcomes, including supervisory data, terminations, and absenteeism. One of the two principal reviews reported on the results of a number of these studies,[39] although they did not evaluate in depth each study's design and conduct.

Measures of counterproductivity used as outcome variables vary considerably. Some measures are specific and discrete (e. g., absenteeism, terminations) and some consist of composites. Some measures are counts from employee records and some are supervisors' ratings. Objective measures of counterproductive behavior include tardiness, absenteeism, accidents, number of worker compensation claims, voluntary turnover, terminations for theft or gross misconduct, and damage to property. Indicators of "productivity," such as mean number of days employed, are also used. Similarly, supervisors' ratings are made of overall performance or misconduct, *or* of more specific

38. It is important to note that the studies used different definitions and measures of theft, and are methodologically flawed.

39. Sackett et al., op. cit., footnote 16. There were no such studies in 1984 when Sackett and Harris conducted their first review. O'Bannon and his colleagues explicitly excluded most studies using counterproductivity as a criterion. There was only one predictive study reviewed by O'Bannon et al. (op. cit., footnote 16) that used terminations as a criterion, and that study focused primarily on terminations for theft.

Table 9- Classification and Misclassification in Five Predictive Validity Studies Using Detected Theft or Close Proxy as Criteria*

Study	Correct classifications: Of those passing test, not detected	Correct classifications: Of those failing test. % detected	Percent of total sample misclassified	Misclassifications: Of those failing test, % not detected	Misclassifications: Of those passing test, % detected
1 & 2	99.6	6.7	46.6	93.3	0.4
	99.0	3.4	57.0	96.7	1.0
3	97.1	7.9	62.8	92.1	2.9
4	94.0	27.3	18.0	72.7	6.0
5	97.9	3.3	40.9	96.7	2.1

* "Passing" and "failing" in these studies reflect cut scores defined for research purposes. These cut scores mayor may not be the cut scores used by any given employer.

SOURCE: Office of Technology Assessment.

measures such as absenteeism and tardiness. This variety of criteria reflects the attempts of researchers to generate useful information. It does, however, make an overall judgment about predictive validity difficult.

Research results from these studies are reported in primarily two ways: (1) in terms of correlation coefficients that serve as a measure of association between integrity test scores and one or more indicators of counterproductive behavior, usually scored continuously; and (2) in terms of proportions of the honest and dishonest individuals who are correctly and/or incorrectly identified by the tests.

As for the theft studies, OTA reviewed a number of counterproductivity studies in order to evaluate their methodology; and as with the theft studies, issues arose with respect to both study design and criterion measurement.

For example, in one study, 169 hotel industry applicants were tested and hired regardless of test scores. The criterion measure was termination. This study, although flawed, suggested somewhat better results, from the point of view of misclassification, than those shown in table 9. First, with respect to persons who "failed" the test: the study showed that among these 53 applicants (31 percent of the total sample), 16 (30 percent) remained employed. Second, among the 116 who "passed," and were therefore presumed honest, 49 (or 42 percent) were eventually terminated. It should be noted, however, that just as detected theft probably underestimates the true amount of theft in the studies reported in table 8, the termination variable in this study probably overestimates dishonesty: there is substantial ambiguity over the causes of termination.[40]

Despite (or because of) flaws in methodology and reporting, the predictive correlational studies reported by Sackett and his colleagues found a range of generally low, but statistically significant, associations between a range of integrity test scores and a wide range of counter-productive measures. Correlation coefficients ranged from 0.16 to 0.62; only one study reported a correlation coefficient greater than 0.35.[41]

40. This is a good example of the trade-off between "criterion variance" and "method variance." See Sackett et al., op. cit., footnote 16, p. 507.

41. From 3 to 38 percent of the variance in counterproductive behaviors would be predicted (explained) by the test scores in a multiple regression model.

It is not possible to ascertain from the studies reporting only correlation coefficients the proportions of honest and dishonest individuals correctly and incorrectly classified. In three studies providing the relevant data, misclassification of dishonest individuals ranges from 17 to 29 percent; in two of these studies, 22 and 29 percent of honest individuals were misclassified. Another study found that the mean number of days employed was significantly higher among those passing the test (95 versus 87 days in the year of the study) .42

Time-Series Designs

Studies that focus on the reduction of organization-level inventory losses and counter-productivity have been termed by some "time-series designs."[43] Almost all of the studies included in the two published independent reviews reported reductions in shrinkage, overall levels of terminations, *or* counterproductive behavior after introduction of the tests.[44] However, flaws in the research designs made it difficult to determine the sources of the change. The most prominent of the flaws was the failure to use appropriate control groups, thus leaving open the possibility that other factors (e.g., seasonal fluctuations in shrinkage; changes in management; perceived changes in company tolerance of theft) accounted for the observed improvements.[45]

In one study,[46] the greatest reduction in shrinkage occurred in the first 2 months after a switch from polygraph to integrity testing screening. The reviewers note, however, that unless there was extraordinarily high turnover, use of the integrity test for selection could not have been the reason for this sudden reduction.[47]

---

42. It may be important to note that Sackett et al. (op. cit., footnote 16) reported both correlation coefficients and dichotomous results for only one study; therefore there is almost no overlap between these types of studies, and results of the studies reporting both types of predictive error may not be generalizable to the studies reporting a single correlation coefficient.

43. O'Bannon et al., op. cit., footnote 16; and Sackett et al., op. cit., footnote 16.

44. Because of differences in measurements used by the various studies, it is not possible to report a meaningful range of results. For example, one study reported a correlation of 0.68 between scores on tests taken by convenience store managers and average monthly store shortage reduction figures. Another reported that 80 percent of all terminations for theft occurred in the control group stores. A third reported a reduction in the termination ratio; a fourth reported both average monthly reductions in terminations for theft and average monthly total voluntary reductions.

45. According to O'Bannon et al. (op. cit., footnote 16), the one study that did use two control groups found that differences in shrinkage among the stores involved in the study were not statistically significant (reported in O'Bannon et al.).

46. Ibid., pp. 88-89.

47. Most employees -- the same ones who were with the company during the baseline measures -- would still be with the company. See ibid., pp. 88-89.

In addition, the following problems were observed in one or more of these studies:

- inappropriate measurement of shrinkage, including shrinkage and cost-savings estimates not based clearly on the study organizations themselves, but on industry averages;[48]
- use of other predictive scales in addition to honesty scales, thus making it difficult to disentangle the effects of the honest scales;[49] and
- concurrent use of polygraph testing for screening a subset of employees.[50]

Reviewers are skeptical about the available time-series studies for these and other reasons, but they believe the results of these studies are grounds for guarded optimism about continued research. While noting that a problem with these studies is the unreliability of the criterion measure (". . . in at least some of the studies it is evident that error is present in the measure of shrinkage. . ."), one reviewer concludes that while ". . . this group of studies cannot be considered unequivocal in demonstrating the validity of honesty tests . . . they do begin to establish a foundation of evidence which may become more convincing as additional studies accumulate."[51]

General Remarks

Industrial and organizational psychologists recognize the difficulty in surmounting methodological barriers to the "ideal" predictive validity study. For example, ". . . the most useful study would be one in which no other selection screening is done, providing a 'pure' examination of the honesty test."[52] The appeal of this model is tempered, however, by the test publishers' claim that their tests are not intended to be the sole (or even the primary) selection criterion.[53] Thus, the truly ideal study would be one in which the various selection procedures continue to be used in combination, but

48. Ibid.
49. Ibid., p. 92.
50. Sackett et al., op. cit., footnote 16.
51. O'Bannon et al., op. cit., footnote 16, p. 92.
52. O'Bannon et al., op. cit., footnote 16, p. 79.
53. See Association of Personnel Test Publishers, Model Guidelines for Preemdoyment Integrity Testing Progams, 1st ed. (Washington, DC: 1990).

which accounts explicitly for the independent effects of the honesty test and for the interaction effects between the test and the other screening procedures. This type of study would not be easy to carry out.

With respect to counterproductivity-based studies using supevisory ratings, in particular, ". . . for a fair assessment to be made, test scores should not be known within the company while the data is being collected . . . [so that the scores cannot] influence the outcome by biasing the opinions of managers toward some employees."[54] In other words, human resource professionals and industrial psychologists recognize a common feature of experiments in the physical and social sciences, i.e., the "double-blind" model. Few of the reported studies indicate whether test scores intended for use in reaching hiring decisions are kept secret from individuals assessing employee performance, and if they were, how it was handled.

Methodological constraints notwithstanding, prominent academic and industrial psychologists, have reviewed the results of the available predictive validity studies. Although these reviews have been conducted by individuals who are generally sympathetic with the objectives of psychological and personnel testing, their findings are couched in cautious tones and their principal conclusion is that better research is very much needed:

> **The most clear cut finding from reviewing predictive validity studies is an observation on the state of this body of research. . . . The field of honesty testing has a great need for producing additional high quality studies in this area.[55]**

54. O'Bannon et al., op cit., footnote 16, p. 79.

55. Ibid., p. 85.

Chapter 3

PUBLIC POLICY IMPLICATIONS OF PRE-EMPLOYMENT HONESTY TESTING

Chapter 3

POLICY ISSUES CONCERNING THE USE OF INTEGRITY TESTS

American society uses many types of tests to assess individual capabilities and attributes, and to inform screening and selection decisions in education and employment.[1] Aptitude tests measure ability in verbal, logical, or mathematical domains; standardized achievement tests tend to focus on attainment of knowledge in more specific areas, usually with reference to defined educational goals; and personality tests are concerned with affective aspects of behavior, such as emotional adjustment, motivation, interpersonal relations, and attitudes.

Although tests can differ fundamentally in their design and in the underlying constructs they measure, they all share a very basic characteristic: they are ". . . imperfect and therefore potentially misleading as measures of individual performance in education and employment."[2]

This chapter addresses several questions for policy makers deliberating the uses of honesty and integrity tests: effects of integrity test fallibility and classification error, potential discriminatory consequences of integrity test use, and privacy considerations. These negative effects must always be weighed against potential benefits to firms and society at large.[3]

---

1. Tests are also used in the biological and medical arena, to provide information on the presence of particular clinical conditions, to forecast future illness based on genetic characteristics, to evaluate the usefulness of a new treatment, and so forth. These tests, like others, are imperfect, and the evaluation of their accuracy and efficacy is critical. For a discussion of one approach to measurement of accuracy and usefulness see, for example, Mark Zweig, "Evaluation of the Clinical Accuracy of Laboratory Tests," Archives of Pathology and Laboratory Medicine, vol. 112, April 1988, pp. 383-386; and for discussion of classification error in AIDS Antibody Testing see Office of Technology Assessment, testimony of Lawrence Miike before the House Committee on Small Business, Subcommittee on Regulation and Business Opportunities, Oct. 19, 1987. Issues pertaining to the accuracy and uses of genetic screening methods are discussed in Dorothy Nelkin and Laurence Tancredi, Dangerous Diagnostics: The Social Power of Biological Information (New York, NY: Basic Books, Inc., 1989).

2. The National Commission on Testing and Public Policy, From Gatekeeper to Gateway: Transforming Testing in America (Chestnut Hill, MA: 1990), p. 6.

3. These benefits include reduced costs of screening and selection, which depend in large part on the tests' predictive accuracy relative to the accuracy of other screening and selection methods; reduced workplace theft and counterproductivity; and, ultimately, increased productivity, which could benefit individual firms and the Nation as a whole. OTA did not evaluate these benefits in detail.

TEST FALLIBILITY: ISSUES FOR CONSIDERATION

Because all tests are imperfect, projections made from test scores are not necessarily accurate representations of test-takers' future behavior or performance. Such 'classification error" is always a possible consequence of test use. But the effects of classification error from different types of tests are not necessarily the same. For example, most people would probably find it unpleasant to be mistakenly classified as below some standard in arithmetic ability necessary to perform successfully in a job or at school, but both the individual and social consequences of being misidentified as carrying (or not carrying) a deadly disease are surely different. For many people, too, there appears to be something special about the potential for erroneous classifications into categories that suggest they are "honest" or "dishonest." Understanding some of the reasons that misclassification from integrity tests can be particularly onerous can be helpful to policymakers.

Self Control and the Presumption of Innocence

As discussed in chapter 1 of this Report, there is considerable debate within the psychological community over the relative importance of personality traits and environments (or situations) in determining human behavior. Another question hinges on the extent to which an individual is able to control a given personality trait, assuming the trait exists.

For example, if a person tests positive for the trait called "dishonesty," i.e., is identified as at high risk for committing certain acts defined as dishonest, he or she may still be able to control future behavior and, in a sense, "overcome" the existence of the trait. And this ability to exercise the requisite self control may also be affected by situational variables. This raises a methodological problem, in that the presence of the "self-control" trait would need to be incorporated in research on test accuracy in predicting propensities to commit dishonest or counterproductive behavior. There has not been research in which this aspect of the problem was addressed explicitly.

Aside from the measurement problem, this issue of self control raises a more basic question.

---

4. This section draws on Mark Kelman, "A General Framework for Evaluating Classification Errors, With Special Reference to Integrity Testing," OTA contractor report, June 26, 1990.

American society does not generally require proof that individuals can exercise control; we rather take that as an implicit assumption. Our society generally prefers to grant people the opportunity to prove themselves as individuals, and shuns prejudgment based on one's identification with a group whose mean level of performance in a given domain (or on a given test) is higher or lower than the mean level of other groups.[5]

A key question, therefore, is whether a trait such as dishonesty is immutable. In the absence of evidence on the immutability of the trait, it can be argued that reliance on an integrity test score could deny an individual's claim to self control, and thus signal a departure from the axiom that society punishes only bad deeds and not bad dispositions. Whether society is willing to tolerate chips in the armor provided by the "presumption of innocence" is a question that Congress will have to consider if it debates policy over the use of integrity tests.

What Happens to the Misclassified?

Even if integrity tests were the sole basis for employment decisions, which no one recommends,[6] applicants who do poorly on the test would not necessarily be excluded from all employment opportunities. If the labor market is functioning properly, and if there are some jobs that simply have lower requirements for honest employees, then theoretically it would be possible for individuals who are labeled dishonest to find work eventually.

In general, though, the relatively lower demand for dishonest workers would theoretically drive down wages and employment opportunities for individuals classified as dishonest. The question becomes, then, whether this situation would persist, i.e., whether individuals who score low on an integrity test would systematically be denied employment in the future. Two possibilities warrant consideration: first, if test results are made available in the labor market, in databases, or through other means, then failing even a single test could have longer term repercussions. (The question of information access is discussed below under "Privacy.")

5. Exceptions are made, however, in cases involving public safety and similar concerns.

6. The test publishers caution employers against relying on any single pre-employment screening device. See Association of Personnel Test Publishers, Model Guidelines for Preemployment Integrity Testing Programs, 1st ed. (Washington, DC: 1990).

Second, if integrity tests are reliable (in the sense that individuals who are tested repeatedly do not vary significantly in their test performance), as the test publishers claim, then their use could create a population of persons who are repeatedly misclassified, and who are systematically denied employment without cause. Alternative methods to screen out dishonest job applicants, such as subjective interviews or letters of reference, are also imperfect and can result in erroneous decisions. They are, however, less likely to be as consistently wrong as integrity tests about specific individuals. Assuming even a modest error rate, widespread use of the tests could deny opportunity to a sizable number of persons.

Aside from potential economic loss -- denial of employment -- based on erroneous test scores, use of integrity tests could cause injury because of the stigma of failure. Comparison with cognitive ability tests can again be illustrative: performance on cognitive tests implies no global judgment of a person, but can indicate that he or she is likely to be more or less productive than someone else in certain jobs. There are no jobs, however, for which dishonesty is either required or preferred. Thus, if individuals learn their test results they could suffer from the implied label; and if scores become accessible to potential employers (other than those who administered the test) or to others in the community, the low scorers could suffer a social stigma.[7]

These potentially stigmatizing effects are made sharper by virtue of the tests' scientific imprimatur. Because many employers will want evidence of a test's accuracy before purchasing it, tests publishers have an obvious incentive to provide evidence that their products have been validated in scientific studies. The result is that individuals are not deemed dishonest or counterproductive by "whim" but rather by dint of a psychological instrument that has been proven in repeated experiments. Thus, while some tests might theoretically result in less overall misclassification than other screening methods, the tests could also intensify the effects of misclassification on those who are misclassified.

Another potential consequence of integrity test error -- and clearly an unintended consequence from the point of view of employers and test publishers as well as test-takers -- is the possibility that erroneous classification of honest persons as dishonest will be self-fulfilling. If low

7. It is important to note that integrity test publishers advise against informing applicants of their test scores. See below, for discussion of shared data, access, and related privacy issues.

scorers are erroneously denied employment, it can be argued that they are paying an unearned economic penalty; if so, they might rationally conclude that they will be made no worse off -- and possibly be made better off -- by engaging in the behavior for which they have already been penalized.[8] The possibility that use of the tests might lead some people to behave dishonestly -- quite the opposite of their intention -- warrants consideration and research.[9] To date there has been no empirical research to test this proposition.

What Happens If Pre-Employment Integrity Tests are Not Used?

A common argument for using integrity tests is that they are "more valid" than other pre-employment screening methods, i.e., that they are more accurate in predicting outcomes of interest. Some test publishers and researchers therefore claim that integrity tests result in fewer classification errors than other selection methods. This claim is not necessarily correct. First of all, most other pre-employment screening methods do not classify people in terms of honesty. They may "misclassify" individuals -- as a poor credit risk or as lacking some skill, for example -- but they do not characterize applicants as not honest. In fact, not all methods result in classification; a random procedure, for example, may reject some individuals without classifying them.

Moreover, where research that compares rates of error of integrity tests with other nonrandom hiring methods has been attempted, it has relied on estimates of the prevailing rate of theft (the base rate) and on estimates of the conditional probabilities (i.e., the frequencies of correctly and incorrectly classified subjects) derived from correlation coefficients reported in other research

8. This reasoning follows in the tradition of the "moral hazard" literature. See, for example, Kenneth Arrow, Essays in the Theory of Risk Bearing (Chicago, IL: Markham, 1971); Mark Pauly, "The Economics of Moral Hazard," American Economic Review, vol. 58, June 1968, pp. 531-537; and Oliver E. Williamson, Markets and Hierarchies (New York, NY: The Free Press, 1975).

9. The fact that tests themselves can stimulate -- and not just measure -- behavior, is a central concern in the educational testing community. See, for example, Edward Haertel, "Student Achievement Tests as Tools of Educational Policy: Practices and Consequences," Test Policy and Test Performance: Education, Language, and Culture, B.R. Gifford (ed.) (Boston, MA: Kluwer, 1989); and A. Collins and J. Fredrikson, "A Systems Approach to Educational Testing," Educational Researcher, vol. 18, No. 9, December 1989, pp. 27-32. Note that neither of these articles addresses integrity tests.

studies.[10] There has been very little comparative research of this type, and it is not possible to draw any firm conclusions.

POTENTIAL DISCRIMINATORY EFFECTS: ISSUES AND EVIDENCE

An important concern about the effects of integrity tests is whether members of various ethnic, racial, or gender groups could suffer from discrimination in hiring because of test results. This is particularly important with respect to protected groups in society, and much of the research that has been conducted on discrimination has focused on so-called "adverse impact" considerations. Indeed, the framework of civil rights laws is a ". . . key consideration in adopting and maintaining a testing program [and] test publishers can market their products more easily if they can advise potential users that their tests do not require legal validation.'"

Many integrity test publishers have conducted adverse impact research. Their studies report a variety of findings: in some cases no statistically significant differences between groups' average test scores are found, in other cases there appears to be a favorable bias toward protected groups (minorities, women, and the elderly), and in other cases minority groups (i.e., Blacks and Hispanics) appear to do less well than whites.[12] Based on the studies supplied by the authors and publishers of honesty tests, their instruments appear to be free of adverse impact.[13]

Four caveats must be noted. First, as stated earlier, research conducted by test publishers, without independent replication, raises credibility issues. '4

Second, ". . . in some cases, the data used to demonstrate lack of adverse impact was not collected in the employment setting and may not accurately represent the way the test will operate

10. See, for example, S. Martin, "Estimating the False Positive Rate for Alternative Measures of Integrity," Journal of Business and Psychology, vol. 4, No. 3, spring 1990, pp. 385-389.

11. R. Michael O'Bannon, Linda A. Goldinger, and Gavin S. Appleby, Honesty and Integrity Testing: A Practical Guide (Atlanta, GA: Applied Information Resources, 1989), p. 93.

12. For a summary of the specific studies providing data on adverse impact see ibid., pp. 94-99.

13. OTA did not independently review all the adverse impact studies. One team of reviewers that did concluded that ". . . a large number of honesty test authors have supplied studies demonstrating lack of adverse impact of their instruments. Many of these studies appear to meet the necessary research standards for this type of research." Ibid., p. 99.

14. "Currently available adverse impact studies do not escape the criticism of some detractors that the majority of research on honesty testing is conducted or sponsored by the testing companies.

with job applicants. Other studies simply provided no information at all about how the data was gathered. A second issue is the size of the samples used. Some studies included only a small number of participants. . . ."[15]

Third, most of the adverse impact research relies on application of the "4/5th rule," a convention suggested by the Equal Employment Opportunity Commission[16] and widely used by employers in evaluating their hiring and promotions practices. This rule of thumb stipulates that a hiring rate for a minority group that is less than 80 percent of the rate for the majority will be regarded as evidence of adverse impact of the hiring system. While the research conducted by integrity test publishers suggests that the tests do not violate this standard, there is debate over its appropriateness as the sole criterion in making judgments of discrimination. Indeed, the Uniform Guidelines note the possibility that ". . . smaller differences in selection rates [than would constitute discrimination under the 4/5th rule] may nevertheless constitute adverse impact, where they are significant in both statistical and practical terms. . . ."[17]

The courts have relied largely but not exclusively on the 4/5th rule approach. "In one case a trial court declined to follow the 80-percent rule where the acceptance rate for minorities was 81.55 percent of that of majority candidates, but there was expert testimony that the disparity nevertheless was statistically significant. At the other extreme, another trial found a clearly 'significant discriminatory pattern' of selection from a test which eliminated about one-fourth of female applicants but only about one percent of male applicants."[18] If the courts shift their stance toward more stringent statistical criteria, more research will be necessary to resolve the question of adverse impact. 'g

---

Suppliers attempting to market their tests have a vested interest in sharing favorable research but may be unwilling to report studies which show adverse impact." Ibid., p. 101.

15. Ibid., p. 93.

16. U.S. Civil Service Commission, Department of Labor and Department of Justice, "Uniform Guidelines on Employee Selection Procedures," Federal Register (Washington, DC: U.S. Government Printing Office, Aug. 25, 1978), pp. 38290-38315.

17. Ibid., p. 38297.

18. David Cathcart and R. Lawrence Ashe (eds.), Five Year Supplement to Schlei and Grossman's Employment Discrimination Law, 2nd ed. (Washington, DC: Bureau of National Affairs, 1989), pp. 39-40.

19. Even if the 4/5th rule is the criterion of choice, research problems remain. Sackett et al., for example, point out that in studies they reviewed, ". . . in no case does the black/white passing rate approach the .80 value used as a rule of thumb for adverse impact determination." But, they add, ". . . the presentation of data at different cut scores does illustrate the difficulty in attempting to compare white/black pass rates across tests. . . ." P. Sackett, L. Burris, and C. Callahan, "Integrity Testing for

Finally, this issue is further complicated because it is not entirely clear whether adverse impact can refer to test scores alone or whether there must be evidence that the scores lead to differences in hiring. Evidence on the precise role of test scores in hiring does not exist in the aggregate, and there have been no individual cases decided in which plaintiffs argue that an integrity test per se was the basis of discrimination.

Because of the questions raised in this discussion, OTA concurs with the finding that while currently available studies may be reassuring to employers, ". . . it cannot be fairly said that the coast is clear."[20] Additional research is required in order to inform policy deliberations concerning discrimination and adverse impact of integrity tests.

PRIVACY

Whether questions on pre-employment tests represent an invasion of privacy is not a new issue.[21] In 1965 the Senate Subcommittee on Constitutional Rights of the Committee on the Judiciary, chaired by Senator Sam Ervin, and the House Special Subcommittee on Invasion of Privacy of the Committee on Government Operations, chaired by Representative Cornelius E. Gallagher,[22] held hearings to determine whether the questions asked on psychological tests used by the Federal Government were an unjustified invasion of the respondent's psyche and private life. The Subcommittees also investigated the validity of these tests and the due process issues involved in test administration. The reactions of the press and public were very critical of the types of questions asked

Personnel Selection: An Update," Personnel Psychology, vol. 42, 1989, p. 498.

20. O'Bannon et al., op. cit., footnote 11, p. 101.

21. See Myron Brenton, The Privacy Invaders (New York, NY: Coward-McCann, 1964); M.L. Cross, The Brain Watchers (New York, NY: Signet, 1963); Vance Packard, The Pyramid Climbers (New York, NY: McGraw-Hill, 1962); Vance Packard, The Naked Society (New York, NY: D. McKay Co., 1964); and William H. Whyte, Jr., The Organization Man (New York, NY: Simon and Schuster, 1956).

22. Rep. Gallagher proposed the study which resulted in the 1964 hearings on "Use of Polygraphs as 'Lie Detectors' by the Federal Government" by the House Committee on Government Operations, Subcommittee on Foreign Operations and Government Information (See House Committee on Government Operations, Report No. 198).

on these psychological tests.[23] Congressional hearings and media attention generated increased involvement by psychologists and the American Psychological Association in public debates and in drafting guidelines for test construction and test use.[24]

In 1967, the Office of Science and Technology of the Executive Office of the President established a panel to examine the issue of privacy and behavioral research, and to propose guidelines for those engaged in behavioral research. The panel defined the right to privacy as". . . the right of the individual to decide for himself how much he will share with others his thoughts, his feelings, and the facts of his personal life."[25]

A critical question, then, is what information a test should try to obtain. In order to protect the privacy of the individual, psychologists maintain that the information must be relevant to the stated purpose of the test.[26] Most psychologists agree that, to ensure that an individual's right to privacy is not violated, ". . . there must be valid psychological reasons for having the particular information sought in making the assessment."[27]

But even a valid psychological reason must be weighed along with social and ethical concerns to determine the appropriate balance between the individual's right to privacy and the employer's right (or need) to choose employees who will not commit certain acts. Each of these competing interests must be analyzed.

Publishers of paper-and-pencil integrity tests argue that applicants do not find taking the test to be offensive. In some cases, the evidence test publishers offer is based on responses to a test question (that is not scored) asking whether the respondent resented answering the questions. One

23. Alan F. Westin, Privacy and Freedom (New York, NY: Atheneum, 1970), especially ch. 6, "Probing the Mind: Psychological Surveillance" and ch. 10, "Prove That You're Adjusted."

24. See, for example, "Special Issue: Testing and Public Policy," American Psychologist vol. 20, No. 11, November 1965, in which testimony from the Senate and House hearings was reprinted, as well as statements by a number of prominent psychologists on the professional responsibilities of psychologists.

25. Executive Office of the President, Office of Science and Technology, Privacy and Behavioral Research (Washington, DC: U.S. Government Printing Office, February 1967).

26. Anne Anastasi, "The Use of Personal Assessment in Industry: Methodological and Interpretive Problems," Personality Assessment in Organizations, H. John Bernardin and David A. Bownas (eds.) (New York, NY: Praeger), pp. 7-10.

27. James Neal Butcher, "Personality Assessment in Industry: Theoretical Issues and Illustrations," Personality Assessment in Organizations (New York, NY: Praeger, 1985), p. 281. See also S.R. Hathaway, J.C. McKinley, and James Butcher, MMPI-2: Minnesota Multiphasic Personality Inventory-2 -- User's Guide (Minneapolis, MN: National Computer Systems, Inc., 1989).

study conducted by employees of an integrity test publishing company found that 82 percent of 224 job applicants who responded to an open-ended question reported no objections to taking the pre-employment honesty test. In a companion study, 11 percent agreed that this type of questionnaire was an invasion of privacy, while 69 percent disagreed. Three percent resented being asked to answer such a questionnaire, while 78 percent did not. Five percent thought that administering this type of questionnaire reflects negatively on the company, while 82 percent thought that it did not. Unfortunately, these results are not conclusive: one cannot rule out the possibility that respondents who have an interest in "passing" the test and being hired are not entirely candid in their answers.

Similar research has involved the reactions of a sample of college students, 84 percent of whom had work experience in industries where honesty tests were common, to taking a paper-and-pencil integrity test. This group was generally more strongly negative. Forty-two percent agreed, either strongly or somewhat, that such a test was an invasion of privacy, while 44 percent disagreed. Twenty-six percent said that they would resent being asked to take such a test, while 49 percent said that they would not. Thirty-three percent thought that administering this type of test reflected negatively on the company, while 43 percent thought that it did not.[28]

Although there is variation in what individuals consider personal, there are some questions that are generally regarded as invasive. There are also other questions about one's personal status that legally cannot be asked, either because they may have a discriminatory effect, e.g., prior arrests, or because the information is not considered relevant or reliable, e.g., religious affiliation.

This leaves a large gray area into which some individuals may feel that some of the attitude and admissions questions asked on integrity tests fall, e.g., "do you always tell the truth?;" "how many people do you like?;" "how strong is your conscience?;" "do you ever feel guilty?;" "do you ever treat people unfairly?;" "do you think your conscience would bother you if you cheated someone who cheated you?;" "how often do you blush?;" "how often have you been so upset that you wanted to leave home?" In addition, open-ended questions, e.g., "tell us what you dislike about yourself," that appear on some tests may also elicit information that individuals would not want to divulge.

28. Ann Marie Ryan and Paul R. Sackett, "Pre-employment Honesty Testing: Fakability, Reactions of Test Takers, and Company Image," Journal of Business and Psychology, vol. 1, No. 3, spring 1987, pp. 253-254.

In pre-employment screening, the individual's right to privacy must be balanced against the employer's (and society's) need for employees who will not steal or otherwise be counter-productive.[29] If it were necessary to demonstrate this need in court, employers might be asked for evidence on the magnitude of the employee theft problem. Additionally, they might be asked if there were other techniques available to screen employees or to monitor workers that posed less of a threat to privacy.

Perhaps the central reason that the privacy debate is difficult to unravel is that although privacy is a fundamental value in our society, it is not well conceptualized and is difficult to define. Three central aspects of privacy do recur, however, in regard to integrity testing.

First, there is the notion that certain types of information are inherently private. Second is the concept of a boundary between the individual and others; people should know the boundary between themselves and others and understand what information is crossing it. The third conceptual issue is the responsibility of organizations with regard to personal information.

Are the test questions themselves invasive? Are they necessary to determine whether a person is honest? Next, and perhaps most important, is the interpretation of the answers. An applicant may believe that his or her answers to a question or series of questions is legitimate, but if the answers are then interpreted to make specific conclusions about propensity for future behavior, the applicant may feel that his or her privacy has been invaded.

With regard to the use of and access to test answers, it maybe useful to consider the code of fair information practices developed in 1973 by an Advisory Committee to the Secretary of Health, Education, and Welfare on automated Personal Data Systems. These principles serve as the basis for information privacy legislation, including the Fair Credit Reporting Act, the Privacy Act, the Financial Privacy Act, and the Video Privacy Act. They are:

- there must be no personal data recordkeeping system whose very existence is secret;

29. The guidelines for use of the MMPI-2, as noted in ch. 1, are explicit on this point, reserving tests that invade privacy for situations of potential pubic hazard. See Hathaway et al., op. cit., footnote 27.

- **there must be a way for an individual to find out what information about him or her is in a record and how it is used;**

- **there must be a way for an individual to prevent information about him or her that was obtained for one purpose from being used *or* made available for other purposes without his or her consent:**

- **there must be a way for an individual to correct or amend a record of identifiable information about him or her; and**

- **any organization creating, maintaining, using, or disseminating records of identifiable personal data must assure the reliability of the data for their intended use and must take precautions to prevent misuses of the data.[30]**

Confidentiality

Although related to the right of the job applicant to decide what information to disclose, confidentiality is distinguished in that it involves the responsibilities of those to whom the applicant has disclosed information. This entails restricting third party access to the information and protecting the security of the information from unauthorized access.

The Equal Employment Opportunity Commission's (EEOC) Uniform Guidelines on Employee Selection Procedures (29 CFR Part 1607.1) require an employer to keep documentation, including the records of the component process, for selection procedures that may arguably have an adverse impact. For this purpose, most employers would be likely to retain information on test results and copies of the tests themselves. EEOC general regulations also require an employer to retain

---

30. U.S. Department of Health, Education, and Welfare, Records, Computers and the Rights of Citizens (Washington, DC: U.S. Government Printing Office, 1973).

applications and supporting material for 6 months; States often have similar requirements. Another EEOC regulation (Form 100, Employer Information Report EEO-1) requires employers to keep information on race, sex, and ethnic background in a separate file from personnel records. There appear to be no legal restrictions on employers retaining integrity test results, and/or copies of the tests themselves, in an employee's personnel file.[31] Some reviewers have expressed concern that extensive files on individuals' minor offenses (gleaned from test answers) could be kept in centralized databanks.

With respect to third-party access to information, the Standards for Educational and Psychological Testing state that:

> Test results identified by the names of individual test takers should not be released to any person or institution without the informed consent of the test taker or an authorized representative unless otherwise required by law. Scores of individuals identified by name should be made available only to those with a legitimate, professional interest in particular cases.[32]

The Model Guidelines of the Association of Personnel Test Publishers suggest that test publishers be consistent with this standard, i.e., that the employer (test user) has an obligation to maintain the confidentiality of the test answers, and that the test results cannot be provided to a third party without the applicant's written permission. However, this policy may not always be communicated to those responsible for administering the test or to test applicants. A review of several guides or manuals for test administrators revealed that confidentiality policies were not stated for the test administrator. A review of the applicant agreement forms on several tests indicates that some include a statement that test results will not be revealed without the permission of the applicant, while others do not. Additionally, there may be no review or audit to ensure that these general policies are being complied with by test users.

Interviews with several test publishers revealed that what happens to the completed test booklet depends on how it was scored. If the test was mailed to the test publisher for scoring, then

---

31. See, however, Association of Personnel Test Publishers, Model Guidelines for Preemployment Integrity Testing Programs, 1st ed. (Washington, DC: 1990), p. 15.

32. Committee to Develop Standards for Educational and Psychological Testing of The American Educational Research Association, The American Psychological Association, and The National Council on Measurement in Education, Standards for Educational and Psychological Testing (Washington, DC: American Psychological Association, 1985), Standard 16.3, Primary, pp. 85-86.

the test publisher kept the booklet. if it was scored by the test user, then the booklet was kept by them. It is not known whether test results or test booklets become part of an employee's personnel record.

Related to the question of third-party access to test results is the issue of non-authorized access to those results. This involves safeguards for the security of test results, especially while being communicated through online telecommunication linkages and stored in computerized databases. The Standards for Educational and Psycholoaical Testing state that:

> Test data maintained in data files should be adequately protected from improper disclosure. Use of time-sharing networks, data banks, and other electronic data processing systems should be restricted to situations in which confidentiality can be reasonably assured.[33]

Some integrity test publishers do maintain computerized databases with information on tests that have been administered. In most cases this information is kept for research purposes. In all cases the information is retrievable by an individual identifier -- in some cases not by name, but by social security number.

Thus, with respect to third-party access to test results and security of test administration and results, it appears that appropriate standards exist for integrity test publishers and test users. However, the extent of adherence to these standards is unknown and there is no mechanism to enforce compliance.

INFORMED CONSENT

Basic to the notion of the fairness of a test or test procedures is the principle that the individual should give his or her informed consent to the test. A critical question is what the individual needs to consent to in order for there to be informed consent. Standard 16.1 of the Standards for Educational and Psychological Testing, which established a general policy of requiring informed consent, exempts those situations in which ". . . consent is clearly implied (e. g., application for employment or educational admissions.)"[34] This exemption assumes a broad concept of "implied

33. Ibid., Standard 16.5, Primary, p. 86.
34. Ibid., Standard 16.1, Primary, p. 85.

consent." One testing expert argues that the individual ". . . should certainly be informed about the purpose of the testing, the kinds of data sought, and the use that will be made of the scores."= She recognizes that in order for the test to be effective, the individual should not know the ways in which responses to specific test items will be interpreted, or be shown test items in advance:

> if an examinee is told in advance that a self-report inventory will be scored with a dominance scale, his or her responses are likely to be influenced by stereotyped (and often erroneous) ideas he or she may have about this trait, or by a false or distorted self-concept.*

Broadly consistent with this advice, the integrity test publishers expect test users to inform applicants about the test and its role in the hiring decision.[37] Not as consistent are some test instructions, which tell employers that job applicants are to be told that the purpose of the test is to gather information on various personal qualifications, attitudes, opinions, and background.

A related question is whether the individual has a choice in whether or not to take the test. Although taking the test is technically voluntary, it is probable for most pre-employment screenings that, if an applicant refuses to take an integrity test, his or her chances of getting that job are significantly reduced.

FURTHER THOUGHTS

This report has challenged some of the basic premises underlying the use of integrity tests in the workplace. In particular, OTA found that integrity testing is based on the belief that workplace behavior is determined largely by individual attributes. However, some researchers concerned with management's interest in limiting workplace counterproductivity suggest a broader view: "It is critical for supervisors to appreciate the complex interrelationship between theft and other forms of non-larcenous counterproductive behavior at work. . . . The factors that influence theft are often the same which generate other manifestations of counterproductive activity. . . . This means that theft and

35. Anastasi, op. cit., footnote 26, p. 8.
36. Ibid.
37. Association of Personnel Test Publishers, op. cit., footnote 31, p. 13.

dishonesty are management problems, not solely the concerns of security or law enforcement personnel."=

If these and other social scientists are correct, and dishonest behavior is largely influenced by workplace environments, then predictive validity studies that do not account explicitly for interactions between individual behaviors and environmental factors are an inadequate basis for assessing the utility of integrity tests. While business managers are of course free to make decisions based on any information, they may wish to press test vendors to clarify the limitations of the available research.

An issue that OTA did not address is whether pre-employment screening in general, and integrity tests in particular, are more efficient than increased investments in detection and security. More research would be required to address this question, which is primarily of interest to business establishments weighing their options.[39]

To the extent that problems of cost and relative effectiveness primarily affect private business decisions and productivity, they are not necessarily public policy matters. Presumably, firms considering investments in various alternatives would weigh their costs and benefits. But it is the Government's role to stay aware of the societal consequences of business decisions, and to share information on the potential risks and benefits of various mechanisms marketed as productivity-enhancing tools. OTA believes that the potentially harmful effects of systematic misclassification, possible impacts on protected groups, and privacy implications of integrity tests combine to warrant further governmental attention.

38. Richard C. Hollinger, Dishonesty in the Workplace: A Manager's Guide to Preventing Employee Theft (Park Ridge, IL: London House Press, 1989), p. 36.

39. One question that would need to be considered is the effect of various pre-employment screening devices on the level of effort made by companies to detect theft. If firms using integrity tests become more complacent about their workplace monitoring, more counterproductivity could result.

U S. Government PRINTING OFFICE" 19900-662-250