QUESTIONNAIRES USED IN THE PREDICTION OF TRUSTWORTHINESS IN PRE-EMPLOYMENT SELECTION DECISIONS: AN A.P.A. TASK FORCE REPORT

Lewis R. Goldberg, Chair
Julia Ramos Grenier
Robert M. Guion
Lee B. Sechrest
Hilda Wing

February 1991

The Task Force was aided enormously by Dianne Lane of the A.P.A. Science Directorate, and we are deeply grateful for her many efforts on our behalf. In addition, we would like to thank Lawrence J. Stricker, who served as the initial Task Force Chair from March through September of 1989.
Table of Contents

Background

Charge to the Task Force  3
Procedures Employed  4
Some Initial Problems  5

Assumptions Guiding Our Evaluation  6

Preliminary Issues
The Demand for Honesty Tests  8
Alternatives to Honesty Tests  9
Paper-and-Pencil Questionnaires vs. the Polygraph  11
Published Standards for Tests  12
Utilities and Disutilities  13
What Is the Construct Being Tested?  16
The Problem of Faking  17

Major Considerations and Recommendations  18

Conformity to Testing Standards  18
Adverse Impact  20
Test Validity  20
General Background  20
Variables Used in Studies of Test Validity  21
Studies of Criterion-Related Validity  22
Studies of Construct Validity  24
The Use of Cutting Scores  28
Marketing Honesty Tests  30
The Need for Increased Openness  31

Recommendations for Specific Problems  33

Foreign-Language Translations  33
Language and Cultural Differences  34
Environmental Factors  35
Training Test Users  35
Test-User Qualifications  36
Selection Decision-making  36
Confidentiality, Privacy, and Informed Consent  37

Summary and Conclusions  37

References  40

Appendix A: Charge to the Task Force  45
Appendix B: Questionnaire Sent to Test Publishers  46
Appendix C: Attendees at the Open Meeting  51
Appendix D: Reviewers of the First Draft of this Report  52
Background

Charge to the Task Force

The Task Force on the Prediction of Dishonesty and Theft in Employment Settings, hereafter referred to simply as the Task Force, was charged by the American Psychological Association (A.P.A.) with the task of reviewing currently available commercial tests used for the purpose of assessing the "honesty and integrity" of employees and prospective employees. Specifically, the Task Force was asked to "gather and examine data relevant to the scientific and social-policy considerations associated with the development and use of predictors of dishonesty, theft, and related behaviors." The complete text of the initial charge is provided as Appendix A to this report.

As is often true of such committee charges, this one turned out to be more inclusive than seemed practical. Although the Task Force was guided by the charge, we did not adhere completely to all of its specific terms, since our work resulted in a different formulation of some of the issues. Moreover, the amount and the nature of the information available required us to define some boundaries for our work and for this report. For example, as reflected in its title, this report is focused exclusively on commercially published questionnaires used for pre-employment selection decisions about the trustworthiness (vs. untrustworthiness) of job applicants. It does not consider tests requiring special apparatus or those designed or used to assess current employees. The primary audiences for this report are the A.P.A. committees and divisions who commissioned it, and therefore we have assumed that most readers will have some knowledge of the basic principles of psychological assessment.

In implementing our charge, we have been guided by the realization that
A.P.A. and the profession of psychology generally has an interest in how psychological tests are marketed, used, and perceived. If tests are improperly used, they can bring the field of psychology into disrepute through its failure to apply its own principles, and that is a basis for interest in testing at the aggregate scientific and professional level. Consequently, we have been attentive to a wide range of issues pertaining to the use of tests, especially those that bear on their fairness to individuals.

**Procedures Employed**

The Task Force reviewed a substantial body of information about the theoretical status of such concepts as honesty and trustworthiness, as well as the scientific literature on procedures for the assessment of these concepts. Through literature searches and previous reviews of the literature, we cataloged and appraised over 300 documents that were directly relevant to our charge. Additional information was solicited from relevant test publishers (including their test manuals, test booklets, unpublished research reports, and marketing materials). We also conducted a questionnaire survey of these publishers; a copy of that questionnaire is included here as Appendix B. In addition, we conducted an open meeting to which interested parties, both test publishers and public representatives, were invited; a list of individuals attending the open meeting is provided in Appendix C. On the basis of all of this information, and our own deliberations about its implications, we prepared an initial draft report, which was circulated with a request for suggestions for revision. We received responses on behalf of five organizations--the American Civil Liberties Union, the Association of Personnel Test Publishers, A.P.A. Division 5 (Evaluation, Measurement, and Statistics), A.P.A. Division 14 (Society for Industrial and Organizational
Psychology), and the A.P.A. Committee on Psychological Tests and Assessment—
and from 41 additional individual reviewers. All of these suggestions have
been carefully considered in drafting this final report.

Although our original charge called for the preparation of both a
scientific paper and a policy paper, we have incorporated these two aspects of
our charge in this report. Indeed, because of the availability of three
recent and extensive qualitative reviews of the literature on honesty tests
(O'Bannon, Goldinger, & Appleby, 1989; Sackett, Burris, & Callahan, 1989; and
Sackett & Harris, 1984) and one quantitative meta-analysis of the criterion-
related validity studies (Ones, Viswesvaran, & Schmidt, 1990), we will not
present a paper-by-paper review of this literature. However, it is important
to understand that we have completed our own independent assessment of each of
the primary sources summarized in these reviews, as well as of new material
not summarized there, including some strongly negative reviews. In addition,
we have carefully considered each draft of the report from the Office of
Technology Assessment, including the final version (U.S. Congress, 1990).

Some Initial Problems

We have had to deal from the beginning with the fact that the domain of
interest is ill-defined and heterogeneous. Numerous firms are producing
diverse instruments for somewhat different uses. The terms "honesty" and
"integrity" are widely employed in the literature on these tests, but so are
other terms such as "counterproductive behavior," "dependability," and "job

---

1 Because the class of measures we are reviewing have typically been
referred to as "honesty tests," we will with some hesitation generally
continue this practice. However, this set of measures is quite diverse in
content, including some constructs of greater breadth, and others of lesser
breadth, than that of honesty per se. Moreover, none of the measures we
reviewed provides an assessment of maximum (rather than typical) performance,
as might be required in some definitions of a "test."
performance," most of which can plausibly be subsumed under the broad concept of Trustworthiness. Honesty tests may be distinguished from one another on a variety of grounds, but one polarity of importance is between (a) tests devised specifically for purposes of predicting honesty or integrity in job performance and (b) broad-bandwidth personality inventories adapted to some degree for the same purposes. The latter may have special keys for scoring new scales or may have formulae for combining existing scales into composite prediction scores. Broad-bandwidth inventories, even when marketed for use in personnel selection, are not likely to be called honesty tests, and therefore we have elected to omit them from our charge.

A considerable amount of evidence has been produced relevant to the reliability, validity, and utility of a few honesty tests, and none for others. Whether the latter tests can ride on the coattails of research on the former is not a matter easily decided. A few tests have adequate documentation; others have virtually none. Promotional materials for some tests seem straightforward and acceptable; for others they are highly questionable, if not expressly fraudulent. Some of the persons employed by honesty-test publishers are members of A.P.A. and subject to its influence; others are not. Thus, this report can set forth few statements of blanket approbation or disapprobation. Readers must be prepared to deal with some ambiguity.

Assumptions Guiding Our Evaluation

We have been guided by the principle that, absent a strong argument to the contrary, the evaluation of honesty tests should not be carried out using criteria more stringent than those that would apply to other types of tests used for making decisions about people. Correlations of the order of .30 or
.40 are regularly accepted as evidence for the validity of measures in these domains, and we saw no reason to belittle such correlations for honesty tests. In other domains, a wide variety of methods are used to provide converging evidence of test validity, methods that we could find no reason to exclude for honesty tests. As an extreme example, evidence for the validity of honesty tests has been adduced from the finding that such tests will distinguish convicted thieves from members of the general population. Although such findings have met with some derision, the method of contrasted groups has a long history of use in test validation. (Certainly if honesty tests did not distinguish convicted thieves from members of the general population, their construct validity would be regarded as highly suspect.)

Perhaps the most common type of evidence for the validity of honesty tests is the demonstration that their scores are related to self-reported offenses, usually obtained under conditions of anonymity or in connection with polygraph examinations. That kind of evidence is sometimes deprecated as if it were trivial and too obvious to be of value. Yet, self-reported involvement in criminal activities is widely regarded as one of the most useful measures of those activities in criminological/sociological studies. Certainly, it would be unlikely that a test that predicted other types of self-reported behaviors (e.g., admissions of racial or sexual discrimination) would be rejected as worthless.

In addition, we assume that all tests should be evaluated, especially with respect to their construct validity, not on the basis of any one study or set of studies, but as the consequence of a pattern of converging evidence from an extensive set of findings. That is the way we believe that the construct validity of honesty tests must be judged. Similarly, with respect
to criterion-related validity, we believe that converging evidence as may be
revealed by sophisticated meta-analyses using a variety of criterion indices
(e.g., Ones et al., 1990) is more persuasive than the findings from isolated
studies.

Finally, and perhaps of the greatest importance, paper-and-pencil honesty
tests should not be evaluated against absolute levels of validity; rather they
must be evaluated comparatively against the validities of other procedures
that would inevitably be used in their stead (e.g., unstructured interviews,
handwriting analyses). For any potential problem with honesty tests, one must
determine the extent to which alternative procedures used for the same purpose
would be similarly indicted. In essence, one must always keep salient the
question, "What would you have them do instead?"

Preliminary Issues

The Demand for Honesty Tests

Probably no one doubts that dishonesty, especially theft, is a serious
problem in some businesses. The hard data available are not extensive, but
they have been widely disseminated. One estimate, from the American
Management Association, is that in 1975 dollars the total cost to business of
crimes other than violence was $40 billion (cited in Clark & Hollinger, 1983).
A survey conducted by an honesty-test publisher found that 62 percent of the
employees in the fast-food industry admitted to stealing company property or
cash (Slora, 1989a); for some other estimates, see McGough and Brown (1990).
We do not vouch for the accuracy of such data, but we understand its citation
as justification for the existence of pre-employment honesty testing, and we
do not question that the problem is a large one. Even if the estimate of loss
due to theft were halved, it would still be an impressive figure. The
justification for honesty testing, then, is based on the premise that test use will result in reduced employee theft.

Certainly these tests are used (O'Bannon et al., 1989; Sackett & Harris, 1985). The honesty-testing industry, using the term broadly, is a large one. By one estimate already several years old (Sackett & Harris, 1985), as many as 5,000 companies use honesty tests in personnel selection, and as many as 5,000,000 people may be tested each year. Obviously, many individuals are at risk for having their employment prospects affected by what they do on a paper-and-pencil honesty test.

Moreover, the use of these tests may increase with the growing concern about the "negligent hiring" doctrine in the courts. At the present time, an employer may be held liable if it can be shown that (a) harm to coworkers, customers, or other third parties has resulted from the actions of an "unfit" employee, (b) the employer could reasonably have foreseen the risk of hiring such an unfit person, and (c) the employer failed to conduct a reasonable inquiry into the employee's background and fitness. Although most negligent hiring cases have arisen from physical harm or sexual misadventures, a more general untrustworthiness (including theft) may constitute a basis for action (e.g., Shattuck, 1989). Honesty test use has been identified, under some legal safeguards, as one form of protection against such charges.

Alternatives to Honesty Tests

We assume that if employers believe that employee theft is a problem, they will take steps to do something about it. If they do not use honesty tests, presumably they will do something else. For some employers, particularly those for whom theft is not a major factor and who are not in a highly competitive segment of business, employee theft could be regarded as
nothing more than part of the cost of doing business, with that cost being passed on to customers. Although such a solution may be relatively painless in the short run, it may be costly in a longer perspective, as well as being unfair to those who must foot the bill for others' dishonesty. Moreover, that solution is not available to businesses that have larger losses and/or that face strong competition.

On the other hand, there are procedures that should always be attempted before pre-employment screening is undertaken. Managers should make clear to employees that they are concerned about theft and that they will do something about it if it is detected. They should also institute reasonable managerial practices that tend to put a damper on theft, such as adequate inventory control, frequent cash register cash-outs, and avoidance of those problems that might lead to employee discontent and cynicism. Indeed, to the extent to which management can elicit feelings of ownership from the employees, theft may be construed as stealing from oneself, and thus be reduced (see Greenberg, 1990). However, if firms have already instituted such practices and employee theft remains a problem, then honesty-test screening may be preferable to the alternatives.

One such alternative is on-the-job surveillance. Surveillance is typically expensive, and it is often offensive. One extreme of surveillance is that used in diamond mines, which, however well justified, one finds discomfiting. A lesser example, but one that illustrates the problems with surveillance, are retail firms, often large discount stores, that will not permit their cashiers to void any transaction, approve any check, or make any other complex move without oversight by the store manager. Replacing honesty-test screening with closer surveillance might not make anyone much happier.
Certainly a need exists for effective methods of screening job applicants to identify persons who are likely to be untrustworthy. Responses to paper-and-pencil tests, although commonly used, seem quite remote from the specific acts that they are supposed to predict. However, the discovery of better screening methods is not likely to be a simple matter. Interviews can be expensive, and the typical unstructured interview is neither reliable nor valid. Background checks are even more expensive. Reference checks are becoming increasingly futile as employers, responding to fears of litigation, refuse to provide any information beyond verification of dates of employment. "Work samples" that provide applicants with the opportunity to steal or cheat would border on entrapment, even if they were otherwise feasible. Structured telephone interviews, or interviews conducted by computer, obviate some forms of interviewer bias, but it is difficult to imagine that the questions actually asked are much different from those used in honesty tests.

**Paper-and-Pencil Questionnaires vs. the Polygraph**

In the past, some firms were persuaded to use polygraphic interrogation for pre-employment screening, a practice that is now legally prohibited for most purposes. The results from a polygraphic examination differ from the scores on a paper-and-pencil honesty test in a host of ways: Compared to honesty tests, polygraphic examinations are (a) physically restraining and intimidating, (b) typically subjectively scored and therefore substantially less objective and reliable, (c) typically interpreted without reference to normative standards, and (d) far more prone to result in a stigmatizing label on those who fail them. In addition, honesty tests provide a continuous range of scores, rather than the discrete categories normally used in polygraphic examinations. Moreover, honesty tests benefit technically from the
accumulated knowledge of test construction acquired from decades of scientific studies of psychometric procedures. It is important to realize that any prohibition of honesty-test screening, unlike the prohibition of the polygraph, could create a niche elsewhere that might be beyond professional and scientific review altogether.

Published Standards for Tests

The Standards for Educational and Psychological Testing (1985) prepared for the American Educational Research Association, the American Psychological Association, and the National Council on Measurement in Education should apply to honesty tests, as should the Principles for the Validation and Use of Personnel Selection Procedures (Society for Industrial and Organizational Psychology, 1987). We will discuss the applicability of the Standards at a later point, but we need to point out in advance that one of the guidelines listed in the Preface to the Standards is a potentially persistent source of difficulty in the honesty-testing industry. Specifically, the Standards require that "test developers, publishers, and users collect and make available sufficient information to enable a qualified reviewer to determine whether applicable standards were met" (p. v).

Among the set of published honesty tests, many are "proprietary" measures. Generally, proprietary implies protection from inspection. Most honesty-test publishers are wary of publishing information concerning the development of their tests, their structure and content, much less the actual test items and scoring keys. We understand the commercial problems, but to the extent to which the publishers make scientific claims, their tests must be judged by the applicable scientific standards. Indeed, the published Standards do not distinguish between proprietary and non-proprietary tests.
To their credit, a few firms have published a number of reports of studies bearing on the reliability and validity of their instruments, but most firms have produced nothing of this sort. Certainly the information available even from the most open of the publishers would not permit an outside reviewer to judge whether the use of their test meets some specific Standards.

Utilities and Disutilities

When a job applicant is correctly identified by an honesty test as likely to be trustworthy and, in part for that reason, is hired, the employer gains a trustworthy employee at relatively low cost. Similarly, when the test correctly identifies an applicant as untrustworthy and the applicant is not hired, the employer may have averted considerable loss, again at a low cost. To the extent that a test helps make decisions that result in such gains, the test is said to have utility.

The idea of utility is reasonably straightforward, although not without complexities. For example, if the manager of a fast-food store concludes from a test score that an applicant is a potential thief, but only because of eating an occasional sandwich without paying for it, the savings may be small. Moreover, if the test serves to reject applicants who steal small amounts but are particularly good workers in other respects (e.g., skills, imagination, energy), net costs may be incurred.

Nonetheless, most criticisms of honesty tests have centered on the disutility called "false positives," namely those persons incorrectly predicted to be potential thieves. The concepts of "positives" and "negatives," and of "true" and "false" categories of each, are based on the unnecessary and unwise assumption of fixed cutting scores on both the test score distribution and the criterion. In effect, both of these continuous
distributions are rescored as dichotomies, and all scores on the same side of a cutting point are treated as if they were identical.

The major source of societal concern, and it is a proper one, is for job applicants who are rejected on the basis of honesty-test scores as potential risks but who would have been trustworthy employees if hired. This false-positive problem seems to involve at least two separate issues, error and labeling. In the first of these issues, it seems unfair if by error two equally trustworthy applicants are placed on different sides of the cutting point, one being hired and one not on the basis of a fallible test score. That problem is, of course, not peculiar to honesty testing. Indeed, the same objection can be raised about any other selection instrument, such as an aptitude test, an interview, or an educational requirement. All are fallible to some degree. Moreover, the problem is probably not any greater in the case of honesty testing; the evidence concerning the validity of honesty tests (which will be discussed later in this report) is about as adequate as that available for most instruments used in employee selection.

As long as the number of potential job applicants is greater than the number of job openings, selection of some nature will be required. Moreover, any fallible selection procedure will result in potentially worthy applicants being rejected. However, any valid selection device will result in fewer false-positive errors than a random or quasi-random procedure such as "first come, first served." For a more detailed presentation of this general principle, see Landy (1985, pp. 211-213) and Martin and Terris (1990).

Because the problem of error is not appreciably different for honesty tests than for other selection instruments, the false-positive arguments against them must rest heavily on the presumed consequences of "failing" an
honesty test and the resultant stigma of being labeled as dishonest. The response of some test publishers to this problem is that (a) applicants do not "fail" an honesty test, and (b) applicants never know why they are not hired. The first part of the response is questionable. The best of the honesty-test manuals do warn that test scores should not be used alone to determine employment. Nonetheless, other practices of test publishers serve to negate this recommendation. First of all, some test manuals, and some research reports, directly refer to "failing" the test. Moreover, most honesty-test publishers recommend the use of cutting scores on their tests, a practice that at least implicitly leads to test "failures." Indeed, one could argue that publishers' advocacy of cutting scores is directly responsible for the "stigma" criticism.

In addition to the problem of labeling, the problem of prediction errors is exacerbated by the use of cutting scores. Specifically, the incidence of false-positive errors is highly influenced by the point on the distribution where the cut is made. Prediction errors for scores close to the cutting point will differ from those far removed from it. The incidence of errors is reduced by recognizing the continuous nature of both the test-score and the criterion distributions. If decisions are based on degree of risk, rather than on category designation, these problems become much less severe.

All of the evidence reviewed by the Task Force suggests that the relations between honesty test scores and criterion behaviors are continuous and monotonic. Because there is no evidence of any "natural" cutting score on any of these tests, any preset cutting score is arbitrary. Worse, the use of any preset cutting score serves to negate the stricture that the test not be used alone, but rather that its scores should be considered along with other
predictor information. That is, the advocacy of a cutting score serves to prevent the development of compensatory selection strategies (such as those based on multiple-regression analyses or even subjective compensatory decisions), and to encourage a multiple-hurdle (or worse a single-hurdle) strategy of personnel selection (see Wiggins, 1973/1988; p. 234).

When cutting scores are not used and honesty test scores are actually combined with other predictor information, then it may be correct to say that applicants do not "fail" the tests. We know of no comprehensive surveys of the ways that honesty tests are used, but in most circumstances selection is probably based on several procedures (e.g., application blanks, ability tests, interviews), any one or combination of which could have been the cause for rejection in a given case. Moreover, it probably is true that most applicants are not, for legal reasons, told why they are not hired. If they are not told and if their records are kept confidential, as they certainly should be, then no labeling effect is likely.

What Is the Construct Being Tested?

As already noted, honesty tests are diverse in their nature, and it is difficult to extract a single construct that is invariant across the large and varied set of tests that we reviewed. "Honesty," "integrity," "theft proneness," "dependability," "counterproductivity," and "antisocial tendencies" are among the terms we encountered. The prediction of substance-abuse problems at work is also an aim of several tests. Reference is even made to "theft of time," and one study proposed as evidence for the validity of an honesty test a relation between test scores and the self-reported tendencies of nurses to extend their break periods. Some test publishers specifically disavow attempts to measure honesty broadly, suggesting that they
are only trying to predict problems on the job, such as theft.

If the stated purpose of a test were limited to predicting the likelihood that an employee would commit a theft, the problem of defining the construct would be much diminished. To the extent, however, that terms such as honesty and integrity are used in describing the tests, those constructs deserve definition. The definitions of such constructs found in manuals and other documents are rudimentary at best. That deficiency does not sharply distinguish honesty tests from a great many others on the market, but it does not positively commend them either. We will have more to say about this matter when we consider issues relating to construct validity later in the report.

The Problem of Faking

A common reaction to paper-and-pencil honesty tests is one of incredulity that they should work at all. Most of them may seem quite transparent, and it seems obvious to many observers that job applicants would not willingly report undesirable behaviors that would ruin their chances for employment; surely they would lie. Although we would not like to discourage continued research on dissimulation, faking may not be as great a problem as most people fear (e.g., Hough, Eaton, Dunnette, Kamp, & McCloy, 1990). Some applicants may not regard the tests as very important, and hence are not motivated to lie to any great extent; others may think they can "outsmart" the tests by admitting to various transgressions (e.g., Cunningham, 1990). Still others may believe that companies can and will check up on what they say and therefore they should be truthful. Finally, at a more theoretical level, we know that people tend to assume that others are much the same as themselves (e.g., Gilovich, 1990; Ross, Greene, & House, 1977); to the extent to which less honest
applicants succumb to this "false consensus" effect, they will assume that their anti-social attitudes and behaviors are quite normal, and therefore they can express them freely.

In general, then, it is possible to achieve undesirable scores on honesty tests for a variety of reasons, just as it is possible to achieve desirable scores by either honest or dishonest means. Although undesirable test scores have been shown to be statistically associated with undesirable criterion behaviors such as theft, it should be clear that this association may be determined by a complex pattern of as yet poorly understood factors. For this reason, we strongly urge the test publishers, as well as independent investigators in personality assessment, to carry out some empirical explorations of the meaning of honesty-test scores. In the interim, however, it is important to bear in mind the fact that successful faking will increase the likelihood of errors in predicting trustworthy employees, the opposite of the "false positive" problem. Faking, therefore, relates primarily to the usefulness of these tests to employers. Presumably, if faking becomes widespread, they will discontinue their use of the tests.

**Major Considerations and Recommendations**

**Conformity to Testing Standards**

As we have already noted, the Standards for Educational and Psychological Testing are applicable to honesty tests. Individual test publishers may not maintain links to any of the three parent organizations, and they are therefore beyond their direct purview in enforcement of the Standards. Nevertheless, we believe that the three organizations individually and collectively have an obligation to judge such tests according to the Standards. Indeed, at least one test publisher has examined both the
Standards and the Principles for the Validation and Use of Personnel Selection Procedures (Society for Industrial and Organizational Psychology, 1987), and evaluated its test against both sets of standards (Jones, 1989; Slora, 1989b). Although we will not evaluate their claims of compliance, we strongly recommend that such analyses be undertaken by all publishers.

Honesty tests vary enormously in their compliance with standards regarding validity. The "manuals" for some of them are so deficient that they obviously would meet none of the standards; others are much more informative. Some validity studies are well planned and implemented, whereas many are not. Most standards for reliability have been more uniformly met, but standard errors of measurement have rarely been reported, especially in relation to the use of "pass-fail" cutting scores and scores demarcating "zones" such as "at risk." Finally, standards having to do with norms have typically been treated lightly, if at all. For example, honesty-test manuals typically provide no information about how so-called risk zones were defined; normative samples are rarely described well enough to make it possible for a potential user even to know the shape of the distribution, much less the composition of the sample with whom the applicants will be compared.

A group of test publishers has, under the auspices of its own organization, the Association of Personnel Test Publishers (A.P.T.P.), published the Model Guidelines for Preemployment Integrity Testing Programs. This effort is laudable, especially to the extent that the new Guidelines elaborate the A.P.A. Standards and add specifically relevant issues. The Guidelines distinguish between published tests that are (a) nonproprietary, for which responsibility for appropriate use falls on the user, and (b) proprietary, which remain under the control of their publishers. According to
the Guidelines, publishers are expected to exercise control over the distribution of nonproprietary tests, but users are responsible for the demonstration of their appropriate use, including the collection of evidence concerning their reliability and validity. In the case of proprietary tests, users purchase the services of the test publisher, who is expected to have the technical expertise necessary for proper test construction, development, validation, and marketing.

The Guidelines include a parallel set of standards for test users, but the procedures for adopting these standards, as well as for enforcing them, are not provided. That is, although the efforts to develop user guidelines are admirable, it is not yet clear how such guidelines can be made to stick. Ideally, test publishers could refuse to sell tests to users who are not willing to abide by the guidelines.

Adverse Impact

Certainly one of the most important issues related to any selection instrument concerns the impact of its use on legally protected groups. Many honesty-test publishers have considered the impact of their tests on women and minority groups, and studies comparing the test scores of different groups suggest that these tests have no adverse impact (e.g., Arnold, 1989; Bagus, 1988; Cherrington, 1989; Terris & Jones, 1982). As with all such issues, however, continued attention is warranted, and individual employers retain the obligation to check for adverse impact in their selection procedures.

Test Validity

General Background. Of the dozens of honesty tests now being marketed, for only a few was research-based information provided to the Task Force. For those few tests, the reported psychometric properties were typically limited
to reliability, usually internal consistency, and validity in the form of correlations with various criteria. Rarely could we find an item analysis with item-total correlations, and we found no reports that showed the kinds of items that were eliminated from the initial pool during the process of test development. Only a few of the tests have been factor-analyzed or in any other way examined for internal structure. As expected, when studied in this way, honesty tests do not turn out to be unidimensional (e.g., Cunningham & Ash, 1988; Harris & Sackett, 1987), which does not differentiate them from other personality scales used in personnel selection. In general, such psychometric data as are reported tend to be satisfactory. For example, internal consistency tends to be quite adequate, and test-retest reliability also appears to be high for those instruments on which it has been examined.

Variables Used in Studies of Test Validity. Validity studies require the specification of some criteria to be related to the characteristics being assessed. Validity studies for honesty tests have tended to rely on seven basic classes of such variables:

1. On-the-job critical incidents, whether of theft, absenteeism, or some other problem. For example, in one study the investigators located employees who had taken an honesty test as job applicants and who later had engaged in some counter-productive behavior that got them discharged; their test scores were compared to those from a normative sample.

2. General performance on the job. For example, department-store employees screened by an honesty test were compared with those not so screened on self-reported rule-breaking on the job, job satisfaction, and overall job stress.

3. Contrasted groups. For example, the attitudes of convicted felons
have been compared with those of nonfelons in regard to theft, violence, and illicit drug use.

4. Comparison with polygraph performance. Because some businesses that previously used the polygraph for pre-employment screening now use honesty tests for the same purpose, it is probably only natural that one form of test validation is the demonstration that the two procedures produce similar results.

5. Induced confessions. Yet another method of validating honesty tests has been to provide the opportunity for individuals to confess to transgressions—e.g., theft, rule breaking, cheating, drug use—and to determine whether such confessions are more likely from persons with low scores on honesty tests. The most common method has been simply to use the promise of anonymity to induce subjects to confess to incidents of theft, alcohol abuse, violence, or industrial damage and waste.

6. Correlations with other tests. Honesty test scores have been correlated with scales from broad-bandwidth personality inventories, as well as with scales purportedly measuring more narrow constructs, such as child-abuse potential.

7. Company losses from inventory. An important criterion from an employer’s point of view is whether screening with an honesty test actually results in a decrease in undesirable behaviors, such as theft. Some studies have examined the impact of honesty screening on inventory "shrinkage."

Studies of Criterion-Related Validity. Two outcomes seem important

---

2 Modern views of validity are stated in terms of different kinds of evidence for the validity of an inference from a test score, rather than in terms of different kinds of validity per se; however, for ease of exposition we will use the more conventional terminology in this report.
from the standpoint of users of honesty tests. First, an employer might be interested in the extent to which the use of honesty tests as part of the selection process will lead to a generally improved workforce. That result might not be specific to honesty tests; insofar as tests assess other job-relevant characteristics such as emotional stability, they might still improve a workforce, and an employer might not be concerned about what the tests were called if they served to produce desirable results. Few studies have used any measures of overall job performance, but Terris and Jones (1984) reported that an honesty-test-screened group of employees had better records and greater job satisfaction. Additionally, a multi-faceted test developed for use in law enforcement has been shown to predict the performance of police trainees and officers over a wide enough range of behaviors to regard it as having value for improving the police workforce (e.g., Shusman, Inwald, & Knatz, 1987).

Much more evidence exists concerning the usefulness of honesty tests and related measures to predict the occurrence of critical events such as theft, drug and alcohol problems, and absenteeism. In their review, O'Bannon et al. (1989) described predictive-validity findings for six different honesty tests and concluded that (a) few predictive studies are free of methodological difficulties and (b) most tests have not been used in predictive studies. We concur with those judgments. However, the evidence for a few tests is reasonably extensive and consistent in a way that would be difficult to account for solely by methodological artifacts.

A third criterion for honesty tests is the "bottom-line" assessment of the extent to which losses owing to theft (shrinkage) are actually decreased. O'Bannon, et al. (1989) described a number of time-series studies carried out on one or another of four tests. Again, although each study is flawed in some
way, they are not all flawed in the same ways, and the findings are reasonably consistent. (In at least one study, the effect of testing was too quick to have been attributable to changes in the workforce, suggesting that one possible mechanism for the influence of honesty tests stems from the knowledge by employees that management is concerned about employee theft.) The conclusion of O'Bannon, et al. (1989) is sufficiently conservative to be acceptable: These studies "... do begin to establish a foundation of evidence which may become more convincing as additional studies accumulate." We strongly recommend that those test publishers who have established this initial foundation of evidence continue their efforts, and that the remaining publishers begin to establish some empirical foundations for their own tests.

Meta-analysis provides a quantitative summary of the existing research on a particular topic. Ones, et al. (1990) have reported such a meta-analysis of the criterion-related validity coefficients of honesty tests. In their analysis, they calculated mean validity coefficients that varied from .19 to .61 for various criterion indices, and from .27 to .71 for various tests and combinations thereof; when corrected for sampling error, the corresponding mean values ranged from .26 to .76, and from .36 to .87. Moderator analyses suggested that the magnitude of the validity coefficients was significantly related to types of tests and types of criteria. If one accepts the assumptions underlying modern meta-analytic methods, then the general usefulness of honesty tests appears to have been established.

Studies of Construct Validity. Establishing the construct validity of a measure is a complex enterprise. The complexity is not reduced if the construct is as ill-defined as are the concepts of "honesty" and "integrity." Test publishers differ radically in the ways that they treat these concepts.
Some state straightforwardly that their test is a measure of honesty, whereas others deny that their test measures such a construct at all. A common claim is that the test helps to locate potential employees who are high risks for dishonest or counterproductive behavior without specifying on just what grounds that is done.

Indeed, some of these tests seem to perform in the way one would expect of tests measuring a construct like "theft potential." People known to be dishonest (convicted thieves) do badly on them in comparison to the general population; males do worse than females; and younger people do worse than older people. None of these findings is a positive indication of theft potential, but all are probably required by our knowledge of that construct. People who admit to dishonesty look bad on honesty tests, as do people who are caught stealing on the job. When apparently dishonest people are removed from the pool of employees, losses from theft tend to decline. Moreover, one of the better analyses of an honesty test, based on Item Response Theory (IRT), found that most items fit reasonably well the specifications of the IRT model (Harris & Sackett, 1987).

Because these tests are so heterogeneous in focus and content, however, it is not possible to assess their construct validity as a class. Obviously, theft potential is not all that these tests measure. Some of them probably tap more general constructs such as dependability, responsibility, and conventional morality—all of which may be subsumed under the broad concept of trustworthiness. On the other hand, it is quite possible that these tests predict more validly at one end of the score continuum than at the other; for example, it may be the case that scores located on the untrustworthy half of the scale tend to be more valid than those from the trustworthy half, a
finding that has characterized some other measures of deviant behavior (e.g., Fisher, 1959). As a consequence, we strongly recommend that investigators devote increased research attention to such construct-oriented issues (e.g., Woolley & Hakstian, 1990).

Although the terms "honesty" and "integrity" have broad and highly desirable meanings in the lay community, they are used in the present context to refer primarily to the theft of employers' property. Whether problems such as on-the-job interpersonal problems, use of alcohol or drugs, absenteeism, and extension of break periods have anything to do with these concepts is not clear. Indeed, if the term "counterproductivity" had initially been employed as a designation of these tests, less emotional heat might have been generated. However, many test publishers may be reluctant to adopt such a term, perhaps because it does not have the marketing value of such terms as "honesty" and "integrity."

In any case, each test publisher should be prepared to define the construct measured by that test and then to assemble and present the evidence bearing on its construct validity. Proper evidence would begin with appropriate psychometric studies of the instruments, including their dimensionality. Item analyses should be carried out routinely and at least the aggregate results reported. Probably detailed results could be made available for inspection by interested persons under fair conditions, while still protecting truly proprietary rights. Evidence related to convergent and discriminant validity should be developed and be made available. Correlations with other measures should be regularly sought and reported.

---

3 If our hypothesis is correct, this report should more accurately (but perhaps more awkwardly) be titled "Questionnaires used in the prediction of untrustworthiness in pre-employment selection decisions."
Moreover, honesty-test publishers need to give more thought to the appropriate criteria by which to validate their tests. Relations between the various criteria need to be studied as well, in order to understand better what is involved in employee honesty. For example, is it relevant to a service industry that honesty-test screening reduces inventory shrinkage in retail stores? Are potentially valuable employees not being hired because they admit they might steal or consume things of small value at their jobs? Are activities such as the theft of goods, cheating on time, and using a company vehicle for personal errands simply different manifestations of the same latent disposition? Will it be necessary to validate honesty tests separately against each criterion (e.g., stealing goods, stealing cash, cheating on breaks, lying about illness)? Additionally, are honesty tests equally valid across the entire range of industries and individual businesses? Does it not seem likely that the validity of an honesty test might depend on the inherent risks involved in stealing, as well as on the security measures in operation in a particular business? It is the responsibility of honesty-test publishers to set forth a rationale for the use of their tests in these diverse contexts.

Finally, evidence of construct validity should include attempts to rule out alternative kinds of individual differences. As we have already noted, existing research suggests that honesty test scores are not related to various racial groups, and other research suggests that such scores are not related to educational level and ability (e.g., Werner, Jones, & Steffy, 1989). However, this type of discriminant validation has received only superficial attention. Measures of constructs that are conceptually related to honesty and integrity should be included in comparative-validity studies, so as to discover those
facets of the total domain of Conscientiousness (e.g., Goldberg, 1990) that are most highly related to various kinds of employee misbehavior.

The Use of Cutting Scores

Many test manuals warn that, for optimal decision-making, scores on honesty tests are not to be used alone, but rather that they are to be combined with other information. Nonetheless, cutting scores of one sort or another are usually provided, whether to divide applicants into two (e.g., pass vs. fail) or three (e.g., high, medium, and low risk) categories. Given the proclivities of overloaded decision makers to arrive at decisions through elimination by aspects (Tversky, 1972), it is probably inevitable that such cutting scores will be used in an absolute way. Honesty-test publishers perpetuate this problem by their frequent references to "failing" and "passing" categories of applicants. That is, although test publishers deny that anyone fails an honesty test, their own publications belie that claim.

Little information is available about the way in which cutting scores have been set, and what is available is disturbing. For example, the cutting scores for some tests appear to have been set purely arbitrarily, on the grounds that the scores chosen seemed about right. Moreover, there is little evidence of empirical work on cutting scores so as to produce the largest proportion of successful employees. The classic references on the theoretical development of optimal cutting scores (e.g., Rorer, Hoffman, & Hsieh, 1966; Rorer, Hoffman, LaForge, & Hsieh, 1966) are not cited in honesty-test manuals, nor are more recent review articles on this topic (e.g., Cascio, Alexander, & Barrett, 1988).

There are a number of psychometric problems associated with the use of any cutting scores. One obvious problem stems from the enormous loss in
information when a continuous distribution of test scores is compressed into a mere two values (Cohen, 1983). Indeed, the two values constitute a binary scale, and it is that truncated scale, rather than the continuous scale, that must be validated against the criterion of interest. In addition, there are problems that stem from the inherent unreliability of any single test score. The conventional indices of test reliability do not necessarily tell much about the likely range of error of estimating a true score relative to any given observed score. The standard error of measurement is the appropriate statistic for that purpose, but estimates of standard errors are rarely reported. Often they cannot even be calculated because standard deviations are not routinely reported in test manuals.4

If test users are to make effective use of cutting scores, then they need to know a great deal more about them than is revealed in most of the test manuals. They need to know something about decision aids such as receiver-operating-curves so that they can see what it costs in potentially successful employees to exclude each potentially unsuccessful one. Cutting scores and their utility depend greatly on the actual incidence of theft (or other critical behaviors) in a particular firm. Test publishers need to provide clear and detailed information about just how cutting scores or boundaries between "zones" are derived (e.g., whether they are empirical, rational, or purely arbitrary). They also need to provide evidence whether the same

4 Ideally, employers should be taught to use prediction equations to combine information in optimal ways. If cutting scores have to be used at all, then the more categories provided the user, the better. That is, categories of high, medium, and low risk are preferable to categories of passing and failing. However, an even better practice would be the provision of overlapping ranges of scores, with specific suggestions about how the scores within each category are to be used in conjunction with other information.
cutting scores will apply to all different businesses and individual companies. For example, is a "high-risk" applicant in the convenience-store business also a high-risk applicant in banks and stock-brokerage firms?

**Marketing Honesty Tests**

Promotional claims for honesty tests, as perhaps for most other procedures used for pre-employment screening, vary from the circumspect to the fraudulent. We have seen a number of promotional brochures that are so clearly excessive and overblown as to make a test expert cringe in embarrassment. In the most flagrantly hucksterish of these, all problems associated with test use are unmentioned, and the purported reduction in actual theft that can be achieved is wildly exaggerated. We recommend that test publishers adopt and enforce standards ensuring that the promotional claims made by each testing organization rest on a firm empirical foundation. This includes expending company resources to train sales representatives about the statements and claims that are appropriate for each test, and to monitor the performance of these personnel.

Moreover, each test publishing company should carefully consider the aims for each of its tests, determine which aims it can support by what evidence, and then examine its promotional material to make sure that its claims are justified. Our position on this issue is determined less by a concern that businesses need to be protected from the testing industry than by more basic considerations of professional standards and the long-run interests of the public in ensuring equitable treatment of its citizens.

In addition, we recommend that A.P.A. publish a set of non-technical guidelines for users of all kinds of pre-employment screening tests, to assist potential users in evaluating claims made on behalf of competing tests. Such
guidelines should provide a realistic appreciation of test limitations, and permit users to distinguish marketing claims held to scientific standards from those that are exaggerated. A useful resource for the development of such a document is the recent report on test-user qualifications prepared by the Joint Committee on Testing Practices (Eyde, Moreland, Robertson, Primoff, & Most, 1988).

The Need for Increased Openness

One serious problem in evaluating proprietary tests is the realization that publishers may have no interest in making negative information available. As we have already noted, we expect that at least some honesty tests would meet reasonable standards of reliability and validity, in part because they capitalize on the long history of test construction in the fields of psychology and education. Nonetheless, we strongly recommend that ways be developed to open up the processes of evaluating honesty tests. Ideally, the test publishers might fund one or more independent test-evaluation organizations. At the very least, test publishers should commit themselves to making available results of all, even negative, research on these tests. Indeed, the A.P.T.P. Guideline 1.c. states in part that test publishers should "Accurately represent research on the psychometric properties (e.g., reliability, validity, fairness, utility) of the integrity test" (p. 9). We believe that the only way to meet that standard is to ensure that all research is reported.

Indeed, at the moment gaining access to research on honesty tests (and many other types of selection procedures) is not always easy. A great deal of it is available only in the form of technical reports from test publishers. It is not reassuring to find references in promotional materials to what
"research shows" without any further information about the nature of that research or the citation of publicly available sources where such information can be obtained. Certainly only a tiny fraction of the research on honesty tests has been published in refereed journals. To their credit, a few publishers do make their research available through technical reports that presumably are available for the asking. For example, the Task Force received over 100 reports from a single publisher. On the other hand, we do not know how easily others (e.g., graduate students) might secure such information, without the clout of A.P.A. behind them.

Commercially published tests are reviewed regularly in the Mental Measurements Yearbooks (e.g., Conoley & Kramer, 1989) and in Test Critiques (Keyser & Sweetland, 1984); these are important sources of information for potential users of tests, as well as for those who might want to do research on them. It is therefore important that all relevant information be made available to test reviewers, including normative data and information about test-construction procedures. We strongly recommend that honesty-test publishers do all in their power to ensure that their tests receive adequate reviews. Any resistance to appropriate peer review of tests is a professional dereliction, and one of considerable consequence. If test publishers wish to maintain their identity with the field of psychological measurement, they must be professionally responsible and accountable, whatever their proprietary interests.

---

5 Test publishers complain, with some justification, that validity studies are difficult to publish these days because of the editorial policies of most journals. However, some honesty-testing investigators have managed to publish reports of their research in present journals. Moreover, an increase in the number of scientifically adequate studies should serve to stimulate the development of new journals.
In general, honesty testing as a scientific specialty and a professional application would be much advanced if opportunities for research were opened up considerably. We believe that safeguards could be developed that would enable graduate students to carry out thesis and dissertation research on these tests, a provision that would surely increase the quantity and scope of research. Independent researchers should, under appropriate arrangements, be given access to these tests for use in studies of honesty and related concepts, and test publishers could also offer to make data bases available for secondary or meta-analyses.

At the moment, most studies of the validity of honesty tests include only a single instrument. Nonetheless, studies of the comparative validity of different honesty tests are ultimately of considerable interest if the scientific basis of the industry is to become more sophisticated over time. Although industry representatives are not keen on the idea of such comparative studies, the day of that sort of accountability seems upon us. With products from cola drinks to luxury automobiles being pitted against each other by their own manufacturers, a defensively negative position resting on "proprietary" arguments does not persuade.

**Recommendations For Specific Problems**

**Foreign-Language Translations**

If honesty tests are to be used internationally, or nationally for entry-level jobs such as in the fast-food industry, they will have to be made available in other languages, particularly in Spanish. The literal translation of English-language questionnaires is not recommended, given the substantial differences in substantive nuance, social desirability, and reading level inherent in the translation process. Methods of back-
translation are neither cheap nor easy, nor do they assure psychometric equivalence. We have had an opportunity to review one Spanish-language translated version of an honesty test and were not impressed with the quality of the translation. Indeed, some of the translated items seem unlikely to elicit the same information as the original English items. We strongly recommend that (a) the equivalence of any test provided in two languages be demonstrated, and (b) such a test not be distributed for use in making selection decisions prior to the determination of its reliability and validity when it is used in each language. Although this is a costly undertaking for test publishers, it is the only fair procedure, and in the long run it may prove less costly than the litigation that could result from problems related to adverse impact and erroneous classifications.

Language and Cultural Differences

Even with the original English-language versions, careful research attention needs to be paid to the nature and importance of differences in the responses of various linguistic and cultural groups to honesty tests. For example, some subcultures may dispose their members to be open and even boastful about transgressions that other subcultures expect people to keep to themselves. Various words may have nuances of meaning for different linguistic or cultural groups that could affect scores of some persons sufficiently to move them across category boundaries. These matters, which are hardly unique to honesty tests, are not understood well enough to say whether present honesty tests are generally fair and unbiased for use with applicants whose backgrounds differ substantially from the U.S. cultural norm.

Moreover, we need further study of the attitudes of examinees toward honesty tests, their understanding of concepts such as confidentiality, and
their expectations about the application process. Some cultural groups (e.g., a particular Native American tribe) may not use a concept such as confidentiality, at least not one that can easily be imparted by a single word or phrase. Just what examinees think might be done with their test responses when they are told that they will be "kept confidential" is a question worth empirical inquiry.

Environmental Factors

The roles of environmental and circumstantial factors in promoting or limiting employment-related honesty need to be better understood. In diamond mines, the temptation to steal, along with the opportunity, is regarded as so great that individual differences in propensity are not considered of any consequence; every worker is treated with suspicion and thus searched. In many banks, tellers may not be allowed to go home at the end of their workday until their transactions have been checked out and their accounts balanced. Certainly a loss of comity is experienced when everyone is treated with suspicion and implied disrespect. On the other hand, to the extent that work environments can be designed so as to promote honest behavior as the best course of action, the need for incivilities of other sorts is lessened and jobs are opened up to a wider pool of applicants.

Training Test Users

Some publishers of honesty tests appear to have a limited view of their responsibilities for test use. In addition to their usual scoring services, they should do more to educate test users about the appropriate and inappropriate uses of tests and test scores. We need to better understand the ways that test scores actually are used, as well as the kinds of knowledge (and misinformation) associated with the use of honesty tests. Moreover, in
an effort to increase the effectiveness of user training, publishers should both develop new written materials and provide their clients with those already available (e.g., O'Bannon et al., 1989). As a later step, they might cooperate in the production of more comprehensive training programs for users, programs that discuss the nature and properties of honesty tests, the assumptions that underlie their use, and the ways in which test results can best be incorporated into selection decision-making.

Test-User Qualifications

By the publishers' own views, as presented in the A.P.T.P. Guidelines, proprietary test publishers make little attempt to assess the qualifications of test users, because they assume that the publisher is responsible for the quality of the test, its administration, and its scoring. We note with approval the inclusion in the Guidelines of specifications for test users, but we suspect that they may be too general (e.g., "should be qualified," "only trained staff") to be of maximum value. Moreover, the publishers themselves should take the responsibility for helping clients comply with their guidelines. A.P.A. has already devoted much effort to studying test-user qualifications and training procedures (Eyde, et al., 1988), work that should be useful to honesty-test publishers.

Selection Decision-making

We have already mentioned the need for honesty-test users to have a better understanding of the effects of their decision-making procedures; they also need to understand the alternatives that might be available to them were they to change those procedures. In general, users should be helped to formulate criteria by which to judge their selection procedures, including a careful consideration of the applicant characteristics they are seeking and
those they truly want to avoid. Moreover, they should be helped to understand, through a full disclosure of the merits and liabilities of the test that they are being solicited to buy, the extent to which the test will help them to realize their aims.

Confidentiality, Privacy, and Informed Consent

Test users must understand the fundamental need for confidentiality in relation to the use of honesty tests. Test publishers should not only assure themselves that users have that understanding, but they should also institute specific procedures for handling test materials and determining that users are in compliance with acknowledged principles for the protection of privacy. Moreover, scores from any psychological test should not be retained by either the publisher or the client in a form that permits that score to be carried forward from one job application to a new one. Finally, although Standard 16.1 of the Standards for Educational and Psychological Testing states that informed consent can be considered implied when job applicants are tested, test users still have some responsibilities. The pressure on applicants to submit to testing must be balanced by an obligation on the part of examiners to explain fully the application procedures.

Summary and Conclusions

The Task Force began its work with no strong presuppositions about paper-and-pencil measures of honesty and integrity, except perhaps for some skepticism about the likely evidence for their validity. We did, however, begin with the position that honesty tests should be judged by the same standards as other measures developed and used by psychologists.

We found that publishers of honesty tests are highly variable in the ways that they describe their tests, in the ways that they document them, and in
the evidence that they assemble to substantiate claims of their validity and utility. For some published measures, almost no evidence at all is available beyond assurances that evidence exists. For a few measures, extensive research has been completed. However, even in the later cases test publishers have relied on the cloak of proprietary interests to withhold information concerning the development and scoring of the tests, along with other basic psychometric information. The entire field of honesty testing would benefit greatly from substantially increased openness.

Published reviews of honesty tests, supplemented by our own analyses of the available literature, suggests that for those few tests for which validity information is available the preponderance of the evidence is supportive of their predictive validity. Assessment of their construct validity is hampered by inadequate formulation of the constructs presumed to underlie test performance. Nonetheless, to the extent that evidence is available, it is consistent with the idea that these tests reflect aspects of personal integrity and dependability, or trustworthiness.

Many test publishers have not come close to meeting standards for the publication, sale, and use of psychological measures, and these deficiencies should be promptly corrected. Many of these tests lack critical documentation of their psychometric properties, and we strongly urge employers not to use such tests. Despite all of our reservations about honesty tests, however, we do not believe that there is any sound basis for prohibiting their development and use; indeed, to do so would only invite alternative forms of pre-employment screening that would be less open, scientific, and controllable.

We strongly recommend that honesty-test publishers should adopt and adhere to the *Standards for Educational and Psychological Testing*. In
addition, we encourage compliance with the independently formulated **Model Guidelines for Preemployment Integrity Testing Programs**. To adhere to such standards, publishers must prepare manuals that provide a clear and detailed account of the development of each measure and a summary of the evidence for its reliability and validity.

In addition, we strongly urge all honesty-test publishers to make available to qualified parties complete and detailed reports of research on their tests, including all negative findings. Constructs presumed to underlie test performance should be explicitly defined. The use of cutting scores should be discouraged, and alternative procedures for using test scores in a compensatory manner should be provided.

In general, we urge test publishers to market their tests in ways that are circumspect and that reflect the actual empirical foundation for any claims made about the tests. When tests are translated into other languages, evidence for the validity of the translated test should be collected and evaluated prior to its marketing. Moreover, we recommend the intensive study of those language and cultural factors that may influence test performance, as well as of any environmental factors that may serve to attenuate test validity.

Finally, we urge the development of more comprehensive procedures for training test users, for ensuring that only qualified persons administer and interpret the tests, for educating test users about alternative selection methods, and— in general— for safeguarding the confidentiality, privacy, and informed consent of all applicants.
References


Appendix A: Charge to the Task Force

With the passage of the Polygraph Protection Act, organizations are seeking alternative means for the prediction of dishonesty and theft in the employment setting. At present there is not a consensus among research psychologists concerning the adequacy of scientific evidence and ethical use of any specific procedure, including paper-and-pencil integrity tests, used for predicting dishonesty, theft, and related behaviors in employment settings.

As a professional organization, A.P.A. has historically been involved with establishing standards for the ethical design, use, and interpretation of standardized tests. A task force on predictors of dishonesty and theft in employment settings will be established to investigate many unknowns associated with such predictors. The task force will gather and examine data relevant to the scientific and social-policy considerations associated with the development and use of predictors of dishonesty, theft, and related behaviors. The task force will identify specific scientific and social-policy issues to address, including: (a) the adequacy of professional policies, procedures, and standards currently in place for proper use of these predictors by organizations; (b) the adequacy of user training and appropriate professional involvement in the development, administration, scoring, and interpretation of scores; (c) the extent that individuals are provided with accurate and appropriate information concerning instructions, the purpose for using the predictors, and the incorporation of test performance on employment decisions; (d) the accuracy of predictors and related issues (e.g., susceptibility to faking, false positives); (e) the appropriateness of criteria used to validate such predictors; and (f) the adequacy of scientific evidence concerning the prediction of dishonesty, theft, and related behaviors with various types of such predictors.

The task force will gather available information on alternatives for the prediction of dishonesty, theft, and related behaviors and solicit input from organizations and individuals who have developed, used, and reviewed these predictors. Members of the task force may be asked to provide A.P.A. with input concerning federal or state legislation, legal challenges, or inquiries from organizations, the media, and the general public.

The task force will prepare a final report that will include: (a) a brief overview of general perspectives on assessment in employment settings; (b) discussion and clarification of relevant scientific and social-policy issues, unresolved issues, and appropriate strategies (methodologies) for addressing these unresolved issues; (c) recommendations to test consumers (e.g., organizations, test takers, general public); and (d) recommendations to test publishers and testing professionals involved in the development and marketing of these predictors. The final report will include a scientific paper and a policy paper. The task force will be comprised of A.P.A. members with representation from the Committee on Psychological Tests and Assessment (C.P.T.A.), Division 5 (Evaluation, Measurement, and Statistics), and the Society for Industrial-Organizational Psychology (Division 14). Appropriate input from other relevant organizations and groups will be welcomed by the task force. The final report will require approval from all three bodies.

May 5, 1989
Appendix B

APA/DIVISION of Evaluation, Measurement, and Statistics/Society for Industrial-Organizational Psychology

Task Force on the Prediction of Dishonesty and Theft in Employment Settings

QUESTIONNAIRE FOR PUBLISHERS OR PROVIDERS

Please answer the following questions in the space provided. You may attach additional comments on separate sheets of paper.

1. GENERAL INFORMATION

1.) Publisher or Provider

2.) Contact person

3.) Title, address, and telephone number of contact person:

4.) Name of measure

5.) The measure is designed for use with:
   ___applicants
   ___job incumbents
   ___both
   ___others (specify)

6.) What year was the measure last revised? ______

7.) The measure is designed to provide the following types of information (check all that apply):

   ___predict individuals' honesty
   ___predict individuals' integrity
   ___predict individuals' reliability
   ___predict individuals' job performance
   ___predict individuals' counterproductivity (e.g., theft, turnover, absence)
   ___predict individuals' attitudes
   ___predict group criteria (specify)
   ___other (specify)
II. PURCHASE AND ADMINISTRATION OF MEASURE

8.) Do you use test purchaser forms or similar methods to determine the qualifications of individuals wishing to purchase and use the measure?
   ___ Yes, use a test user qualification form (enclose)
   ___ Yes, use other methods (specify)
   ___ No

9.) Do you conduct any screening for organizational purchases?
   ___ Yes (specify)
   ___ No

10.) What qualifications are required for individuals to administer the measure? (check all that apply and explain)
   ___ Successful completion of training by test publisher (specify)
   ___ Graduate degree specializing in (specify)
   ___ License or certification in (specify)
   ___ Completion of graduate courses in (specify)
   ___ Completion of undergraduate courses in (specify)
   ___ Other evidence or knowledge, training, and experience (specify)
   ___ Agreement to comply with "sound testing practices" (specify)
   ___ Supervised experience with this measure (specify)
   ___ Other (specify)

III. SCORING AND RESULTS

11.) Please check all available methods for scoring the measure
   ___ Computerized scoring (telephone, modem, and mail-in services by publisher)
   ___ PC software (computer software that user has purchased)
   ___ Hand scoring by user (with scoring keys from publisher)
   ___ Other (specify)
12.) The following questions concern the scores reported by the publisher. Please check all that apply.

A.) How are scores reported to the test user?
   ___ subtest scores
   ___ a total score
   ___ other (specify)

B.) What types of derived scores are reported?
   ___ standard scores
   ___ percentile scores
   ___ raw scores
   ___ other (specify)

C.) What is the reference group for the derived scores?
   ___ local group only (specify)
   ___ national group (specify)
   ___ other (specify)

D.) Do you provide any of the following additional information on an individual’s score?
   ___ confidence range or band
   ___ separate norms by gender
   ___ separate norms by race
   ___ separate norms by industry
   ___ separate norms by occupation
   ___ other separate norms (specify)
   ___ other

IV. INTERPRETATION AND REPORTS

13.) What qualifications are required for individuals to interpret the results for the measure (check all that apply and explain)?

   ___ successful completion of training by publisher (specify)
   ___ graduate degree specializing in (specify)
   ___ license or certification in (specify)
   ___ completion of graduate courses in (specify)
   ___ completion of undergraduate courses in (specify)
   ___ other evidence or knowledge, training, and experience (specify)
   ___ agreement to comply with "sound testing practices" (specify)
__supervised experience with this measure (specify)________________________

__other (specify)______________________________________________________

14.) Do you offer individual or group training for users?
   __Yes (specify)_______________________________________________________
   __No

15.) Do you furnish interpretive reports/results on each individual's scores to the organization?
   __Always
   __If requested
   __Never

16.) What types of interpretive reports are provided?
   __Narrative report
   __Summary descriptions of subscale/global scores
   __Graphic presentations
   __Other (specify)______________________________________________________

17.) Are your interpretive reports computer-based?
   __Yes
   __No (explain)_______________________________________________________

18.) Are cutoff or pass scores provided by the test publisher? (Please check all that apply)
   __Yes, a general cutoff score may be provided.
   __Yes, a specific cutoff scores may be provided for a user.
   __Yes, the publisher may assist a user in setting a cutoff score.
   __No, cutoff scores are inappropriate for most applications of the measure.

V. APPLICATIONS AND SERVICES

19.) Check the approximate number of clients tested with this instrument during the past 12 months:
   __less than 999
   __1,000 to 4,999
   __5,000 to 9,999
   __10,000 to 14,999
   __15,000 to 19,999
   __Over 20,000

20.) Indicate the approximate number of different organizations/businesses using this measure in the past 12 months? _______
21.) Check all of the following that are considered an appropriate use of the measure:
   _______ Initial screening of job applicants.
   _______ Final hurdle in screening of job applicants.
   _______ Screening of incumbents for retention.
   _______ Assess organizational dynamics (e.g., theft rate, motivation)
   _______ Other (specify) ________________________________

22.) Do you provide advice to users concerning issues such as race or sex bias and other Equal Employment Opportunity concerns?
   _______ Yes (specify) ________________________________________
   _______ No

23.) Do you provide feedback to individuals completing the measure?
   _______ Always
   _______ Sometimes, when requested by test user
   _______ Never

24.) Do you employ doctoral level research psychologists and measurement specialists to develop the measure and provide services to users.
   _______ Yes, as full-time employees
   _______ Yes, as consultants
   _______ No

25.) Do you have specific policies or procedures for users and individuals taking the measure to register complaints and resolve problems?
   _______ Yes (specify) _______________________________________
   _______ No

26.) Are you willing to furnish the scoring keys and computer algorithms used for scoring and interpreting the measure?
   _______ Yes
   _______ No

Please send the completed questionnaire and other material to:

Dianne Lane
Science Directorate
American Psychological Association
1200 Seventeenth St., NW
Washington, D.C. 20036
(202) 955-7653
APPENDIX C

ATTENDEES AT THE MARCH 31, 1990 TASK FORCE OPEN MEETING

David W. Arnold, Reid Psychological Systems
Philip Ash, Ash, Blackstone and Cates
Gerald L. Borofsky, Bay State Psychological Associates
Michael R. Cunningham, University of Louisville
Linda Goldinger, private consultant
William G. Harris, Stanton Corporation
John W. Jones, London House, Inc.
Gary Kay, Police and Public Safety Section, Division 18
Martin L. Kurke, Police and Public Safety Section, Division 18
John B. Miner, Intergram, Inc.
Marc Minor, Team Building Systems
Robert Perloff, University of Pittsburgh
Richard L. Rees, Personnel Resources Consultants
Nancy Robinson, Amtrak
Leonard Saxe, Brandeis University
Lance Seberhagen, Seberhagen and Associates
William Terris, London House, Inc.
APPENDIX D

THE TASK FORCE WOULD LIKE TO THANK REVIEWERS OF AN EARLIER DRAFT:

David W. Arnold, Reid Psychological Systems
Phillip Ash, Ash, Blackstone and Cates
Association of Personnel Test Publishers
Gerald L. Borofsky, Bay State Psychological Associates
Peter D. Bullard, American Tescor, Inc.
David P. Campbell, Center for Creative Leadership
Lee J. Cronbach, Stanford University
Michael R. Cunningham, University of Louisville
Robyn M. Dawes, Carnegie-Mellon University
Heber W. Eber, Psychological Resources, Inc.
Lorraine D. Eyde, U.S. Office of Personnel Management
David Faust, University of Rhode Island
Donald W. Fiske, University of Chicago
Fred Fuller, Alert Personality Profile, Inc.
Linda Goldinger, private consultant
Harrison G. Gough, University of California at Berkeley
Marilyn K. Gowing, U.S. Office of Personnel Management
William G. Harris, Stanton Corporation
Robert Hogan, University of Tulsa
Lloyd Humphreys, University of Illinois
Robin Inwald, Hillsion Research
Douglas N. Jackson, University of Western Ontario
John W. Jones, London House, Inc.
Benjamin Kleinmuntz, University of Illinois at Chicago
Richard Kilmoski, on behalf of Division 14 Scientific Affairs Committee
N. Knuth, P.O.S. Corporation
Richard L. Lanyon, Arizona State University
Maurice Lorr, Catholic University of America
David T. Lykken, University of Pittsburgh
Lewis L. Maltby, American Civil Liberties Union
Scott L. Martin, London House, Inc.
Robert Mauro, University of Oregon
Paul McReynolds, University of Nevada
John B. Miner, Intergram, Inc.
Warren T. Norman, University of Michigan
Daniel J. Ozer, Boston University
Robert Perloff, University of Pittsburgh
Lawrence A. Pervin, Rutgers University
Richard L. Rees, Personnel Resources Consultants
Leonard Saxe, Brandeis University
Frank Schmidt, University of Iowa
Lance W. Seberhagen, Seberhagen and Associates
Lawrence J. Stricker, Educational Testing Service
Mary L. Tenopyr, on behalf of Division 5 Committee of Public Affairs
Jerry S. Wiggins, University of British Columbia