dence was to be obtained from randomized, controlled research, however, that recommendation could change. To argue from a different evidentiary standard without making this clear or presenting why the different standard is preferred is to play somewhat fast and loose with the issues.

In summary, although Muñoz et al.'s (1994) article was not labeled as the "official" American Psychological Association position, it can be confused as such because of its publication with articles representing the agency guidelines (Clinton, McCormick, & Besteman, January 1994; Schulberg & Rush, 1994). The position represented does a disservice to psychology by appearing to place its own interests above those of depressed patients.

REFERENCES


2. Integrity tests predict non-self-report broad composites of counterproductive behaviors with operational validities of .39 and .29 (depending on the type of test).

3. Integrity test validities generalize across tests, jobs, organizations, and settings.

Our second comment addresses what integrity tests measure. Camara and Schneider (1994) stated that (a) "there is little agreement on the behaviors of interest and the precise definition of integrity" (p. 115); (b) the construct of integrity is overly broad and ill defined" and poorly understood (p. 115); and (c) "researchers have generally agreed that there is insufficient evidence to reach definitive conclusions" about the construct of integrity (p. 117). Ones (1993) recently completed a dissertation focusing on the construct validity of integrity tests. Findings from this dissertation have not been published yet. However, this dissertation won Division 14's (Society for Industrial and Organizational Psychology) 1994 best dissertation award. In this dissertation extensive evidence for convergent and discriminant validity of integrity tests is presented. Results were based on more than 8,000 correlations from various primary and secondary databases. An important finding was that integrity tests displayed significant convergent validity (i.e., a large general factor across integrity tests was confirmed). Of the Big Five dimensions of personality, conscientiousness has the highest correlation with integrity tests. However, a linear composite of the conscientiousness, agreeableness, and emotional stability dimensions of the Big Five described integrity better than any one dimension. With integrity tests, the increased breadth of predictor construct coverage appeared to translate itself into better prediction of job performance.

The major problem with the research assessing the construct validity of integrity tests has been a focus on narrow (i.e., construct limited) personality variables (Ones, Schmidt, & Viswesvaran, 1993). Typically, the question has been whether integrity tests measure any one of the Big Five factors or any facets within the Big Five factors. Until our research on the construct validity of integrity tests, no one had asked whether integrity tests might measure a construct broader than any one of the Big Five traits (Ones, Schmidt, & Viswesvaran, 1994). In all the writings on integrity tests, there is an implicit assumption that broader constructs are less useful than narrow or specific constructs in prediction. Perhaps this state of affairs is a reflection of the remnants of the behavioral fractionation that has afflicted social sciences for many years. Our findings that the broad construct of integrity is useful, meaningful, and practical for prediction
holds promise for the development of general and parsimonious theories of human motivation and work behavior.

Our third comment addresses Camara and Schneider's (1994) concerns that most of the research on integrity tests is conducted by test publishers. There is no inherent reason why the credibility of in-house research should be questioned. Many test publishers are I/O psychologists in good standing. In addition to I/O psychologists, many personality psychologists with good reputations are tied directly to or serve as consultants to various integrity test publishers. Furthermore, to date we have seen no evidence that negative findings have been withheld. All research is sponsored to some degree, and all primary studies are potentially affected by biases and errors (Hunter & Schmidt, 1990; Schmidt, 1992). There is also a catch-22 here. APA test standards and APA ethical standards require test authors to conduct research on the validity of the instruments they offer. Yet, when they do so, their studies are rejected as being of no scientific value because of their connections to the instrument. So they are damned if they conduct validity studies and damned if they don't.

We hope the empirical findings on integrity tests will contribute to advances in understanding the role of personality in behaviors on the job. Future reviews of integrity testing should base their conclusions on all available data.

REFERENCES


Why Integrity Testing Remains Controversial

Scott O. Lilienfeld
Department of Psychology,
Emory University

George Alliger and Krystin Mitchell
Department of Psychology,
University at Albany,
State University of New York

Camara and Schneider's (February 1994) article "Integrity Tests: Facts and Unresolved Issues" provided a balanced overview of current issues pertaining to the use of integrity tests in pre-employment screening. Nevertheless, Camara and Schneider's title and their discussion of the American Psychological Association (APA) task force report (Goldberg, Grenier, Guion, Sechrest, & Wing, 1991) may leave readers with the impression that the authors have addressed most of the principal unresolved issues concerning integrity tests. To the contrary, we maintain that Camara and Schneider's review failed to get to the heart of the major concerns voiced by scientists and educated laypersons regarding the use of these measures.

Although integrity testing has provoked considerable debate among both researchers and the general public, the substantive issues underlying this debate have not always been apparent. For example, Schmidt, Ones, and Hunter (1992) asserted that "although integrity testing, for reasons that are not entirely clear, remains controversial, the APA report may lead to greater acceptance" (p. 641, italics added). We contend that three key issues concerning integrity tests, which were largely neglected by Camara and Schneider (1994), remain unresolved. The persistence of these issues helps to explain why integrity testing remains controversial.

1. Although there is compelling evidence that many integrity tests possess above-zero validity for certain criteria, such as counterproductive behavior in the workplace (Ones, Viswesvaran, & Schmidt, 1993), there are virtually no data regarding the types of classification errors made by these measures. Most discussions of integrity tests assume that the errors made by such tests are random. But many of the criticisms directed toward integrity tests, such as the assertion that they are biased against highly religious or moral individuals (Guastello & Rieke, 1991; Lykken, 1981), imply that the classification errors made by these tests may be systematic. These arguments echo recent concerns that most paper-and-pencil measures relevant to conscientiousness, including integrity tests, assess conventionality and conformity rather than morality (Loevinger, 1994).

Moreover, these criticisms underscore the potential for the exclusion of certain individuals from occupations on the basis of characteristics that are largely or entirely unrelated to honesty. Indeed, integrity test scores are positively correlated with measures of social closeness, traditionalism, and other traits that are not clearly relevant to integrity (Lilienfeld, Andrews, Stone-Romero, & Stone, 1994). Although Camara and Schneider (1994) referred to the problems of labeling and false positive errors that have dogged integrity tests (p. 115), the more critical question concerns whether such errors reflect the systematic misclassification of some honest individuals as dishonest.

2. Camara and Schneider (1994) also did not address the issue of fakability—in other words, the extent to which integrity tests are susceptible to impression management. This possibility, which has been the focus of surprisingly little research, is consistent with findings that integrity test scores are positively correlated with scores on lie scales (Guastello & Rieke, 1991). In one of the few investigations of the fakability of integrity tests, Ryan and Sackett (1987) found that an integrity test modeled after existing measures was highly susceptible to faking, although the scores of participants asked to take the test "as if applying for a job" were more similar to the scores of participants asked to respond candidly than to the scores of participants asked to fake their answers. Finally, it should be noted that because many integrity tests do not contain lie scales, respondents on such tests may be able to engage in response distortion without risking detection.

A related issue that has received no research attention concerns the coachability of integrity tests, or the extent to which these tests are vulnerable to training by means of a standard set of instructions. Because most integrity tests, especially those that are overt or with a clear purpose, are highly similar to one another in content, the finding that integrity tests are amenable to coaching implies that they would be at risk for becoming obsolete as soon as the heuristics for passing them become widely known. Similarly, one of the primary liabilities of the