Development and Validation of the Select Assessment® for Employee Reliability

Authors: Amie D. Lawrence, Ph.D., Kristin Delgado, M.S., Doug Wolf, M.S., Mavis Kung, Ph.D. & Matthew O'Connell, Ph.D.

SELECT PERSPECTIVES

Background

Estimates of the cost of employee theft range from $10 to $200 billion dollars annually ("Employee Theft: Legal Aspects - Estimates Of Cost," n.d.). Very little is known about the true prevalence of employee theft or exactly how costly it is for employers. What we do know is that it is expensive and a key issue for almost all organizations. I/O psychologists have been helping to identify individuals who are more likely to engage in theft and other counterproductive work behaviors for many years through the use of integrity tests. The predictive value of integrity tests has been established (e.g., Ones, Viswesvaran & Schmidt, 1993) in the academic literature.

Focus of Paper

This paper primarily focuses on the development, scientific underpinnings and appropriate application of the Select Assessment® for Employee Reliability assessment. As a starting point, we present a model of integrity and its relationship to employee theft and counterproductive behaviors. We then present a study that evaluates the effectiveness of this tool in predicting employee theft and other counterproductive behaviors.

Defining Integrity

Integrity researchers tend to disagree to some extent as to what integrity is, but they do seem to agree that it is not a unidimensional construct. Integrity seems to be a compilation of subfactors that work together to predict counterproductive or unethical work behaviors. Some researchers have compared the items in standard integrity tests to the five factor model of personality and identified relationships with C (Conscientiousness), A (Agreeableness) and ES (Emotional Stability) (Marcus, Hoft & Riediger, 2006; Murphy & Lee, 1994; Ones, 1993). In an attempt to better understand the factor structure of well-known integrity tests, Wanek, Sackett and Ones (2003) conducted a judgmental sort of 798 items from seven published integrity tests and found 23 distinct composites. A principal components analysis of these 23 indicated four components to integrity: antisocial behavior (e.g., theft admissions, association with delinquents), socialization (e.g., achievement orientation, locus of control), positive outlook (e.g., viewing people as basically good and the world as basically safe) and orderliness/diligence. Van Iddekinge, Taylor and Eidson (2005) also rationally sorted the honesty scale in PSI customer service assessment (PSI-CS), and reported similar results to Wanek et al. (2003). Other researchers feel that these four factors do not adequately cover integrity. Connelly, Lilienfeld and Schmeelk (2006) found integrity test scores related less to moral reasoning and more towards psychopathic personality, which includes such as behaviors of self-centered nature, willingness to manipulate others, externalizing blame and an impulse to flout social norms. Lee, Ashton and colleagues have introduced the HEXACO model of integrity (Lee, Ashton & de Vries, 2005; Lee, Ashton & Shin, 2005). Their research suggests that Honesty-Humility (H-H) should be added to the five factor model to adequately cover all aspects of integrity. The low end of this sixth construct resembles many of the psychopathic behaviors described by Connelly et al. (2006). Future research should continue to examine the underlying make up of integrity.

Measuring Integrity

A majority of integrity measurements fit cleanly into two categories: overt and covert. Overt measures of integrity ask individuals direct questions about their past unethical behavior or their attitudes towards unethical behavior. There is no subtlety or pretense to these items. Individuals find them to be very face valid and tend not to have strong negative reactions to them (Berry, Sackett & Weinman, 2007; Jones, Ash & Soto, 1990). Covert integrity items are typically personality-based and are more indirect in their connection to counterproductive behavior. Commonly used covert personality items might ask about one's impulsivity or general outlook on people and the world. These personality constructs are supported empirically by being significantly related to counterproductive work behavior, but test takers are less likely to identify them as integrity items.

At the time of this paper, only two published studies were found that chose to look outside of these standard measurements and develop situational judgment items to measure integrity. Becker (2005) developed a situational

SELECT INTERNATIONAL

judgment test of employee integrity. He found significant relationships with his test and managerial ratings of career potential, leadership activities and job performance. A video situational judgment test was developed to specifically measure integrity in a sample of Dutch policemen (de Meijer, Born, van Zielst & van der Molen, 2010). A confirmatory factor analysis showed solid construct validity for the SJT measure.

What Does Integrity Predict?

From the beginning, integrity tests were developed with counterproductive behaviors in mind. The main goal of this kind of testing is to prevent unethical and undesirable behavior in the workplace. Extensive reviews since the late 1980's have supported the validity of integrity testing, regardless of type (e.g., Ones, Viswesvaran & Schmidt, 1993; U.S. OTA, 1990; Sackett & Wanek, 1996; APA Goldberg et al., 1991). Ones et al. (1993)'s meta-analysis remains to be the most comprehensive quantitative review to this day. They meta-analyzed 665 validity studies and reported similar validity for the overt and covert tests in predicting job performance (.33 for overt vs. .35 for covert). Their results did find that overt tests tend to predict counterproductive behaviors (e.g., absenteeism), excluding theft behavior, better than covert tests (.39 for overt vs. .29 for covert). Additionally, overt tests significantly predicted theft behavior (determined from external sources) (.13) as well as admissions to theft (.42).

At the time of the Ones et al. paper, an insufficient number of studies were available to examine the relationship between theft and covert integrity tests. Our review located only one empirical study by Neuman and Baydoun (1998) that compared overt and covert tests with regard to theft. Their results indicated that covert tests were better than overt in predicting theft (.22 overt vs. .37 covert). They also reported higher correlations with counterproductive behaviors for covert test as compared to overt (.31 overt vs. .42 covert). The use of a student sample may limit the generalizability of these results. More research is needed to determine the relationship between covert measures of integrity and theft behavior.

Our Model

Taking into account the results of Wanek et al. (2003) and the suggestion by other researchers to examine psychopathic personality traits, we identified six factors that we felt would adequately cover the "construct" of integrity and maximize our prediction of counterproductive work behaviors. We chose to add a narcissism factor as it is a key component of a psychopathic personality (Harpur, Hakstian & Hare, 1988). We also chose to pull impulsivity out as a separate factor. Some of the results in the Wanek et al. (2003) study indicated that impulsivity could have been contributing to some intercorrelations. Additionally, impulsivity has been shown to be related to counterproductive work behaviors in previous research (Henle, 2005). Based on a review of the extant research we came up with the following model of integrity which covers the underlying competencies, the measurement model and the behavioral outcomes.

Assessment Methodologies

After the underlying model of integrity had been established, the next step in the development of the Select Assessment for Employee Reliability was to determine the most appropriate means of measuring those key factors. It's not enough to have a good theoretical model. To make that model truly useful requires designing methodologies for accurately and reliably measuring the variables in the model.

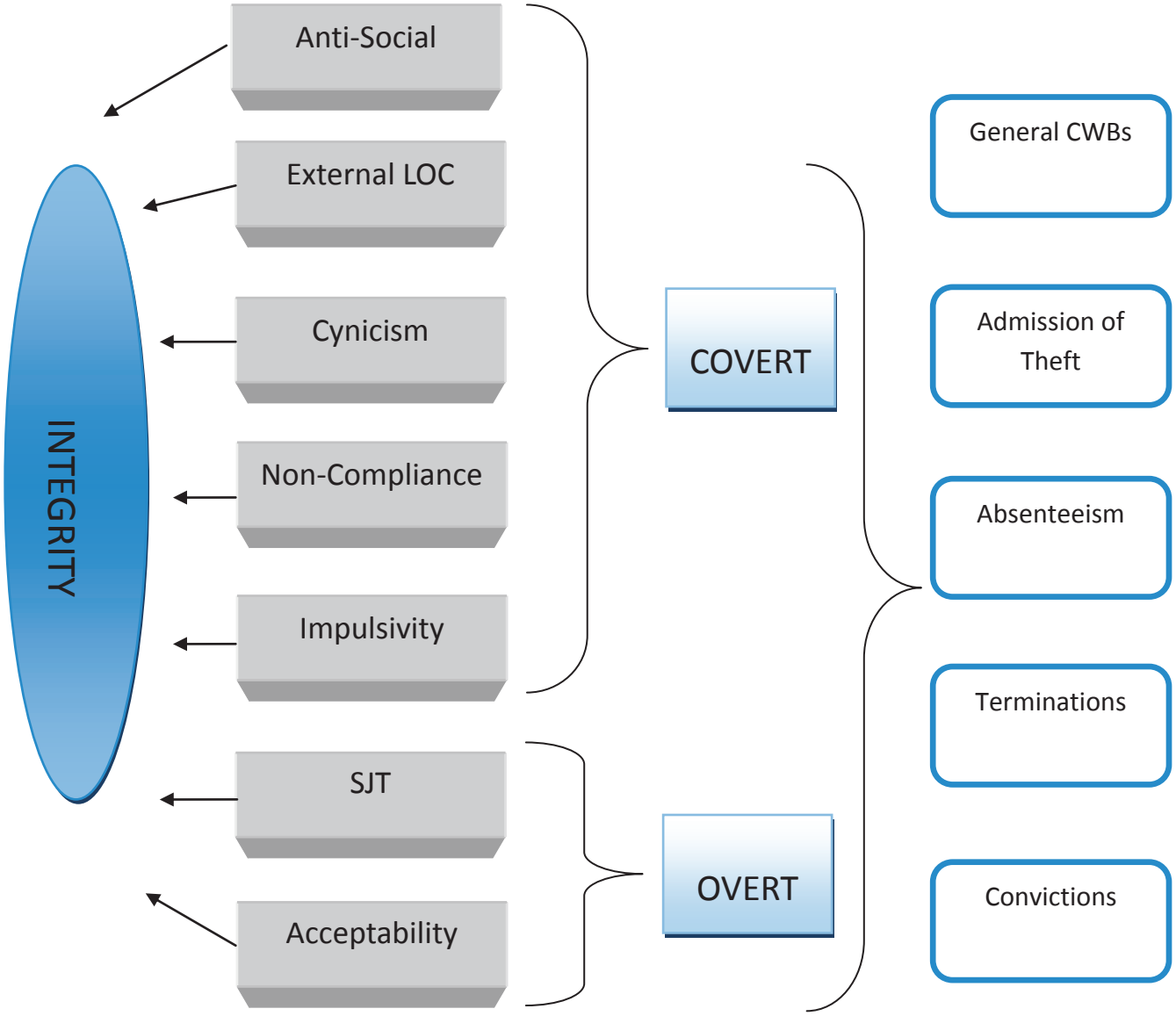
No single assessment method is acceptable for measuring every variable. Some assessment methods, e.g., social intelligence tests, personality inventories, cognitive ability tests, tests of knowledge, interactive simulations or interviews, are better at measuring different competencies than are others. Using multiple tools for evaluating a competency not only increases the reliability or consistency of the assessment but also increases the validity or accuracy of the assessment. A person may do poorly or may do well on one particular assessment tool due to a variety of reasons, some related to their actual ability or skill level and some not related to their skill or ability. The pattern of their scores across a variety of well-developed tools is going to provide the best assessment of that variable.

In the Select Assessment for Employee Reliability we utilize three primary assessment methodologies to measure the key variables in the model. These include:

- Self-Report Personality Scales
- Behavioral Self-Report Scales
- Situational Judgment Scales

Each of these methods is discussed in more detail in the next section.

Final Integrity Measurement Model



Self-Report Personality Scales

Since the early 1990's a large body of research has consistently found that self-report personality measures are valid predictors of job performance across a variety of job settings (cf. Barrick, Mount & Judge, 2001; Hurtz & Donovan, 2000). The field of Industrial/ Organizational psychology has been particularly interested in the usefulness of the Big Five personality dimensions (aka the Five Factor Model, or FFM) in predicting performance (cf. Barrick & Mount, 1991; Digman, 1991; Tett, Jackson & Rothstein, 1991). The FFM provides a useful model for describing normal personality structure, but many components of it are not particularly useful (or at least complete) in predicting safety behavior or accidents. For instance, as mentioned earlier, only three of the five Big Five factors, Conscientiousness, Agreeableness and Emotional Stability are related unethical work behaviors and integrity tests (Marcus et al., 2006; Murphy & Lee, 1994; Ones, 1993).

Additionally, the Big Five model is incomplete when it comes to predicting integrity and unethical behavior. It fails to cover important traits such as locus of control, cynicism and a number of factors traditionally associated with psychopathic personality (cf. Connelly et al., 2006).

At Select International, we have taken a philosophy in which we utilize personality scales, or any assessment methodology for that matter, that are effective at measuring the variables of interest to our models. The FFM is a useful, general model. As noted above, however, it is neither complete nor particularly useful when attempting to maximize the prediction of factors related to integrity, employee theft or counterproductive work behaviors. We have used self-report personality items in our assessments since 1994. We have a large database of established items that have been validated in dozens of studies and taken by hundreds of thousands of individuals.

For the Select Assessment for Employee Reliability assessment we took a multi-step approach to creating the personality scales. The first step was to identify existing items that would work in the model. We then reviewed analyses related to the performance of those existing items in terms of their accuracy, their distributions in large applicant samples of information and if there was any evidence of differential item functioning (DIF) against minorities. The best functioning items were selected, i.e., the most accurate with the most variance across the entire range of responses. Any items that demonstrated DIF were also removed. This resulted in a set of personality items covering a majority of the key sub-factors. It was not a complete model, however. There were some gaps in covering the subfactors. To cover these gaps we undertook the process of writing new items.

A team of five I/O psychologists (three Ph.D.'s & two M.S. level) wrote a series of new personality items that tapped the sub-factors in the integrity model. Initial item pools were reviewed by the team and the best items were selected using expert judgment. As with our other assessments, all personality items are rated on a Likert-type rating scale. This type of rating scale asks candidates to read a statement and rate their level of agreement on a 6-point scale ranging from Strongly Agree to Strongly Disagree. Likert-type scales are very common for the measurement of personal beliefs. Research has shown that personality scales written in this manner are consistently positively related to performance in a variety of settings (Barrick & Mount, 1991). The final set of personality items covering the entire assessment model was comprised of 116 items.

While personality measurement has been around for many years, there are a number of critics who feel that the use of self-report personality items in a personnel selection context is inappropriate (Guion & Cranny, 1982; Morgeson et al., 2007). The concern is that applicants are given the opportunity to misrepresent themselves to the organization or "fake" the test items. We concede that any type of self-report test item provides this opportunity to an applicant; however, the academic research in this area has not supported the concerns. Many studies have shown personality scales to be valid predictors of performance in applicant settings (Barrick & Mount, 1991,1996; Ones, Viswesvaran & Schmidt, 1993; Tett et al., 1991; see more in a review by Ones, Dilchert, Viswesvaran and Judge, 2007). There is some evidence to suggest that assessment validity may be minimally suppressed by the occurrence of socially desirable responding (Barrick & Mount, 1996; Hough, Eaton, Dunnette, Kamp & McCloy, 1990; Ones, Viswesvaran & Reiss, 1996). There is additional research that provides suggestions for minimizing the effects of misrepresentation and socially desirable responding, such as presenting different instructional sets (e.g., Converse et al., 2008; Dwight & Donovan, 2003; McFarland, 2003; Pace & Borman, 2006). We have applied many of these techniques over the years to obtain the maximum validity possible. As a leader in assessment technology, we conduct research and collaborate with other researchers to investigate the questions in this area of research. As part of an ongoing research program, we have collaborated with top researchers in this field to investigate and evaluate the best strategy for reducing the impact of applicant faking on hiring decisions. In general, numerous research studies have shown the following:

- The inclusion of corrections for Social Desirability[1] or other types of faking indexes do not improve validity (cf. Christiansen, Goffin, Johnston & Rothstein, 1994; Leite & Beretvas, 2005; Ones et al., 1996).
- The inclusion of Social Desirability corrections does not help reduce mean differences between demographic groups.
- Social Desirability is not significantly related to actual faking (Peterson, Griffith, Isaacson, O'Connell & Mangos, 2011).
- As of today, there is no methodology available that offers an acceptable level of accuracy for identifying applicants who purposefully inflate their scores on personality items.

At this time, there is wide scale acceptance in the field of I/O psychology that the use of personality in personnel selection settings provides predictive value and little to no adverse impact despite the issues with self-report methodology (Barrick & Mount, 2005; Hough & Dilchert, 2010). Select continues to conduct applied research and partner with academic researchers to stay at the forefront of the research in this area.

Behavioral Self-Report Scales

Biodata has a long and successful history as an applicant screening device (Mumford & Stokes, 1992; Stokes & Cooper, 1994) but has evolved away from its origin, the scored application blanks, and bears more similarity to traditional personality measures. This change has occurred for largely two reasons. First, the legal implications of empirically scored questions found on application blanks such as "Do you own your home?" or "What was the size of your high school?" are obvious. Second, empirically keyed biodata items have been widely criticized for their lack of theoretical contribution to our understanding of what actually predicts performance (e.g., Nickels, 1994; Pace & Schoenfeldt, 1977; Tenopyr, 1994). To aid in this task, a taxonomy describing the attributes of biodata items has been developed by Mael (1991) and a construct-oriented approach to scale development has been advocated (Hough & Paullin, 1994).

Not only has biodata been a consistent predictor of job related criteria (Hunter & Hunter, 1984; Reilly & Chao, 1982), some research has indicated that biodata items may be resistant to faking (Atwater, 1990; McManus & Masztal, 1993) especially when a warning is used (Schrader & Osburn, 1977). Additionally, biodata scales have been shown to add significantly to the prediction of performance in a broad range of jobs from clerical through sales, army recruits through air traffic controller above and beyond cognitive ability and the five factor personality scales (cf. Dean & Russell, 1998; Mael & Ashforth, 1995; Allworth & Hesketh, 1998; Mount, Witt & Barrick, 2000). There is also evidence that empirically-keyed biodata scales demonstrate a high level of generalizability across organizations, demographics, tenure and work experience (Rothstein, Schmidt, Erwin, Owens & Sparks, 1990; Carlson, Scullen, Schmidt, Rothstein & Erwin, 1999).

Behavioral self-report items that are used in the Select Assessment for Employee Reliability fall under the general rubric of biodata items. We refer to this as an "acceptability of behavior" measure consisting of 30 behavioral items related to integrity. Participants are asked to rate each behavior in terms of the behavior's acceptability, using a 6-pt response scale (1=very unacceptable, 6=very acceptable). This measure was considered to be an overt measure, as specific attitudes about specific ethical/unethical behaviors were assessed. Given its success and historical prevalence as an employee selection instrument, we feel that the inclusion of such behavioral self-report items supplement that which can be obtained through personality measures.

Situational Judgment Scales

Situational judgment tests (SJT) are simulations requiring the respondent to exercise judgment when responding to hypothetical problem situations that occur in work settings. The use of SJTs dates back to the 1920's (McDaniel, Morgeson, Finnegan, Campion & Braverman, 2001). Procedures for developing this type of test item are discussed in several studies (Motowidlo, Dunnette & Carter, 1990; McDaniel & Nguyen, 2001; Smith & McDaniel, 1998).

SJTs have become popular measures for gathering respondent's knowledge of how to handle particular situations and/or their behavioral tendencies in these situations. The popularity of these instruments has led to research investigating the nature of the construct(s) measured by these items.

<u>What do SJTs measure?</u>

Some authors of SJTs have asserted that their tests measure a single construct (e.g., practical judgment [Cardall, 1942], managerial success [Campbell, Dunnette, Lawler & Weick, 1970], and tacit knowledge [Sternberg & Wagner, 1993]). On the other hand, three recent meta-analyses (McDaniel et al., 2001;

[1] *Most of our assessments contain a commonly used social desirability scale. The results are used strictly for academic research purposes and are not shared with clients or candidates. We continue to collaborate with researchers in this area to better understand the issue of applicant misrepresentation and socially desirable responding.*

McDaniel & Nguyen, 2001; Nguyen & McDaniel, 2001) have shown that situational judgment measures typically measure several well-established constructs including cognitive ability, conscientiousness, emotional stability and agreeableness. In addition to these constructs, job knowledge appears to have a relationship with SJTs as suggested by the modest correlations between job experience and SJTs. Although evidence supports the consistent relationship between these constructs and SJTs, the magnitude of correlations across studies varies substantially, even after correcting for artifacts. For example, although most SJTs have moderate correlations with cognitive ability, any given test might show a very large correlation with cognitive ability, or a very low correlation with cognitive ability. Since SJTs measure a variety of constructs and different tests assess these constructs to varying degrees, we join others in arguing that SJTs are best viewed as measurement methods and not measures of a single construct (Chan & Schmitt, 1997; Clevenger, Pereira, Wiechmann, Schmitt & Harvey, 2001; McDaniel & Nguyen, 2001; McDaniel et al., 2001; Nguyen & McDaniel, 2001; Weekly & Jones, 1999; O'Connell, Hartman, McDaniel, Grubb III & Lawrence, 2007).

Validity and Subgroup Differences of SJTs

SJTs have gained increasing popularity in recent years. This popularity has been driven both by the validity of the tests (McDaniel et al., 2001) and by findings of smaller mean differences among racial subgroups as compared to traditional cognitive ability tests (Motowidlo et al., 1990; Motowidlo & Tippens, 1993; Pulakos & Schmitt, 1996; Clevenger et al., 2001; O'Connell et al., 2007).

Since the validity of SJTs has been established (McDaniel et al. 2001), it is important to understand how these tests add utility to a selection battery by further understanding the relationship between SJTs and other constructs, such as personality and cognitive ability. Specifically, it is important to identify the incremental validity that can be attributed to these types of instruments over other established predictors used in employee selection (Chan & Schmitt, 2002). In five samples, Weekly and Jones (1997, 1999) found significant incremental validity for SJTs over cognitive ability and job experience. Clevenger et al. (2001) used three samples to examine the incremental validity of situational judgment measures over cognitive ability, conscientiousness, job experience and job knowledge and reported incremental validity in two of the three samples. Chan and Schmitt (2002) found an SJT to have substantial validity in predicting task performance, and overall job performance. However, the SJT used in that study reported an unusually small correlation with cognitive ability (r = -.02). O'Connell et al. (2007) found that the SJT added incrementally above cognitive ability as well as a combination of personality measures.

SJTs in the Select Assessment for Employee Reliability

Select set out to develop a series of integrity-related SJTs that would measure key factors of the assessment model. A team of five I/O psychologists (three Ph.D.'s & two M.S. level) with extensive experience in item development wrote 20 integrity-related scenarios encompassing 68 individual ratings. These items were then reviewed individually and rated for appropriateness and relevance to the integrity factor in question as well as rating it using the response scale that would be used by respondents. Items that were not uniformly agreed upon in terms of relevance were rewritten, removed or replaced. Items that had more than one point difference on a 6-point rating scale were also rewritten, removed or replaced. In the end, a total of 12 hypothetical scenarios were retained covering a total of 48 items.

There are multiple approaches to gathering responses from participants using SJTs. One approach is to ask respondents to pick the best and/or worst alternative, or the alternative they would most likely and/or least likely perform if they were in that situation (cf. Weekley, Ployhart & Holtz, 2006; Motowidlo, Dunnette & Carter, 1990). SJTs, such as these, that force applicants to choose one or two response options from a larger set are typically scored according to how effective subject matter experts judged the chosen response options to be. Applicants earn high SJT scores when options they pick as best or as responses they would most likely carry out are options that subject matter experts (SMEs) judged to be high in effectiveness. They also earn high SJT scores when options they pick as worst or as responses they would least likely carry out are options that SMEs judged to be low in effectiveness.

Instead of asking for a forced choice between multiple response options, some SJTs ask applicants to rate all the options for their effectiveness (e.g., Chan & Schmitt, 2002; O'Connell et al., 2007). Effectiveness ratings can be scored in various ways that involve a comparison between applicants' ratings and SMEs' judgments of response option effectiveness. Applicants earn high SJT scores when their ratings of option effectiveness are similar to SMEs' judgments.

Over the past 15+ years, we at Select have experimented with multiple formats for scoring SJTs, including the best/worst, a forced ranking, as well as rating either the effectiveness or their

level of agreement with the statements. Our research has shown that using a multiple response rating, such as a 6-point rating scale, provides the most variance and ultimately the highest level of accuracy and reliability. For this particular assessment, we utilized a 6-point agreement scale, ranging from Strongly Disagree to Strongly Agree.

The evidence for the effectiveness, in terms of predictive validity, reduced adverse impact and face validity is strong and consistent. We have included this SJT section in the Select Assessment for Employee Reliability to supplement and enhance the information from the other assessment methods. This measure is considered to be an overt measure because specific attitudes regarding integrity are assessed.

Validating the Select Assessment for Employee Reliability

As part of an ongoing effort to establish the validity of the Select Assessment for Employee Reliability the following study was conducted. This study looked at the criterion-related validity assessment.

One of our goals in the study was to evaluate and compare the efficacy of the overt and covert measures of integrity. We believe that both types of measures (overt and covert) are important to the prediction of unethical behavior. Each covert integrity sub-factor was also examined to better understand what aspects of integrity are leading to different types of undesirable work behavior.

Method

Sample

The basic psychometric data were gathered from a sample recruited through various media outlets (e.g., social media, blogs). Individuals were compensated $20 for participation in the study. Participants were first asked to respond to a pre-qualifying questionnaire. The pre-qualifying questionnaire consisted of 43 low integrity behaviors, including counterproductive work behaviors and criminal activity. Participants were asked to indicate whether they had ever engaged in each behavior in the past. From the prequalifying data, 200 individuals were invited to participate in the full pilot study. Ninety-six individuals responded and were included in the final sample.

Measures

Predictors

The Select Assessment for Employee Reliability, described earlier was utilized in the study.

Criteria

Counterproductive work behaviors. Data from the pre-qualifying questionnaire were used as the counterproductive work behavior criteria. The CWB measures consisted of 40 behavioral statements related to counterproductive work behaviors (e.g., played computer/phone games while at work). Participants were asked to indicate whether they had ever engaged in each behavior in the past. See Appendix A for a full list of these behaviors.

Theft behavior. Data from the pre-qualifying questionnaire were used as the theft behavior criteria, which consisted of seven behavioral statements related to theft behavior (e.g., stolen company equipment or merchandise). Participants were asked to indicate whether they had ever engaged in each behavior in the past.

Absenteeism. Attitudes toward absenteeism were assessed using a one-item measure: "How many unscheduled or unexcused absences per year do you think are acceptable?" Participants responded by entering one whole number.

Convictions and terminations. Participants were asked to indicate whether they had ever been 1) terminated from a job and 2) convicted of a crime.

Results

Scale Construction of the Personality-Based Measures

We first conducted a principle components analysis using Kaiser criteria with varimax rotation on the 116 personality-based items. The results yielded a six factor solution. We examined the rotated factor matrix and, in general, the principle components fit our general factor structure, with the exception of the Narcissism factor. Thus, we then examined the individual factor loadings to determine which items to retain for the final personality scales. Only items that obtained a factor loading of >.40 and did not show any significant cross-loadings were retained for the final scales. Means, standard deviations, intercorrelations and reliabilities for the final scales are shown in Table 1 below. These

results led us to reconsider the original proposed factor structure, specifically regarding the Narcissism factor. None of the Narcissism items appeared to fit into a common factor structure and no unique Narcissism factor emerged through our analyses. Thus, the Narcissism factor was eliminated from our original model. In order to further assess the structure of the data, we then conducted a confirmatory factor analyses using AMOS and maximum likelihood estimation to assess how well the data fit the revised five-factor model (Antisocial, External LOC, Cynicism, Noncompliance and Impulsivity). The results of the CFA indicated the data fit the model well (see Table 2).

Relationships with Criteria

The next step was to see how well our final scales predicted the criteria of interest. As previously mentioned, our personality-based measures were considered to be covert, as they did not specifically address behaviors or attitudes toward integrity, but instead assessed certain personality dimensions we consider to be related to integrity. Using these personality dimensions, we generated an overall Integrity composite score to examine how well the combination of all personality-related integrity measures would predict various integrity-related criteria. Thus, we first evaluated how well each of the personality-based integrity measures predicted the criteria on its own, as well as an overall unit-weighted composite score.

Table 3 shows the correlations between the final predictor scales and the five different criteria: Counterproductive work behaviors (CWBs), theft behaviors, terminations, convictions and absenteeism. In general, the personality-based measures strongly predicted all criteria with the exception of convictions. The only personality measure that was significantly related to convictions was cynicism.

Next we used a multiple regression approach to test whether the overt integrity measures could predict unique variance in the five criteria above and beyond the personality-based (covert) integrity measures. Thus, we regressed each outcome variable on the personality-based integrity composite score in the first step, and the overt measures (acceptability and situational judgment scores) in the second step. Results of these analyses are presented in Table 4. As shown, the overt measures contributed significant incremental variance for general CWBs (F (1, 91) = 4.38, p <.05) and theft (F (1, 91) = 3.87, p <.05), but not for terminations (F (1, 91) = 1.33, p >.05), convictions, (F (1, 91) = 1.26, p >.05) or absenteeism (F (1, 91) = 0.37, p >.05).

Table 1
Means, Standard Deviations, and Zero-Order Correlations for Final Scales (N = 96)

Variable	*M*	*SD*	1	2	3	4	5	6	7
Antisocial (20 items)	4.35	0.63	(.91)						
External LOC (10 items)	4.14	0.55	0.57*	(.76)					
Impulsivity (10 items)	4.35	0.71	0.61*	0.51*	(.74)				
Noncompliance (17 items)	4.53	0.55	0.73*	0.60*	0.58*	(.87)			
Cynicism (13 items)	3.69	0.61	0.55*	0.58*	0.41*	0.45*	(.85)		
Acceptability (30 items)	2.21	0.60	0.63*	0.56*	0.56*	0.69*	0.36*	(.93)	
Situational Judgment	4.92	0.67	0.46*	0.50*	0.48*	0.57*	0.26*	.78*	(.71)

*Note: Reliabilities are presented in parentheses on the diagonal. *p < .05.*

Table 2

Results of Confirmatory Factor Analyses					
Model	χ^2	df	CFI	RMSEA	SRMR
Antisocial	122.20	88	.94	.06	.04
External LOC	189.78	58	.94	.07	.05
Impulsivity	152.12	58	.91	.08	.08
Noncompliance	110.40	80	.96	.04	.04
Cynicism	196.55	67	.90	.08	.07
Note: CFI, comparative fit index; RMSEA, root mean square error or approximation; RSMR, standardized root mean square residual					

Table 3

Correlations between final predictor scales and criteria					
	CWBs	Theft	Terminations	Convictions	Absenteeism
Covert (Personality-based)					
Anti-social	.41*	.37*	.21*	.17	.26*
External LOC	.35*	.27*	.27*	.09	.23*
Impulsivity	.39*	.39*	.27*	.13	.12
Noncompliance	.44*	.45*	.18	.11	.25*
Cynicism	.29*	.14	.19	.20*	.18
Composite	.47*	.40*	.28*	.18	.24*
Overt					
Acceptability	.39*	.44*	.14	.21*	.17
Situational Judgment	.16	.27*	.03	.23*	.19

*$p < .05$

Discussion

The results of this study shed some light on the measurement of integrity as well as its relationship with multiple integrity-related criteria. First of all, we found support for a five factor personality-based solution of our covert integrity measure. The proposed factor of narcissism did not adequately fit the model and was removed; however, anti-social lifestyle, another key component of psychopathy, did remain in the model. These results provide support for the four factors identified by Wanek et al. (2003) and for the inclusion of impulsivity as a separate factor.

We then turned our attention to the measurement of integrity by investigating the covert factors and developing two overt measurements (acceptability and situational judgment). We were interested in determining the unique value that overt and covert measures play in predicting integrity-related criteria, with a specific look at a newly developed overt situational judgment measure. Our study provided a look at several integrity-related criteria: general CWBs, theft, terminations, convictions and absenteeism. Previous research has examined CWBs, theft and absenteeism, but few with both overt and covert measures. Our results show that covert integrity measures were significantly related to all of our criteria except convictions. Our integrity composite measure was not

significantly related to convictions, but this may have been due to low endorsement rate for that item (n=4; 4.2%). This study suggests that personality-based measures of integrity can provide a strong understanding of someone's propensity to engage many different kinds of unethical or undesirable work behaviors.

The validity of our overt measures was also examined. These measures were more strongly related to theft and convictions as compared to the covert measures. Additionally, the acceptability measure was significantly related to general CWBs. A series of multiple regressions determined that the variance in the overt measures provided unique prediction over and above covert measures in the prediction of general CWBs and theft. Our results provide evidence that the two types of integrity tests provide different predictive information about individuals. Used in conjunction, they provide a more accurate picture of one's propensity to engage in unethical behavior. Our situational judgment measure, while not related to all criteria, was significantly related to theft and convictions. The inclusion of an SJT measure for integrity is supported, as it helps to add significant variance over and above personality.

Overall, the results were quite positive and demonstrated that the Select Assessment for Employee Reliability is a valid predictor of employee theft and other counterproductive work behaviors.

Table 4
Results of Multiple Regression Analyses

	Model	Variables	B	SE	β	R^2	ΔR^2
DV: CWB	1	Personality	5.49	1.08	.47*	.21*	
	2	Acceptability	4.16	1.60	.43*	.26*	**.07***
		Situational Judgment	3.51	1.25	.41*		
DV: theft	1	Personality	1.34	.31	.40*	.16*	
	2	Acceptability	1.29	.47	.47*	.23*	**.07***
		Situational Judgment	.53	.37	.21		
DV: terminations	1	Personality	1.07	.32	.35*	.08*	
	2	Acceptability	.05	.12	.09	.10*	**.03**
		Situational Judgment	.13	.09	.24		
DV: convictions	1	Personality	.07	.04	.18	.03	
	2	Acceptability	.02	.06	.05	.06	**.03**
		Situational Judgment	.05	.05	.16		
DV: absenteeism	1	Personality	1.51	.63	.24*	.05*	
	2	Acceptability	.52	.97	.10	.06	**.01**
		Situational Judgment	**.65**	**.76**	**.14**		

Note. *$p < .05$. DV = dependent variable.

References

Becker, T.E. (2005). Development and Validation of a Situational Judgment Test of Employee Integrity. International Journal of Selection and Assessment, 13(3), 225-232.

Berry, C. M., Sackett, P. R. & Wiemann, S. (2007). A review of recent developments in integrity test research. Personnel Psychology. 60. 271-301.

Connelly, B. S., Lilienfeld, S. O., & Schmeelk, K. (2006). Integrity tests and morality: Association with ego development, moral reasoning, and psychopathic personality. International Journal of Selection and Assessment, 14, 82 -86.

De Meijer, L.A.L., Born, M.P., van Zielst,J. & van der Molen, H.T. (2010). Construct-Driven Development of a Video-Based Situational Judgment Test for Integrity: A Study in a Multi -Ethnic Police Setting. *European Psychologist, 15*(3):229–236

Employee Theft: Legal Aspects – Estimates of Cost. (n.d.). Retrieved from http://law.jrank.org/pages/1084/ Employee-Theft-Legal-Aspects-Estimates-cost.html.

Goldberg, L. R., Grenier, J. R., Guion, R. M., Secherest, L. B., & Wing, H. (1991). Questionnaires used in the prediction of trustworthiness in preemployment selection decisions. Washington, DC: American Psychological Association.

Henle, C.A. (2005). Predicting workplace deviance from the interaction between organizational justice and personality. *Journal of Managerial Issues, 17,* 247-263

Harpur, T.J., Hakstian, A.R., & Hare, R.D. (1988). Two-factor conceptualization of psychopathy: construct validity and assessment implications. Psychological Assessment: a Journal of Consulting and Clinical Psychology, 1: 6-17.

Jones,J.W., Ash, P., Soto, C & Terris, W. (1991). Protecting job applicants' privacy rights when using preemployment honesty tests. In Jones, J.W.(Ed.), Preemployment honesty testing: Current research and future directions. p. 229-238. New York, NY: Quorum Books.

Lee K, Ashton MC, de Vries RE. (2005). Predicting workplace delinquency and integrity with the HEXACO and five-factor models of personality structure. Human Performance, 18, 179–197.

Lee, K., Ashton, M. C., Morrison, D. L., Cordery, J., & Dunlop, P. D. (2008). Predicting integrity with the HEXACO personality model: Use of self- and observer reports. Journal of Occupational and Organizational Psychology, 81, 147-167.

Marcus, B., Hoft, S., & Riediger, M. (2006). Integrity tests and the five factor model of personality: A review and empirical test of two alternative positions. International Journal of Selection and Assessment, 14, 113-130.

Murphy, K. R., & Lee, S. L. (1994). Does conscientiousness explain the relationship between integrity and job performance? International Journal of Selection and Assessment, 2, 226 -233.

Neuman, G. A., & Baydoun, R. (1998) An empirical examination of overt and covert integrity tests. Journal of Business and Psychology. 13. 65-79.

Ones, D.S. (1993). The construct validity of integrity testing. Unpublished Ph.D. Dissertation, University of Iowa.

Ones, D.S., Viswesvaran,C., and 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, 78, 679-703.

Ones D. S., Viswesvaran, C, & Schmidt F. L. (2003). Personality and absenteeism: A meta-analysis of integrity tests. European Journal of Personality, 17, 19–38.

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.

U.S. Congressional Office of Technology Assessment. (1990). The use of integrity tests for pre-employment screening (Report No. OTA-SET-442). Washington, DC: U.S. Government Printing Office.

Van Iddekinge C. H, Taylor M. A, Eidson CE. (2005). Broad versus narrow facets of integrity: Predictive validity and subgroup differences. Human Performance, 18, 151–177.

Wanek J. E, Sackett P. R, Ones D. S. (2003). Towards an understanding of integrity test similarities and differences: An item-level analysis of seven tests. Personnel Psychology. 56, 873–894.

Appendix A: Pre-qualification Questionnaire

Please indicate whether or not you have done the following behaviors:
• taken work supplies home for personal use without permission.
• quit a job without giving a two-week notice.
• cheated on a test or exam in school.
• left work early without permission.
• spread rumors about co-workers.
• intentionally made things more difficult for co-workers I did not like.
• taken longer breaks at work than I was supposed to.
• browsed the internet for personal use while at work.
• played computer/phone games while at work.
• blamed a co-worker for mistakes.
• misused a discount privilege.
• gossiped about employees.
• called in sick when I was not.
• stolen company equipment or merchandise.
• hidden when I didn't feel like working.
• intentionally worked slowly, in order to get overtime.
• lied to cover up for mistakes.
• overcharged on services to profit myself.
• lied about hours worked.
• made discriminatory and/or harassing comments to my co-workers.
• cursed at someone at work.
• taken property from work without permission.
• spent time daydreaming or fantasizing instead of working.
• helped a co-worker cover up an error.
• stolen items from work that would not be missed.
• taken short naps while on the clock.
• intentionally done something to get back at a rude customer.
• used an illegal drug or consumed alcohol on the job.
• helped another person or advised them how to take company property or merchandise.
• defaced, damaged or destroyed property belonging to a co-worker.
• defaced, damaged or destroyed property belonging to the company.
• intentionally withheld information from a supervisor or co-worker.
• conducted personal business during work time.
• used company resources you weren't authorized to use.
• missed work without calling in.
• returned an item that had been used.
• exchanged an item that had been damaged.
• damaged someone else's car in the parking lot without leaving a note.
• been terminated from a job.
• been arrested and charged with a crime.
• been convicted of a crime.
• checked personal e-mail while at work.
• updated social media while at work.