

## **Pre-Employment Integrity Testing in Israel: A Validation Study**

**Saul Fine**  
**Midot, Ltd.**

**Baruch Nevo**  
**University of Haifa, Israel**

**Merav Hemi**  
**University of Haifa, Israel**

*Few prior studies have examined the validity of integrity testing in Israel, where these tests are prevalent. Among 201 Israeli students, overt integrity was found to have statistically significant correlations with: self-reported counterproductive work behaviors (-.29), prior dismissals (-.18), and a simulated theft scenario (-.20), but not with a simulated deception scenario. In addition, the integrity test scores showed no indication of adverse impact or test bias for gender, age, or national origin. Overall, this study supports the international literature on this topic, and provides evidence towards the usefulness of integrity testing for personnel selection in Israel.*

### **INTRODUCTION**

Employee integrity may very well be one of the most robust personality constructs in personnel psychology (Gottfredson, 2002). Job analyses and employee surveys alike have consistently identified integrity as critical job requirement for a broad range of jobs and industries around the world (Bartram, Lindley, Marshall, & Foster, 1995; Fine, 2008; Kouzes & Posner, 2009; Scholarios & Lockyer, 1999). In addition, integrity often appears as a fundamental component of leading companies' organizational values (AMA, 2002; Fulmer & Conger, 2004). Indeed, it would be difficult to imagine an organization for which the integrity of its employees is not of paramount importance for its success. As a result, and considering a fundamental tenet of personnel selection is to predict job performance based on key job competencies (Guion, 1991), in some respects, it might be considered inappropriate for organizations *not* to assess a trait as important as this one (Martin & Orban, 1995).

Despite the rational and empirical basis for assessing integrity among new job applicants and the increasingly common practice for doing so in organizations around the world (Miner & Capps, 1996), supporting research regarding the effectiveness of integrity tests internationally is still very much lacking. In addition, only a small percentage of the current research on this topic has included primary samples from outside of North America (Berry, Sackett, & Weimann, 2007). Despite the widespread use of integrity in Israel, for example, and the recognition of integrity as a critical job requirement in Israel (Solomon, 2010), little local validity evidence is available to support their usefulness.

## **Integrity Testing**

Integrity testing has seen steady growth in both practice and research in the past two decades, with increased evidence regarding the usefulness of integrity testing for screening out job applicants with risks towards future involvement in counterproductive work behaviors (CWB; Ones & Viswesvaran, 2001; Wanek, 1999). In addition, while a recent meta-analysis by Van Iddekinge, Roth, Raymark, and Odle-Dusseau (in press), has raised concerns regarding the validity of integrity tests for predicting job performance, the validity for predicting counterproductive work behaviors remains supported.

In terms of the content of integrity tests, these tests typically include direct questions regarding attitudes and opinions towards CWB. Integrity tests of this type are known as “overt” tests, and may be distinguished from personality-based tests which measure psychological traits that are primarily related to facets of conscientiousness, agreeableness and emotional stability (Sackett, Burris, & Callahan, 1989). The predictive validities of overt and personality tests are largely similar (Ones, Viswesvaran, & Schmidt, 1993), while overt integrity tests have been reported to predict CWB beyond that of traditional Big Five scales (Berry et al., 2007). Overt integrity tests, which are more commonly used in Israel, are based primarily on attitude-behavior theory (Ajzen & Fishbein, 1977), whereby individuals who believe CWBs are normative or justifiable, and/or are lenient towards their offenders, are considered to be a greater risk for engaging in such behaviors themselves (Murphy, 1993; Wanek, 1999). A typical item from a standard overt integrity test, for example, might be the degree to which an individual agrees or disagrees with the statement, “most employees will steal from employers at least once,” presumably tapping the perceived normativity of employee thefts.

It should be noted, however, that CWB does not merely refer to employee crimes. CWB is considered to include a wide range of inappropriate and/or undesirable work behaviors that may target an organization (CWB-O; e.g., drugs, sabotage, absenteeism, and negligence) as well as other individuals (CWB-I; e.g., harassment, disobedience, and backstabbing), with varying degrees of severity (Robinson & Bennett, 1995). It should also be noted that these behaviors are prevalent, with the great majority of U.S. employees reportedly having engaged in some form of CWB, roughly one third admitting to having stolen from their employers (Hollinger & Clark, 1983; Sackett & Devore, 2001), and similar results from Israel (Lachman & Shpigel, 2004). Therefore, considering the damages that may be caused by CWB, which have been valued in the hundreds of billions of dollars annually (Murphy, 1993), organizations clearly have a strong interest to reduce the extent of CWBs.

It may not be surprising, therefore, that integrity tests are in used for personnel selection around the world. Although survey studies have found integrity test usage to be less extensive than alternative assessment tools (Ryan, McFarland, Baron, & Page, 1999), estimates indicate that millions of tests are administered each year in the U.S. alone (Camara & Schneider, 1994). No specific research is available on the extent of integrity testing in Israel, but based on the first author’s experience, many of the large organizations in Israel, and certainly a majority (if not all) of the large banking institutions in the country, use integrity tests as part of their standard assessment and selection processes. In addition, several Israeli-developed integrity tests are commercially available, which are used nearly exclusively.

## **Studying Integrity Test Validity**

In one of the most criticized aspects of integrity test research, studying criterion based validity presents something of a methodological challenge. Specifically, objectively defined criteria of CWB are not always reliable or observable, due in part to their notoriously low base rates and even lower detection rates (Camara & Schneider, 1994). In addition, supervisor appraisals of integrity are far less readily available, and may also suffer from “conceptual limitations” (Berry et al., 2007, p. 280). Instead, a commonly used alternative criterion for measuring integrity test validity in field studies has been the use of personal admissions (Sackett & Wanek, 1996). Personal admission criteria are self-reports of specific past behaviors (e.g., prior involvement in workplace theft), or the consequential outcomes of these behaviors (e.g., prior dismissals). In some respects, it is possible to consider admissions as proxies for external evidence of prior involvement in these behaviors. In this way, admissions facilitate measures of “postdictive” validity (Jones, 1991).

The use of self-reported postdictive criteria measures such as these is not without criticism, however. In particular, the reliability of admission criteria may be criticized for their susceptibility to faking in high-stakes scenarios. Paradoxically, researchers have found that verifiable questions about one's past behaviors are typically responded to honestly (Becker & Colquitt, 1992). Therefore, despite the apparent ability to distort responses, there is less evidence that job applicants actually do so. Stuman and Sherwyn (2007) offer that the reasons for these unintuitive behaviors may be due to job applicants' rationalizations for responding honestly, such as: trivializing their deeds due to their perceived normativity, justifying their actions due to extenuating circumstances, under-reporting (i.e., admitting to only a portion of their behaviors), and providing honest admissions due to concerns that their responses will be verified. Accordingly, it is perhaps understandable why some studies have found that faking in general on integrity tests has minimal effects on overall validity levels (Ones & Viswesvaran, 1998a; Ones, Viswesvaran, & Reiss, 1996).

Another criticism for using admission criteria is that of "common method bias," which is created when both the predictor and the criterion are derived from individual self-reports, and is exacerbated by content overlap in the measures. Common method bias might be responsible for spuriously high correlations, while the lack of reliability in external measures may lead to spuriously low correlations. Indeed, postdictive validity coefficients are typically much larger ( $\rho = .42$ ) than those found for externally derived criteria ( $\rho = .15$ ; Van Iddekinge et al., in press), clearly indicating the need to include both criterion source types when validating these tools.

In addition to postdictive admissions, and in light of the difficulty to reliably observe CWB in occupational settings, a growing number of researchers have chosen to adopt strategy of studying CWB in controlled research designs (Berry et al., 2007). In these experimental settings, participants are presented with opportunities to engage in simulated acts of CWB, while being observed by the researchers without the participants' awareness. Examples of such designs include giving participants an opportunity to change answers or scores on a performance task when they think no one is looking, giving them access to a test's answer key and observing whether or not they use it to correct their answers, giving them unsupervised access to food or small money and observing who takes without permission, etc. (Nicol & Paunonen, 2002; Mikulay & Goffin, 1998). Overall, integrity tests have found low to medium validities for simulated behavioral criteria such as these (e.g., scenarios of deception  $r_s = -.08$  to  $-.25$ , and stealing  $r_s = -.04$  to  $-.32$ ) depending on the given tasks and situations (see Berry et al., 2007, for a summary).

In order to have practical relevance, simulated elicitation of counterproductive behaviors may wish to measure realistic occupational scenarios (e.g., observing acts of theft as opposed to acts of cheating on an academic test). In addition, any single-act criteria such as these would need to be interpreted cautiously, as they themselves may be questioned in terms of measurement error and overall reliability. Therefore, including multiple such criteria is preferable.

### **Integrity Testing in Israel**

Given the importance of integrity as a key competency requirement for jobs around the world (Kouzes & Posner, 2009), the validity of integrity tests can be reasonably hypothesized to be generalized cross-culturally. Moreover, based on the robust meta-analytic data from the U.S. by Ones et al. (1993) and Van Iddekinge et al. (in press) and validity generalization theory (Schmidt & Hunter, 1977) one could hypothesize integrity test validity to be transportable across contexts and settings. However, as long as the primary samples of these meta-analyses were derived in similar geographies, empirical evidence is still required to generalize the validity to new cultures (Herriot & Anderson, 1997; Lievens, 2008). Such appears to be the case with integrity testing.

Initial evidence has been found for the cross-cultural validity of integrity testing in predicting CWB admissions, although this area is still very much understudied. In one such study, Fortmann, Leslie, and Cunningham (2002) found that their overt integrity test maintained significant validity coefficients for CWB admissions and supervisor ratings among samples from Argentina, Mexico, South Africa and the US. In another study, Marcus, Lee, and Ashton (2007) reported similar validities for integrity against self-reported CWB admissions among German and Canadian undergraduates. In one of the only Israeli based

studies located, Fine, Horowitz, Weigler, and Basis (2010) found evidence for the validity of integrity in predicting criminal admissions among a sample of retail employees in Israel ( $r = -.35$ ). In another such study, Fine (2010), cross-culturally compared integrity test validities for general CWB admission criteria in the banking sector, and reported a significant validity for the Israeli sample ( $r = -.32$ ), as well as for the other countries studied. While encouraging, the above studies were primarily based on self-reported admissions of CWB, and have yet to incorporate controlled designs of actually elicited CWB behaviors. In addition, since the above studies were carried out on actual job applicants, incentives to falsify self-reported criteria information and the possible contamination of the validity results cannot be refuted. Controlled experimental designs are therefore important to corroborate prior cross-cultural findings.

In addition to validity evidence, meta-analyses by Ones & Viswesvaran (1998b) have found integrity tests to be non-discriminating against protected minority groups for age ( $>40$ ), gender (females), and race (non-whites) in a variety of settings. In fact, females tend to score slightly higher than males ( $d = .16$ ), and older individuals slightly higher than younger ones ( $d = .08$ ). That integrity tests are non-discriminating is, of course, essential to ensure fair and legal hiring practices when used in personnel selection around the world (e.g., SIOP, 2003; US EEOC, 1978) as well as in Israel (i.e., Israeli Ministry of Industry, Trade and Labor, 1988; Israel Psychological Association, 2004). However, the effects of integrity on local subgroups outside of the U.S., in terms of age, gender and race are still largely unknown. Further local research in this area is therefore important.

Building on the current integrity testing literature, the present study set out to examine the degree of usefulness of integrity testing in Israel in a controlled setting, using both interpersonal and organizational oriented CWB criteria in the form of admissions and simulated behaviors.

## METHOD

### Sample

The sample for this study included 201 students from a large public Israeli university. 49.3% of the sample ( $N = 99$ ) were males and 50.7% females ( $N = 102$ ). 86.1% of the sample ( $N = 173$ ) were native Israelis and the remaining 13.9% ( $N = 28$ ) were immigrants born outside of Israel, primarily from the former Soviet Union. Participants' mean age was 25.5 ( $SD = 2.6$ ), each with at least one year of job experience. The sample was represented by the following areas of study at the university: social sciences (36.3%), humanities (30.3%), health and welfare (13.9%), law (12.4%), education (5.5%), and other (1.6%).

### Measures

#### *Integrity*

Integrity was measured using the *IntegriTEST*, a web-based overt integrity test, developed in Israel (Fine, 2008). The test includes 96 items, which measure opinions (i.e., punitiveness, pervasiveness, and rationalization) towards CWBs. The test has one overall scale and 4 sub-scales: Theft, Report, Norms and Bribery. The test's admissions section, which relates to past behaviors, was *not* included in the overall score for this study. The overall score consists of three ordinal recommendation ratings, as follows: (1) "Not recommended," (2) "Marginal" and (3) "Recommended", where (2) is the typical cutoff score. Only the test's overall score was used in this study. Sample items include questions such as: "If a colleague was to offer you drugs, would you complain to their superiors?"; "Do you think a person has the right to break laws they don't agree with?"; and "Should a company employ someone who is known to have been fired for stealing in the past?" Each item includes a 3-point response scale ("Yes", "No" and "???"). The overall score has a test-retest reliability of .70 over a period of between 2 and 8 weeks, based on a sample of 346 job applicants. A convergent validity of .52 was found between the overall integrity score and independently collected pre-employment security interviews among 217 job applicants in a large bank. And, among 1,232 job applicants in a large retail chain, divergent validity was found between integrity scores and mental ability levels ( $r = .03$ ).

### *Counterproductive Work Behaviors (CWB)*

The criterion used for this study was a self-report measure of past involvement in deviant workplace behaviors. Items were based on a translated version of the 19-item CWB behavioral admissions checklist developed by Bennett and Robinson (2000). The CWB admissions index included 12 items measuring organizational deviance (CWB-O; i.e., acts targeting the organization such as theft, fraud, bribery, sabotage, absenteeism; Cronbach alpha = .82), and seven items measuring interpersonal deviance (CWB-I; i.e., acts targeting other members of the organization such as fighting, rudeness, offending, etc.; Cronbach alpha = .80). Confirmatory factor analysis using structural equation modeling (SEM) with latent variables produced a good fit for the two main factors of CWB,  $X^2(150) = 264.1, p < .001$ ; RMSEA = .06 (90% CI = .49 to .74); CFI = .90. Participants indicated the frequency with which they had engaged in deviant behaviors by responding to each item on a 7-point Likert type scale: (1) “never”, (2) “once a year”, (3) “twice a year”, (4) “several times a year”, (5) “monthly”, (6) “weekly”, and (7) “daily”. Sample items include: “I act rudely toward someone at work”; “I say hurtful things to people at work”; “I come to work late without permission”; “I neglect to follow my boss’s instructions”. Overall CWB was measured as the mean of both facets (CWB-O and CWB-I), representing an overarching latent construct (Berry, Ones, & Sackett, 2007).

### *Prior dismissals*

Prior dismissals were measured by four self-reported biographical statements each having two response options (1) “yes” and (0) “no”. The items tap four possible reasons for having been fired in the past: Inappropriate behavior, conflicts with superiors, suspicion of misconduct, committing an offense (i.e., “have you ever been fired for...”). The overall score was dichotomized based on the endorsement of one or more items. Reliability KR-20 = .64.

### *Theft scenario*

A dichotomous behavioral criterion of organizational deviance, representing monetary theft was created by using a simulated theft paradigm. In this paradigm, participants were purposely (and significantly) overpaid for their involvement in the experiment, after which time the administrator reported whether or not the participants returned the extra money (any time thereafter for the duration of the experiment). Specifically, each participant was told before and after the experiment that they would be paid 60 shekels (approximately 15 U.S. dollars) for their participation in the experimental trial. Upon completion of the experiment, the administrator reminded each participant individually by saying, “here is your payment of 60 shekels” when handing them an envelope with their payment. In each envelope were four notes of 20 shekels each (i.e., 80 shekels), instead of the appropriate three notes of 20 shekels, in order to appear as if an accounting error had occurred (based on Cunningham, Wong, & Barbee, 1994).

### *Deception scenario*

A dichotomous behavioral criterion of interpersonal deviance representing dishonest reporting was created by asking participants to report their psychometric university entrance exam (PET) scores. Subsequently, the reported scores were matched against the university’s database, and positive discrepancies (i.e., where reported scores were higher than actual scores) were noted. The PET, developed by the National Institute of Testing and Evaluation (NITE), in Israel, consists of three subtests: verbal reasoning, quantitative reasoning, and English language, which are aggregated into an overall composite score of general scholastic aptitude. This composite score is reported as a standardized score, ranging from 200 to 800, with an approximate mean of 500 and standard deviation of 100. Reliability and validity indices for the PET have been studied at length (e.g., Beller, 1994; Nevo & Oren, 1986).

### **Procedure**

Participants were recruited to participate in this study via notices that were posted around the university campus advertising the experiment as a study in attitudes towards behaviors at work. Participants were tested in a computer lab in groups ranging from one to ten, and underwent other

experimental trials in addition to those described here. At the start of this trial, the participants were specifically instructed to respond honestly to the questionnaire despite the sensitivity of the content. In addition, they were promised complete privacy of their personal responses, and used dummy-coded identification numbers throughout the experiment. The integrity measure was completed online, while the CWB questionnaire was completed separately in paper and pencil format. All participants were explained the nature of the experiment in advance, signed a written consent form to participate in the experiment and have the data used for research purposes, and were given the opportunity to discontinue the experiment at any time. Participants also granted the researchers permission to retrieve PET scores from the university database. The researchers gave the participants ample time to return the money in the simulated theft scenario (up to several weeks), and did not approach anyone who did not return the money.

## RESULTS

### Descriptive Statistics

The descriptive statistics of each measure can be found in Table 1. The mean integrity score was 2.16 ( $SD = .92$ ), which reflects: 51.7% high or “recommended” scorers ( $N = 104$ ), 12.4% moderate or “marginal” scorers ( $N = 25$ ), and 35.8% low or “not recommended” scorers ( $N = 72$ ), based on the tests standard norms. In terms of the criteria, the mean CWB score was 2.05 ( $SD = .77$ ), indicating that most participants had engaged in at least some form of CWB at least once. In addition, 13.9% of the sample ( $N = 28$ ) admitted to having been fired from a previous employer. Regarding the simulated theft, just under one-third (i.e., 30.2%,  $N = 57$ ) of the sample did *not* return the extra money. Finally, actual PET scores ranged from 392 to 763, with a mean of 622.74 ( $SD = 84.60$ ). 39.3% of the participants ( $N = 77$ ) inaccurately reported their PET scores (to at least some degree) by an average difference of 25.5 points ( $SD = 38.2$ ).

**TABLE 1**  
**DESCRIPTIVE STATISTICS AND INTER-CORRELATIONS**

	<i>N</i>	Mean	<i>SD</i>	1	2	3	4	5	6	7
1. Integrity	201	2.16	0.92	(.70) <sup>a</sup>						
2. CWB-I	201	2.12	0.89	-.22**	(.80)					
3. CWB-O	201	2.05	0.77	-.27**	.42**	(.82)				
4. CWB-overall	201	2.08	0.70	-.29**	.86**	.82**	(.85)			
5. Dismissal	201	1.14	0.35	-.18*	.11	.18*	.17*	(.65)		
6. Theft	189	1.70	0.46	-.20**	.03	.23**	.15*	.06	(--)	
7. Deception	196	1.39	0.49	-.01	.16*	.03	.12	.10	-.01	(--)

\* $p < .05$ , \*\* $p < .01$ . *Notes:* Reliability estimates are shown in the diagonal; Integrity scores ranged from 1 (low) to 3 (high); CWB ranged from 1 (never) to 5 (daily); Dismissal, Deception, and Theft: 1 (no), 2 (yes). <sup>a</sup> Source: Fine, 2008.

### Validity Indices

As shown in Table 1, integrity test scores were significantly correlated with CWB (-.29). Integrity scores were also correlated with prior dismissals (-.18), as well as the simulated theft scenario (.20). However, integrity scores were not significantly correlated with the deception scenario. Among the criteria, self-reported CWB was correlated with the behavioral theft criterion (.15), but not with deception (.06). Interestingly, however, on the facet level, CWB-O was more strongly correlated with the theft scenario (.23) than was CWB-I (.03). And, CWB-I was significantly correlated with deception (.16), while CWB-O was not (.03).

In accordance with these validity levels, significant differences in the incidence of CWB by test scores can be expected. Specifically, 19.4% ( $N = 14/72$ ) of the individuals with low (i.e., “not recommended”) integrity scores, admitted to having engaged in counterproductive work activity on a *weekly basis or more*, compared to just 6.7% ( $N = 8/104$ ) among the individuals with high (i.e., “recommended”) scores. In other words, there was nearly a three times greater probability for being consistently involved in CWB among individuals with low integrity scores,  $X^2 = 6.54, p < .01$ .

Similarly, in terms of the simulated theft, individuals with low integrity scores were nearly twice as likely to *not* have returned the additional money given to them, compared to individuals with high integrity scores. Specifically, 40.6% of the low integrity test scorers ( $N = 28/69$ ) chose not to return the money compared to 21.1% ( $N = 20/95$ ) of the high integrity test scorers,  $X^2 = 7.36, p < .01$ .

Finally, 20.8% ( $N = 15/72$ ) of individuals with low integrity test scores admitted to having prior dismissals compared to just 7.9% ( $N = 8/104$ ) of individuals with high integrity test scores, a ratio of more than two and a half to one,  $X^2 = 6.47, p < .01$ .

### Fairness

No evidence of adverse impact due to test score differences was found for the subgroup studied (see Table 2). Since the sample had no participants over the age of 37, the typically studied protected subgroup of individuals over the age of 40 was not available. Instead, differences between particularly young participants (<25) were investigated. This was not done from an adverse impact perspective, of course, as young individuals are not a protected group, but rather to investigate possible differences due to young test takers’ lesser work experience and/or different possible outlook regarding CWB that might have influenced their responses (Ones & Viswesvaran, 1998b). No such differences were found ( $d = 0.09$ ). Females’ integrity scores were slightly *higher* than males ( $d = 0.25$ ). And, scores from new immigrants were not significantly different from those from native Israelis ( $d = 0.11$ ). As a rule of thumb, effect sizes ( $d$ ) of 0.2 and below are considered to be small, generally indicating no meaningful operational differences (Cohen, 1988). Similarly, none of these subgroup differences were found to violate the 4/5<sup>th</sup> rule for estimating adverse impact. The 4/5<sup>th</sup> rule is a method for evaluating adverse impact, which states that protected subgroups cannot be hired at a rate less than 80% that of the majority group (Bobko & Roth, 2004; US EEOC, 1978). As shown in the Table 2, the passing ratios of this test were within this criterion for all subgroups.

**TABLE 2**  
**INTEGRITY TEST SUBGROUP DIFFERENCES**

Group	<i>N</i>	Mean <sup>1</sup>	<i>SD</i>	<i>t</i>	<i>d</i>	4/5 <sup>th</sup> rule
Age	<25	70	2.13	0.54	0.09	95%
	≥25	131	2.21			
Gender	Male	99	2.04	-1.81	-0.25	119%
	Female	102	2.27			
National Origin	Israeli	173	2.17	0.54	0.11	87%
	Other	28	2.07			

\* $p < .05$ , \*\* $p < .01$ ; <sup>1</sup>Integrity scores ranged from 1 (low) to 3 (high).

Test fairness may also be measured by test bias, i.e., the differential prediction of the criterion between subgroups, which may lead to inaccurate or unfair selection decisions as a result of prediction errors (Aguinis & Smith, 2007). A series of moderated regression analyses were carried out for groups of gender, age, and country of origin, whereby a significant change in  $R^2$  after entering the subgroup variable and/or an interaction variable between the subgroup and the test would be an indication of test bias (Lautenschlager & Mendoza, 1986). Across all of the criteria and subgroups, none showed significant changes in  $R^2$  ( $p > .05$ ), indicating no evidence of test bias.

## DISCUSSION

This study was one of the first controlled investigations of integrity test validity in Israel. The scores from a widely used Israeli integrity test were studied against a set of criteria that can be described in terms of two main parameters: measurement method (self-report vs. behavioral) and construct orientation (interpersonal vs. organizational). Studying these variables together allowed for a more comprehensive look at the criterion-related validity of the test, as well as a unique opportunity to compare some of the various measures typically used to validate pre-employment integrity tests, which have seldom been examined collectively.

In terms of the descriptive statistics, it is worth noting that among actual applicant populations, the integrity test's score distributions are distinctly more positively skewed, with approximately 15% "not recommended" scores (Fine, 2008), compared to the 36% found here. This difference is reasonable, given both the lower stakes situation of this experiment and the explicit instructions to respond honestly, compared to actual occupational settings, which may otherwise have elevated scores, but in-turn resulted in a lower power for the analyses. However, as a result, the findings here may not be directly applicable to true high stakes selection scenarios. In terms of the criteria, the rates of self-reported CWB were in accordance with the literature, which report the majority of employees to have engaged in at least some degree of CWB (Murphy, 1993). In addition, the rate at which participants chose not to return the extra payments (roughly 30%) is in line with an estimated rate of employee thefts overall (Hollinger & Clark, 1983) as well as the return rates of similarly designed experimental trials (Cunningham et al., 1994).

### Validity in Israel

The overall validity results are congruent with those from prior meta-analyses, in terms of the approximate magnitudes of the coefficients for both the self-report and external criteria (e.g., Cunningham et al., 1994; Ones et al., 1993; Sackett & Harris, 1984). By replicating prior findings from primarily North American data in this way, this study has shown (to at least some extent) that integrity test validities may be transportable to the Israeli workforce. Indeed, according to Lievens (2008), the validity of broad constructs, such as integrity, should generalize cross-culturally and be transportable, as the present results would suggest.

To be sure, individuals with low overt integrity test scores are not necessarily considered to be unethical or dishonest. This would neither accurately reflect the construct measured, nor the level of its validity. Overt integrity tests measure opinions towards CWB, such that extreme opinions indicate a greater risk of involvement in CWB, relative to those without such opinions. Although assessing risk in this way may be an effective method to reduce CWB, a significant percentage of low scorers will still not be involved in any form of deviant behaviors (Sackett & Wanek, 1996). While this (i.e., false positives) may be true for any selection device, due to possible stigmatizations that may result from misinterpreting integrity results in this way, practitioners should be especially careful.

In light of the criticisms surrounding the use of self-report criteria of CWB in integrity test validity studies (Berry et al., 2007), the fact that this study included criteria of both self-reports (i.e., CWB and prior dismissals) and behavioral observations (i.e., theft and deception) is especially significant. Specifically, based on the validity indices, the findings of this study imply that despite their methodological challenges, admissions may be a viable alternative criterion for initial validity evidence when external criteria are not available. In addition to the integrity test validity, this contention is also based on the inter-relationships between the nature of the criteria themselves. Convergence was found between the two "interpersonal" criteria, and also between the two "organizational" criteria. That is to say, the behavioral scenario of deception (an interpersonal offense) was positively related to self-reported variable of CWB-I, and the behavioral scenario of theft (an organizational offense) was positively related to self-reported variable of CWB-O. Together, these relationships, while moderate in magnitude, provide important reliability evidence, while also supporting prior findings that self and other-reported CWB measures significantly converge on at least some facets (Fox, Spector, Goh, & Bruursema, 2007).



An interesting point that was brought up by the U.S. Congress Office of Technology Assessment report (1990) on integrity testing regarding admissions criteria is worthy of discussion. They ask that if self-reported past behaviors are to be considered good criteria of CWB for studying integrity test validities, companies should consider using these admissions themselves for selection purposes, instead of integrity tests (p. 52). In response, it should first be noted that admissions surveys represent just one possible criterion for demonstrating integrity test validity, and are by no means ideal, due to the methodological limitations already mentioned above. Second, there are a number of reasons why organizations may not wish to administer admission surveys of this kind to job applicants: a) their items are fully transparent and may appear to be more prone to faking than integrity tests; b) their items may be perceived to be more invasive or unfair than integrity tests, and elicit more negative applicant reactions as a result; c) they rely on applicants having sufficient past job experience, and may therefore be inappropriate for new or young hires. Finally, several integrity tests include separate admissions scales for this very purpose (Sackett et al., 1989), but certainly not as substitutes for the other scales.

Regarding another issue related to self-reported criteria, by recognizing some degree of content overlap between the predictor criterion in these cases, the relationship between the two may be considered one of construct validity rather than criterion validity (Ones et al., 1993). As such, the magnitude of the correlation coefficients found here (.20s - .30s) may be considered low for construct validity. On the other hand, many integrity test scores (including the one studied here) are derived from additional measures, beyond the directly self-reported item responses (e.g., response latencies, lie scales, avoidance responses, consistency measures, etc.). Therefore, since the content overlap represents only part of the test score's measurement, high construct validities are not necessarily to be expected. In order to examine this point more thoroughly, new integrity scores were created based only on the direct item responses, and yielded substantially higher construct validities than the original test scores. Specifically, validity significantly increased from -.29 to -.49 for the self-reported CWB criterion ( $t = 2.98, p < .01$ ), although only slightly for the self-reported dismissal criterion (-.18 to -.22,  $t = .53, p > .05$ ). Similarly, the correlation with the behavioral theft criterion increased slightly (and not significantly) using the normative-only scores (from -.20 to -.26,  $t = .78, p > .05$ ), and the deception criterion remained approximately zero. While the improvement was not consistent, from this, one might be tempted to prefer the use of the test's normative scoring method. However, this conclusion would be premature for reason that the present results were achieved in a controlled laboratory study, whereas in real occupational settings, job applicants have stronger incentives to try and "fake" the test. Therefore, the additional scoring measures above and beyond the direct responses are perhaps critical to dealing with faking in real occupational settings - whether or not they are successful in doing so, is unfortunately debatable (Griffith & Peterson, 2006), and a question for another study.

The reason why integrity was not able to significantly predict the behavioral deception criterion is not entirely clear. While the literature has reported comparably low validities for similar deception scenarios (see Berry et al., 2007), at least three possible explanations can be offered for this study's results. First, unlike the theft scenario, where participants had a monetary incentive to "steal", participants had no strong incentive to be deceitful on their PET scores, especially given the fact that they were promised anonymity in the study. Second, at least some of the participants may not have honestly remembered their true test scores and thereby unwillingly (rather than deceitfully) misreported their scores. Indeed, five participants reported *lower* scores than their actual scores, which might indicate such types of errors. Third, since participants were not explicitly asked to write down their "exact scores," it is possible that some of them decided to round their score for convenience. Accordingly, the majority of differences between reported and actual scores were less than 10 points (i.e.,  $1/8^{\text{th}}$  of an *SD*). To be sure, while it is unlikely that any one of the above explanations was responsible for the lack of validity in the deception criterion, it is certainly possible that a combination of them was able to depress the observed correlations to at least some degree.

### **The Utility of Integrity Testing**

Whenever examining the validity of a selection solution in an occupational setting, it is important to consider its monetary utility. Utility refers to the financial benefit the selection solution has on the organization (i.e., its return on investment, or ROI), and is calculated as the monetary return from using the test minus its costs (Cascio, 1998). While utility is typically considered in terms of *increased* productivity gained from high performing employees hired via the selection system, here we relate to it in terms of counter-productivity *saved* from not hiring potentially problematic employees. In terms of the operational usage of integrity testing, this study implies a significant financial utility, whereby individuals who were “not recommended” by the test were two to three times more likely to have been involved in deviant behaviors than individuals who were “recommended”. Accordingly, as deviant work behaviors can lead to substantial financial losses to an organization, when used in the selection process, integrity tests can be seen as a potentially valuable method to mitigate these losses.

With that, it should be clear that not all utility estimates adequately take into consideration the affects of other acts of CWB (e.g., sabotage, absenteeism, negligence, etc.), turnover costs (which are at least 50% of an employee’s annual salary; Johnson, Griffeth, & Griffin, 2000), lost productivity, possible litigation fees, lowered morale, etc., and thus underestimate the true utility of integrity testing. Indeed, according to Ones and Viswesvaran (2001), “it would be a mistake to continue focusing on theft prediction as the only criterion to judge the usefulness of integrity tests in the personnel selection domain” (p. 33). Therefore, the benefits of prevented CWB via integrity testing may be even more substantial than some utility estimates would indicate. In all events, it is critical for organizations to understand the financial implications integrity testing can have.

### **Subgroup Differences and Test Bias**

A final issue investigated in this study was integrity test fairness. As described above, no evidence was found that might have led to adverse impact in applied settings, for the available groups of gender, age, or national origin. As a result, these results indicate that the prior research in this area carried out primarily in the United States (e.g., Ones & Viswesvaran, 1998b), may be generalizable cross-culturally. The results also add to prior findings in terms of studying differences among very young individuals as opposed to older ones, an issue raised previously in the literature (Ones & Viswesvaran, 1998b). Moreover, based on these results, when used together with traditional assessment tools, it is reasonable to assume that integrity testing in Israel can actually help alleviate some of the adverse impact commonly exhibited by cognitive based tests (Hunter & Hunter, 1984; Ones et al., 1993). Finally, while more specific local Israeli minority groups were not studied here, prior evidence of non-test bias between Arab and Jewish Israelis using this test has been established elsewhere (Fine & Basis, 2011).

### **Concluding Remarks**

Overall, this study provides initial empirical evidence for the validity, utility and fairness of integrity testing in Israel, and implies that integrity tests are likely to be effective tools for use in personnel selection in international settings. While the investigation of additional samples, tests, criteria, and longitudinal predictive research paradigms is still required before final conclusions can be made, the immediate findings are supportive of the many organizations currently using integrity tests in Israel today.

### **NOTE**

An earlier version of this paper was accepted for publication (in Hebrew) in the Israeli behavioral science journal, *Megamot*.

### **REFERENCES**

Aguinis, H., & Smith, M. A. (2007). Understanding the impact of test validity and bias on selection errors and adverse impact in human resource selection. *Personnel Psychology*, *60*, 165–199.

- Ajzen, I., & Fishbein, M. (1977). Attitude-behavior relations: A theoretical analysis and review of empirical research. *Psychological Bulletin*, *84*, 888-918.
- American Management Association. (2002). *Corporate values survey*. Retrieved May 28, 2009, from <http://www.amanet.org/research>.
- Bartram, D., Lindley, A. P., Marshall, L., & Foster, J. (1995). The recruitment and selection of young people by small businesses. *Journal of Occupational and Organizational Psychology*, *68*, 339-358.
- Becker, T. E., & Colquitt, A. L. (1992). Potential versus actual faking of a biodata form: An analysis along several dimensions of item type. *Personnel Psychology*, *45*, 389-406.
- Beller, M. (1994). Psychometric and social issues in admissions to Israeli universities. *Educational Measurement: Issues and Practice*, *12*, 12-20.
- Bennett, R. J., & Robinson, S. L. (2000). Development of a measure of workplace deviance. *Journal of Applied Psychology*, *85*, 349-360.
- Berry, C. M., Ones, D. S., & Sackett, P. R. (2007). Interpersonal deviance, organizational deviance, and their common correlates: A review and meta-analysis. *Journal of Applied Psychology*, *92*, 409-423.
- Berry, C. M., Sackett, P. R., & Wiemann, S. (2007). A review of recent developments in integrity test research. *Personnel Psychology*, *60*, 271-301.
- Bobko, P., & Roth, P. (2004). The four-fifths rule for assessing adverse impact: An arithmetic, intuitive, and logical analysis of the rule and implications for future research. In J. Martocchio (Ed.), *Research in personnel and human resources management* (Vol. 23, pp. 177-197). New York: Elsevier.
- Camara, W. J., & Schneider, D. L. (1994). Integrity tests: facts and unresolved issues. *American Psychologist*, *49*(2), 112-119.
- Cascio, W. F. (1998). *Applied psychology in human resource management (5th ed.)*. Englewood Cliffs, NJ: Prentice-Hall.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences (2nd ed.)*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Cunningham, M. R., Wong, D. T., & Barbee, A. P. (1994). Self presentation dynamics on overt integrity tests: experimental studies of the Reid Report. *Journal of Applied Psychology*, *79*(5), 643-659.
- Fine, S. (2008). *IntegriTEST: Technical manual 3.0*. Midot, Ltd.
- Fine, S. (2010, July). *Cross-cultural validity of integrity testing: A tale of three banks*. Paper presented at the 27th International Congress of Applied Psychology, Melbourne, Australia.
- Fine, S., & Basis, L. (2011, May). *Fairness between Arabs and Jews in pre-employment integrity testing*. Paper presented at the 15th annual conference of the European Association of Work and Organizational Psychology, Maastricht, The Netherlands.

- Fine, S., Horowitz, I., Weigler, H., & Basis, L. (2010). Is good character good enough? The effects of situational variables on the relationship between integrity and counterproductive work behaviors. *Human Resource Management Review*, 20(1), 73-84.
- Fortmann, K., Leslie, C., & Cunningham, M. (2002). Cross-cultural comparisons of the Reid integrity scale in Latin America and South Africa. *International Journal of Selection and Assessment*, 10, 98-108.
- Fox, S., Spector, P. E., Goh, A., & Bruursema, K. (2007). Does your coworker know what you're doing? Convergence of self- and peer-reports of counterproductive work behavior. *International Journal of Stress Management*, 14(1), 41-60.
- Fulmer, R. M., & Conger, J. A. (2004). *Growing your company's leaders: How organizations use succession management to sustain competitive advantage*. New York: AMACOM.
- Goldberg, L. R., Grenier, J. R., Guion, R. M., Sechrest, L. B., & Wing, H. (1991). *Questionnaires used in the prediction of trustworthiness in pre-employment selection decisions: an APA task force report*. Washington DC: American Psychological Association.
- Gottfredson, L. S. (2002). Where and why g matters: Not a mystery. *Human Performance*, 15(1/2), 25-46.
- Griffith, R. L., & Peterson, M. H. (2006). *A closer examination of applicant faking behavior*. Greenwich, CT: Information Age Publishing.
- Guion, R. M. (1998). *Assessment, measurement and prediction for personnel decisions*. Mahwah, NJ: Lawrence Erlbaum Associates
- Harold, C. M., McFarland, L. A., & Weekly, J. A. (2006). The validity of verifiable and non-verifiable biodata items: An examination across applicants and incumbents. *International Journal of Selection and Assessment*, 14(4), 336-346.
- Herriot, P., & Anderson, N. (1997). Selecting for change: How will personnel and selection psychology survive? In N. R. Anderson & P. Herriot (Eds.), *International handbook of selection and assessment*. London: Wiley.
- Hollinger, R. C., & Clark, J. P. (1983). *Theft by employees*. Lexington, MA: Lexington.
- Hunter, J. E., & Hunter, R. F. (1984). Validity and utility of alternative predictors of job performance. *Psychological Bulletin*, 96, 72-98.
- Israel Psychological Association. (2004). *Professional code of ethics for psychologists in Israel*. Retrieved December 24, 2009, from [www.hebpsy.net](http://www.hebpsy.net).
- Israeli Ministry of Industry, Trade and Labor. (1988). *Employment (equal opportunities) law*. Retrieved December 24, 2009, from [www.tamas.gov.il](http://www.tamas.gov.il).
- Johnson, J., Griffith, R. W., & Griffin, M. (2000). Factors discrimination functional and dysfunctional sales force turnover, *Journal of Business & Industrial Marketing*, 15(6), 399-415.
- Jones, J. W. (1991). *Preemployment honesty testing: current research and future directions*. New York: Quorum.

- Kouzes, J. M., & Posner, B. Z. (2009). To lead, create a shared vision. *Harvard Business Review*, 87(1), 20-21.
- Lachman, R., & Shpigel, S. (2004). *Counterproductive behavior in organizations: Opinion survey 2001-2003*. Technical Report #2.04. Transparency International, Israel.
- Lautenschlager, G. J., & Mendoza, J. L. (1986). A step-down hierarchical multiple regression analysis for examining hypotheses about test bias in prediction. *Applied Psychological Measurement*, 10, 133-139.
- Lievens, F. (2008). Research on selection in an international context: Current status and future directions. In M. M. Harris (Ed.), *Handbook of research in international human resource management* (pp. 107-123). New York: Lawrence Erlbaum Associates.
- Marcus, B., Lee, K., & Ashton, M. C. (2007). Personality dimensions explaining relationships between integrity tests and counterproductive behavior: Big Five, or one in addition? *Personnel Psychology*, 60, 1-34.
- Martin, S. L., & Orban, J. A. (1995). Are fundamental job requirements neglected in selection system? *Journal of Business and Psychology*, 9(4), 345-353.
- Mikulay, S. M., & Goffin, R. D. (1998). Measuring and predicting counterproductivity in the laboratory using integrity and personality testing. *Educational and Psychological Measurement*, 58, 768-790.
- Miner, J. B., & Capps, M. H. (1996). *How honesty testing works*. Westport, CT: Quorum.
- Murphy, K. R. (1993). *Honesty in the workplace*. Pacific Grove, Ca: Brooks/Cole Publishing.
- Ones, D. S., & Viswesvaran, C. (1998a). The effects of social desirability and faking on personality and integrity assessment for personnel selection. *Human Performance*, 11(2/3), 245-269.
- Ones, D. S., & Viswesvaran, C. (1998b). Gender, age and race differences on overt integrity tests: Results across four large-scale job applicant datasets. *Journal of Applied Psychology*, 83(1), 35-42.
- Ones, D. S., & Viswesvaran, C. (2001). Integrity tests and other criterion-focused occupational personality scales (COPS) used in personnel selection. *International Journal of Selection and Assessment*, 9, 31-39.
- Ones, D. S., Viswesvaran, C., & Reiss, A. D. (1996). Role of social desirability in personality testing for personnel selection: The red herring. *Journal of Applied Psychology*, 81, 660-679.
- Ones, D. S., Viswesvaran, C., & Schmidt, F. L. (1993). Comprehensive meta-analysis of integrity test validities: Findings and implications for personnel selection and theories of job performance. *Journal of Applied Psychology (Monograph)*, 78, 679-703.
- Nevo, B., & Oren, C. (1986). Concurrent validity of the American Scholastic Aptitude Test (SAT) and the Israeli Inter-University Psychometric Entrance Test (IUPET). *Educational & Psychological Measurement*, 46(3), 723-725.
- Nicol, A. M., & Paunonen, S. V. (2002). Overt honesty measures predicting admissions: An index of validity or reliability. *Psychological Reports*, 90, 105-115.

- Robinson, S. L., & Bennett, R. J. (1995). A typology of deviant workplace behaviors: A multidimensional scaling study. *Academy of Management Journal*, 38, 555-572.
- Ryan, A. M., McFarland, L., Baron, H., & Page, R. (1999). An international look at selection practices: Nation and culture as explanations for variability in practice. *Personnel Psychology*, 52, 359-391.
- Sackett, P. R., & Devore, C. J. (2001). Counterproductive behaviors at work. In N. Anderson, D. Ones, H. Sinangil, & C. Viswesvaran (Eds.), *Handbook of industrial, work, and organizational psychology, 1*, 145-164.
- Sackett, P. R., & Harris, M. M. (1984). Honesty testing for personnel selection: A review and critique. *Personnel Psychology*, 37, 221-246.
- Sackett, P. R., & Wanek, J. E. (1996). New developments in the use of measures of honesty, integrity, conscientiousness, dependability, trustworthiness, and reliability for personnel selection. *Personnel Psychology*, 49, 787-829.
- Sackett, P. R., Burris, L. R., & Callahan, C. (1989). Integrity testing for personnel selection: an update. *Personnel Psychology*, 42, 491-529.
- Schmidt, F. L., & Hunter, J. E. (1977). Development of a general solution to the problem of validity generalization. *Journal of Applied Psychology*, 62, 529-540.
- Schmidt, F. L., & Hunter, J. E. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin*, 124, 262-274.
- Scholarios, D., & Lockyer, C. (1999). Recruiting and selecting professionals: contexts, qualities and methods. *International Journal of Selection and Assessment*, 7, 142-169.
- Society for Industrial and Organizational Psychology (SIOP). (2003). *Principles for the validation and use of personnel selection procedures (4<sup>th</sup> ed.)*. Bowling Green, OH: Society for Industrial and Organizational Psychology.
- Solomon, I. (2010). Looking for work? Leave your management skills at home [in Hebrew]. *The Marker*. Retrieved from August 5, 2010, from [www.themarker.com](http://www.themarker.com).
- Stuman, M. C., & Sherwyn, J. D. (2007). *The truth about integrity tests: The validity and utility of integrity testing for the hospitality industry*. The Center for Hospitality Research, Cornell University.
- U.S. Congressional Office of Technology Assessment. (1990). *The use of integrity test for pre-employment screening* (Report No. OTA SET-442). Washington, D.C.: U.S. Government Printing Office.
- U.S. Equal Employment Opportunity Commission. (1978). *Uniform guidelines of employee selection procedures*. Federal Register 43, 38290-38315.
- Van Iddekinge, C. H., Roth, P. L., Raymark, P. H., and Odle-Dusseau, H. N. (in press). The criterion-related validity of integrity tests: An updated meta-analysis. *Journal of Applied Psychology*.
- Wanek, J. E. (1999). Integrity and honesty testing: What do we know? How do we use it? *International Journal of Selection and Assessment*, 7(4), 183-19.