The Holden Applicant Reliability Measure as a Tool for Detecting Deception Angela Book ^{a*}, Ronald R. Holden ^b, Ashley Hosker-Field ^a, Nathalie Gauthier ^a, & Tabitha Methot ^a

Abstract

Two studies evaluated the utility of the Holden Applicant Reliability Measure (HARM; Holden, 2000) as an integrity measure for use in personnel selection. Because there is the risk of faking with overt tests, such as the HARM, we administered the test via computer, which has been shown to reduce the tendency for socially desirable responding (Vereecken & Maes, 2006). The HARM is a 100-item instrument measuring counterproductive employee behaviors, such as alcohol and drug use, absenteeism, and unauthorized resource usage. Respondents answer true or false to each item. In Study 1, 300 undergraduate participants completed the computer-administered HARM twice; once honestly, and once attempting to appear as well-adjusted as possible without being caught faking. Participants in Study 2 were individuals employed, seeking employment, or interning at Latin American Polygraph Institute in Columbia. As hypothesized, both studies found that HARM subscale scores differed significantly between honest and faking administrations. More specifically, all counterproductive behaviors were underreported in the faking conditions. Logistic regression analysis was conducted to determine whether the combination of HARM subscales accurately predicted the instructions to fill out the measure honestly or to fake good in order to maximize chances of employment. In both samples, the predictive utility of the HARM was found to be good to excellent. The findings suggest that the HARM is a useful measure for establishing prospective employee integrity/reliability. Counterproductive behaviors were also found to be negatively correlated with personality scales of Honesty-Humility, Agreeableness, and Conscientiousness, as measured by the HEXACO-PI-R (Lee & Ashton, 2004).

Keywords: integrity testing; personnel selection; HARM; HEXACO

Counterproductive behaviors in the workplace are both difficult to detect, and extremely costly to organizations (Ryan & Sackett, 1987), with cost estimates being in the range of billions of dollars per year (Hefter, 1986). For example, substance use by workers has been shown to be related to increased absenteeism and turnover (Normand, Salyards, & Mahoney, 1990), as well as accidents, medical costs, and worker compensation claims (Lehman & Simpson, 1992). Due to the impact of counterproductive behaviors on organizations (especially large financial implications), selection tests that evaluate employee integrity and/or honesty have become an important part of personnel selection procedures. Such tests of integrity have been developed and administered in order to reduce the cost associated with counterproductive behaviors (Hogan & Hogan, 1989; Wanek, 1999). Integrity tests are often administered during pre-employment screening (Ryan & Sackett, 1987), and the use of such tests has been increasing in popularity since the 1980s (Rudner, 1992; Sackett, 1994). Rudner reports that initial results of research were promising in that screening with integrity testing does appear to improve employee productivity under certain conditions. First, the integrity test needs to be implemented properly. Further, the test must have empirically demonstrated validity and reliability. Finally, Rudner suggests that, rather than using the integrity test in isolation, such tests should only be used in conjunction with other screening procedures.

a Brock University, St. Catharines, Ontario, Canada

b Queen's University, Kingston, Ontario, Canada

^{*} Corresponding author: abook@brocku.ca. These studies were funded by Limestone Technologies, who owns and markets the HARM.

Initially, organizations tended to utilize the polygraph to establish truth/deception in job selection contexts. These endeavors, however, met with little success, and were extremely expensive to implement within the context of personnel selection. Thus, alternatives to the polygraph were sought, and amongst them was integrity testing (Rudner, 1992). Given the ease of administration and relatively low cost, integrity testing has become a popular alternative to the polygraph.

Ones and Viswesvaran (1998) note that integrity tests were developed on general populations, making them more generalizable than clinical measures. Further, most of integrity tests are paper-pencil or computer-administered self-report measures, making them relatively inexpensive and easy to administer (Ones et al). Integrity tests were also specifically developed to predict counterproductive behaviors, and are, thus, more relevant to the workplace. Rudner (1992) suggests that a test be chosen for relevant content (i.e., to the job being applied for), giving the procedure further validity. Finally, Wanek (1999) points to the importance of choosing a test that is consistent with the company image and is appropriate for the population being tested.

Originally called "honesty testing", this type of screening is now labeled "integrity testing" (Wanek, 1999), and has also been referred to as dependability, trustworthiness, conscientiousness, or reliability testing (Sackett & Wanek, 1997). Integrity tests fall primarily into two categories. "Overt" or "clear purpose" tests ask direct questions about counterproductive behaviors in the workplace, and, therefore, make it obvious to the respondent exactly what is being measured. A second set of tests are personality-oriented tests (or "veiled purpose" tests) and these are less direct in their approach and rely on various personality traits, such as conscientiousness to predict counterproductive behaviors.

For a measure to be useful, it needs to be both reliable and valid. Previous research on the reliability of integrity testing has been promising. Ones, Viswesvaran, and Schmidt (1993) conducted a meta-analysis examining the reliability and validity of such tests. Results indicated that both types of integrity tests had high internal consistency reliability (Cronbach alpha coefficients were .82 and .77 for overt and personality oriented tests, respectively). These tests were also stable over time, exhibiting high test-retest coefficients measured between 1 and nearly 2000 days apart (adjusted coefficients were .94 and .88 for overt and personality oriented tests, respectively). The authors concluded from these findings that integrity testing meets the standards for reliability.

For a measure to exhibit validity for job selection contexts, there must be a substantial correlation with job productivity, speaking to predictive validity (Brogden, 1949; Taylor & Russell, 1939). Several meta-analyses and reviews have been conducted that support the criterion-related validity of integrity testing (e.g., Hogan & Hogan, 1989; Inwald, Hurwitz, & Kaufman, 1991; McDaniel & Jones, 1988; Ones et al., 1993). In their meta-analysis (described above), Ones et al. also examined the validity of integrity testing instruments. Effect sizes were generally medium to large for the relationship between instruments and counterproductive behaviors. Importantly, validity estimates were higher for overt tests than for personality-oriented measures. Ones et al., as such, found strong support for the validity of integrity testing in selection contexts.

A more recent meta-analysis was conducted by Van Iddekinge, Roth, Raymark, and Odle-Dusseau (2012), examining 104 studies with 134 independent samples. They also found moderate relationships between integrity tests and criterion variables, and as with Ones et al. (1993), the validity coefficients were higher for overt tests than for personality-oriented measures. It should be noted that publisher involvement in research also had an effect, with publisher data showing higher validity coefficients than independent researchers (Ones et al.). That being said, integrity tests may provide the largest increment in validity over and above cognitive testing, according to Schmidt and Hunter (1998).

Although it appears that integrity testing can be both reliable and valid, many researchers are concerned about the issue of fakability. Faking does tend to distort responding on self-report personality assessments (see Holden & Book, 2012 for a review of the research), and one might expect this to be a problem for integrity tests as well given: a) the self-report nature of the instruments and; b) that items on overt tests ask applicants to report on sensitive issues such as drug use, absenteeism, and theft. Although integrity testing does appear to have validity (e.g. Hough, Eaton, Dunnette, Kamp, & McCloy, 1990), susceptibility to faking is a compelling concern that needs to be addressed (Holden, 1995). Some research suggests that these tests are not easily faked (e.g., Ash, 1974; Morey, 1981), however such studies typically utilize extreme groups designs, with one group being, for example, inmates, and the other being recruited from the general population.

One study that used comparable groups was conducted by Ryan and Sackett (1987). One-hundred and forty-eight students completed an honesty testing measure under 1 of 3 sets of instructions (respond honestly, fake good, or respond as if applying for a job). Participants in the "fake good" condition were asked to make themselves look better than they actually are, while students in the "job" condition were asked to answer as though applying for a job. In both conditions, it was expected that participants would enhance their good qualities and minimize their bad qualities. Students in the "applying for a job" condition responded to the items in a similar manner to the honest group. This suggests that faking in a job application context may be subtle (and thus, difficult to detect). When simply asked to "fake good", students' responses were significantly more positive than the other two conditions, which did not differ from one another. The fact that the "applying for a job" condition mirrored the honest condition suggests that these instruments are susceptible to faking; thus making it plausible to include it in the administration of the test.

One avenue for potentially reducing socially desirable responding (and faking) is to administer sensitive tests (i.e., integrity tests) via computer (Vereecken & Maes, 2006). Computer administration can be seen as less personal, less judgmental, and allow for greater privacy. Such a procedure may reduce the likelihood of biased responding on sensitive issues. Wright, Aquilino, and Supple (2001) found that adolescents were more likely to self-report substance use in computer administration than in the traditional paper-and-pencil administration. In another study, Vereecken and Maes (2006) compared computer administration and paper-and-pencil versions of the same test. In a sample of over 5,000 adolescents, mode of administration did not have a significant effect on responding, except for affect-related items, where adolescents were more likely to give socially desirable answers on the paper-and-pencil version than they were during computer administration. These findings suggest that computer administration may be useful in reducing the effect of faking on integrity tests.

In the present set of studies, we evaluated the Holden Applicant Reliability Measure (HARM; Holden, 2000) as an integrity test to be utilized in personnel selection. The HARM assesses eight dimensions of on-the-job employee counterproductivity, and is comprised of 100 true/false items. HARM subscales are Alcohol Use, Interpersonal Conflict, Unauthorized Absenteeism, Missing Deadlines, Drug Use, Unauthorized Resource Usage, Dishonesty, and Arrival Tardiness. Higher scores indicate increased problems (e.g., higher scores on alcohol use indicate increased alcohol usage). Internal consistencies for HARM subscale scores are acceptable to excellent, with all subscales having coefficient alpha reliabilities above .76, with the exception of Arrival Tardiness, which has been shown to have a coefficient alpha of .64 (Holden, 2000). In a recent study, Lambert, Arbuckle, and Holden (2016) found that the HARM significantly predicted whether participants had been asked to respond honesty versus faking-good (attempting to appear better than one actually is).

The HARM is an overt integrity test, and, therefore, items are quite obvious in what they are measuring, giving the measure face validity. Importantly, the HARM has also demonstrated construct validity in that it is related to various measures of antisocial behavior and traits, including primary and secondary psychopathy (Levenson Self Report Psychopathy Scale; Levenson et al., 1995) and social symptomatology (Holden, Starzyk, Edwards, Book, & Wasylkiw, 2003). Table 1 provides correlations between the HARM subscales and measures of antisociality for the HARM validation sample of 300 individuals who were actively seeking employment (Holden et al.). Most of the relationships are mod-



erate in strength. Given the fact that all of the relationships were significant and most were moderate indicates that the HARM effectively predicts other measures of antisocial behavior, supporting its construct validity.

The current set of studies examined the utility of the HARM to detect faking in a job application context. Study 1 was conducted with undergraduate student participants, while Study 2 was conducted with actual job applicants, employees, and students at the Latin American Polygraph Institute (and employees at private companies in Columbia) in an effort to evaluate the generalizability of findings from the first study. In general, we expected respondents to produce higher HARM scores (total and subscales) when responding honestly than when they were instructed to look as well-adjusted as possible without being caught at faking. To evaluate this hypothesis, we used a repeated-measures design with each participant filling out the HARM twice (once honestly, once faking) with order being counterbalanced. We also hypothesized that the HARM subscale scores could be used to correctly classify whether or not participants were being honest.

Although the criterion-related validity of integrity testing is well established, some research suggests that these relationships can be explained based on personality traits. For example, integrity tests appear to have a moderate correlation with conscientiousness (Barrick & Mount, 1991), which is clearly related to counterproductive behavior. Further, Marcus, Lee, and Ashton (2007) directly tested whether criterion-related validity was explainable by personality traits. The authors found that the validity of overt tests was explained best by Honesty-Humility (part of the HEXA-CO model of personality), while the Big Five personality traits best predicted the validity of personality-oriented measures. Because personality (particularly as measured by the HEXACO) appears to be an important aspect of construct validity in the context of integrity testing, we examined the construct validity of the HARM in terms of its correlations with basic personality traits, as measured by the HEXACO (Lee & Ashton, 2004). The HEXACO has three factors that directly map onto traits measured by the Big Five: a) Extraversion (X; tendency to be confident, sociable, and energetic vs. unsociable, lack of liveliness and positivity), b) Conscientiousness (C; tendency to be organized, disciplined, and deliberative vs. careless, impulsive, and disorganized) and; c) Openness to Experience (O; tendency to be inquisitive, imaginative, and absorbed in art and nature vs. conventional, uncurious, and disinterested in aesthetic characteristics). Two of the other HEXACO factors are similar to their Big Five counterparts; Emotionality (E; tendency to be fearful, anxious, empathetic and sentimental vs. unworried, emotionally detached from others, and fearless) and Agreeableness (A; tendency to be forgiving, cooperative, and even-tempered vs. angry, unforgiving, and critical of others). These factors are rotated versions of Big Five Neuroticism and Agreeableness, respectively. The final HEX-ACO factor (Honesty-Humility (H)), however, measures fairness, sincerity, greed avoidance, and modesty, which is not captured by the Big Five factors. It is unsurprising then, that HEXACO personality, assessed with the HEX-ACO Personality Inventory (HEXACO-PI-R; Lee & Ashton, 2004) has been shown to outperform the Big Five in accounting for behaviors related to dishonest and manipulative (and, conversely, honest and cooperative) behaviors (e.g., Lee, Ashton, Morrison, Cordery, & Dunlop, 2008; Lee et al., 2013; Lee, Gizzarone, & Ashton, 2003). Given these findings, we hypothesized that Honesty-Humility would be related to HARM subscale scores, in that people higher on H are less likely to engage in negative behaviors in the workplace. As well, Agreeableness and Conscientiousness have obvious theoretical links to workplace behavior. People high on Agreeableness are less likely to have interpersonal conflicts, for example, and people high on Conscientiousness should be more concerned with behaving properly in the workplace, thus engaging in fewer negative workplace behaviors. We had no expectations for Emotionality, Neuroticism, or Openness.

Method

Participants

Sample 1. Three hundred undergraduate students were recruited through a psychology research participant pool at a Canadian university (mean age = 21.93 years, *SD* = 1.34). Sample 2. Participants for this sample were 156 students, employees, and referrals of the Latin American Polygraph Institute, as well as employees of private companies in Columbia. The sample consisted of 86 men and 70 women, aged 17 to 59 years (M = 32.36, SD = 10.11).

Materials. The Holden Applicant Reliability Measure (HARM; Holden, 2000), described in detail above, was administered to participants in the present study. For Study 2, where participants' first language was Spanish, the HARM was translated into Spanish by a native Spanish speaker, and was tested out on a pilot sample to ensure that the wording was appropriate. For both studies, HARM scores were transformed into T-scores.

HEXACO. The 100-item version of the HEXACO-PI-R (Spanish translation; Ashton & Lee, 2004) was used to assess six personality factors: Honesty-Humility, Emotionality, Extraversion, Agreeableness, Conscientiousness, and Openness to Experience. Participants responded to items on a five-point scale from 1 (*strongly disagree*) to 5 (*strongly agree*).

Procedure

Study 1

Participants were instructed to complete the HARM twice, once under instructions to answer honestly, and once under instructions to maximize the possibility of being selected as part of a personnel selection process. Order of administration was counterbalanced to control for order effects. Half of the participants filled out the questionnaire honestly first, while the other half completed the HARM under deceptive conditions first. The HARM was computer-administered given the obvious and sensitive nature of the questions, in an attempt to reduce the susceptibility to faking (Vereecken & Maes, 2006). For the condition maximizing selection, instructions were as follows:

NOW, PLEASE READ THE FOLLOW-ING INSTRUCTIONS TWICE

For the next questionnaire, assume

that you are in a situation where it would benefit you greatly to appear very well adjusted. Therefore, please respond so that you present yourself as someone without psychological problems or personality faults. In other words, try to fake the questionnaire so that the results will show that you are better than you really are. Although you may feel that you would never represent yourself dishonestly, please try to do so for this research study. However, beware that the questionnaire has certain features (which you want to avoid) designed to detect "faking". Do your best to fake out the questionnaire.

Study 2

In Study 2, the participants underwent the Study 1 procedure (using a Spanish translation of the HARM), and also completed the HEXACO-PI-R (Spanish Form; Lee & Ashton, 2004).

Results

Study 1

Because the current study utilized a repeated-measures design (participants filled out the HARM under two different sets of instructions), we first wanted to determine whether participant responses differed between the two administrations. All paired samples t-tests were significant (see Table 2). Thus, participants in the truthful conditions responded differently than those in the deceptive conditions. More specifically, all subscale scores were higher in the honest administration.

To determine whether HARM scores could correctly classify honest versus faking instructions, we conducted two logistic regression analyses. In the first analysis, , we used the honest responses from half of the sample (n = 150) and the deceptive responses from the other, distinct half (n = 150) of the total sample. Complementarily, the second logistic regression used the deceptive answers from the former half (n = 150) of the sample, and the honest responses from the latter half (n = 150) of the sample. In these analyses, all HARM



subscales scores were used to classify whether a respondent had been answering honestly or faking.

Both logistic regressions were statistically significant, $_{\chi}2$ (7) = 177.29, p < .001, and x^{2} (7) = 270.72, p < .001, indicating that the HARM subscales together predicted whether participants were in the honest condition or had been asked to maximize chances of employment. Sensitivities were excellent (93% and 95%) for both analyses, and specificities were very good to excellent (80% and 85%, respectively). Overall accuracies were also excellent (87% and 89%, respectively). Absenteeism, Unauthorized Resource Usage, Dishonesty, Arrival Tardiness, and Interpersonal Conflict subscales all significantly contributed uniquely to predicting instruction condition (honest vs. faking; p < .05).

The above analyses indicate that the HARM can be a valuable tool in personnel selection contexts, but a cutoff T-score is necessary in order to apply it to real situations and individuals. Based on the above analyses, we recommend using an average T-score of 42 (and lower) as an indicator of faking. Using an average T-score of 42 resulted in a sensitivity of .97 and a specificity of .99.

Study 2

Because Study 1 was conducted using an undergraduate sample, Study 2 used a broader sample of employees, potential employees, and student interns at the Latin American Polygraph Institute in Columbia. As in Study 1, we used a repeated-measures design with participants completing the HARM under two sets of instructions. HARM subscale scores were again significantly higher in the honest condition, as can be seen in Table 3.

Our second objective was to determine whether we could classify respondents as responding honestly versus faking good using HARM scores. To examine this logistic regression analyses were again applied. In the first logistic regression, we used the honest responses from one half of the sample and the deceptive responses from the other, distinct half of the total sample. In the second, we used the deceptive answers from the first half of the sample, and the honest responses from the second half. HARM subscale scores were entered as predictors, with instructional condition being the dependent variable.

As in Study 1, both logistic regressions were statistically significant, $_{\chi}2$ (7) = 140.34, p< .001, and $_{\chi}2$ (7) = 125.72, p < .001, indicating that the HARM subscales together predicted whether participants were in the honest condition or had been asked to maximize chances of employment. Sensitivities (83% and 79%), specificities, (72% and 75%), and overall accuracies (78% and 77%) were relatively high, mirroring the results from Study 1. Unauthorized Resource Usage, Dishonesty, and Arrival Tardiness subscales were all significant contributors to predicting of instructional condition (honest versus faking; p < .001).

As in Study 1, we determined that an average T-score of 42 should be used as a cutoff for faking on the HARM. Average T-scores of 42 or below produced a sensitivity of 1.00 and a specificity of 1.00.

Construct Validity of the HARM

Our final hypothesis was that HARM subscales would be negatively related to scales of Honesty-Humility, Agreeableness, and Conscientiousness (as measured by the HEXACO; Lee & Ashton, 2004). As can be seen in Table 4, relationships between the HARM and HEXACO subscales were relatively consistent, whether looking at the truthful scores or the deceptive scores. Under Honest instructions, Honesty-Humility was negatively correlated with Unauthorized Resource Usage and Dishonesty, while in the Deceptive condition, it was related negatively to Interpersonal Conflict and Dishonesty. Under Honest instructions, Agreeableness was negatively related to Interpersonal Conflict, Unauthorized Absenteeism, Unauthorized Resource Usage, Dishonesty, and Arrival Tardiness. Similarly, under Faking instructions, Agreeableness was negatively associated with both Interpersonal Conflict and Dishonesty. Conscientiousness was associated with Interpersonal Conflict, Dishonesty, and Arrival Tardiness, but only under Faking instructions. Interestingly, Arrival Tardiness scores were also negatively related to Extraversion and Openness under Faking instructions.

Discussion

The purpose of the present study was to evaluate the HARM (Holden, 2000) as a tool for assessing the integrity of prospective employees during screening procedures. As hypothesized, participants in both samples had significantly higher scores on all subscales when answering the instrument honestly than when attempting to appear well-adjusted. That is, participants tended to underreport the extent to which they engaged in various counterproductive workplace behaviours when instructed to respond in a way that would make them appear better than they really are. The clear distinction between individuals' scores in the current research implies that subtle faking was not an issue, in spite of instructions that implied the ability to detect faking. Previous research has, in fact, found that subtle faking does tend to be a problem in job application scenarios (Ryan & Sackett, 1987), but biased responding is reduced when questionnaires are computer administered (Vereecken & Maes, 2006; Wright, Aquilino, & Supple, 2001). As such, our decision to use the computer administered HARM appears to have lessened the problem of "subtle" faking, showing large differences in scores between honest and faking administrations. That being said, we did not directly compare computer administration to traditional paper/pencil administration.

The subscales of the HARM predicted whether an individual was answering honestly or faking good in both student and field samples. That is, the HARM subscale scores combined to predict whether or not the participant was responding to maximize their suitability as an applicant (specifically, attempting to appear well-adjusted). Using a cutoff T-value of 42, we were able to correctly classify 97 to 100% of participants. This finding aligns with previous research on integrity testing. A number of reviews and meta-analyses support the reliability (e.g. Ones et al., 1993) and validity of integrity testing, especially for overt integrity tests (Hogan & Hogan, 1989; Inwald, Hurwitz, & Kaufman, 1991; McDaniel & Jones, 1988; Ones et al., 1993; Van Iddekinge, Roth, Raymark, & Odle-Dusseau, 2012), with validity coefficients tending to be medium to large in size.

Further, in the field study, we found that HARM subscale scores were correlated with expected personality traits, as measured by the HEXACO. Under Honest instructions, Honesty-Humility was negatively correlated with Unauthorized Resource Usage and Dishonesty, while in the Deceptive condition, it was related negatively to Interpersonal Conflict and Dishonesty. Under Honest instructions, Agreeableness was negatively related to Interpersonal Conflict, Unauthorized Absenteeism, Unauthorized Resource Usage, Dishonesty, and Arrival Tardiness. Similarly, under Faking instructions, it was negatively associated with both Interpersonal Conflict and Dishonesty. Conscientiousness was associated with Interpersonal Conflict, Dishonesty, and Arrival Tardiness, but only under Faking instructions. Interestingly, Arrival Tardiness scores were also negatively related to Extraversion and Openness under Faking instructions. The findings for Honesty-Humility, Agreeableness, and Conscientiousness are in agreement with previous research findings that indicate these traits are related to antisocial behaviors (Lee, Ashton, Morrison, Cordery, & Dunlop, 2008; Lee et al., 2013; Lee, Gizzarone, & Ashton, 2003).

Results indicate that the HARM would be a valuable tool in personnel selection contexts. There is a clear difference in HARM scores between honest and faking administration, and the HARM scales accurately distinguish between honest and deceptive conditions. Because counterproductivity has a large impact on both organizations and society (Lehman & Simpson, 1992; Normand, Salyards, & Mahoney, 1990; Hefter, 1986), it is important to have such tools available. Additionally, integrity testing seems to have a positive impact on employee productivity under certain conditions, including using a measure that is both reliable and valid (Rudner, 1992).

Limitations of the Study

The present set of studies may have some potential limitations. First, because participants in Study 1 were undergraduate students, the results of that study may not generalize to samples of individuals who are actually seeking employment. That being said, the original validation study (Holden, 2000) was conducted with individuals who were



seeking employment, bolstering the reliability of the findings. Study 2 was conducted using a sample of potential and actual employees/interns, where the findings from the first study were replicated, confirming the utility of the HARM in real-world applications.

Although the results of the computer-administered HARM are compelling, we did not directly examine the difference between computer and traditional administration. Future research should be designed to directly compare the two administration modes.

Future Research

There are a number of research questions that will be important to investigate in future research. For example, while we did examine the construct validity of the HARM using HEXACO personality variables, researchers should examine how HARM scores relate to workplace behaviors in an employment context. Related to this, there should be attempts to examine the use of the HARM to determine integrity in various employment contexts, including law enforcement and private sector companies.

As well, Rudner (1992) states that for integrity testing to be useful, it should not conducted in isolation. Therefore, future studies should examine the utility of the HARM when it is administered as part of a larger selection package, including personality and cognitive abilities measures.



References

- Ash, P. (1974). Convicted felons' attitudes toward theft. Criminal Justice and Behavior, 1, 21-27. doi: 10.1177/009385487400100104
- Barrick, M., & Mount, M. (1991). The Big Five personality dimensions: A meta-analysis. *Personnel Psychology*, 44, 1-26. doi: 10.1111/j.1744-6570.1991.tb00688.x
- Brogden, H. (1949). When testing pays off. *Personnel Psychology*, 2, 171-183. doi: 10.1111/j.1744-6570.1949.tb01397.x
- Hefter, R. (1986). The crippling crime. Security World, 23, 36-38.
- Hogan, J., & Hogan, R. (1989). How to measure employee reliability. *Journal of Applied Psychology*, 74, 273-279. doi: 10.1037/0021-9010.74.2.273
- Holden, R. R. (1995). Response latency detection of fakers on personality tests. Canadian Journal of Behavioural Science, 27, 343-355. doi: 10.1037/0008-400X.27.3.343
- Holden, R. R. (2000, June). Development and preliminary data for the Holden Applicant Reliability Measure (HARM). Presented at the Canadian Psychological Association Annual Convention, Ottawa, Canada.
- Holden, R. R., & Book, A. (2012). Faking does distort self-report personality assessment. In M. Ziegler, C. MacCann, & R. Roberts (Eds.), *New perspectives on faking in personality* assessment (pp. 71-84). New York: Oxford University Press.
- Holden, R. R., Starzyk, K. B., Edwards, M. J., Book, A. S., & Wasylkiw, L. (2003, June). The Holden Applicant Reliability Measure (HARM). Construct validity in a university sample. Presented at the Canadian Psychological Association Annual Convention, Hamilton, Canada.
- Hough, L. M., Eaton, N. K., Dunnette, M. D., Kamp, J. D., & McCloy, R. A. (1990). Criterionrelated validities of personality constructs and the effect of response distortion on those validities. *Journal of Applied Psychology*, 75, 581-595. doi: 10.1037/0021-010.75.5.581
- Inwald, R. E., Hurwitz, H., Jr., & Kaufman, J. C. (1991). Uncertainty reduction in retail and public safety-private security screening. *Forensic Reports*, 4, 171–212.
- Lambert, C. E., Arbuckle, S. A., & Holden, R. R. (2016). The Marlowe-Crowne Social Desirability Scale outperforms the BIDR Impression Management Scale for identifying fakers. *Journal of Research in Personality*, 61, 80-86. doi: 10.1016/j.jrp.2016.02.004
- Lee, K. & Ashton, M. C. (2004). Psychometric properties of the HEXACO personality inventory. *Multivariate Behavioral Research*, 39, 329-358.
- 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.
- Lee, K., Ashton, M. C., Wiltshire, J., Bourdage, J. S., Visser, B. A., & Gallucci, A. (2013). Sex, power, and money: Prediction from the Dark Triad and Honesty-Humility. *European Journal of Personality*, 27, 145-154. doi:10.1002/per.1860
- Lee, K., Gizzarone, M., & Ashton, M. C. (2003). Personality and the likelihood to sexually harass. *Sex Roles*, 49, 59-69.



- Lehman, W. E. K., & Simpson, D. D. (1992). Employee substance use and on-the-job behaviors. *Journal of Applied Psychology*, 77, 309-321. doi: 10.1037/0021-9010.77.3.309
- Levenson, M. R., Kiehl, K. A., & Fitzpatrick, C. M. (1995). Assessing psychopathic attributes in a noninstitutionalized population. *Journal of Personality and Social Psychology*, 68, 151-158.
- Marcus, B., Lee, K., & Ashton, M. (2007). Personality dimensions explaining relationships between integrity tests and counterproductive behavior: Big Five or one in addition? *Personnel Psychology*, 60, 1-34. doi: 10.1111/j.1744-6570.2007.00063.x
- McDaniel, M. A., & Jones, J. W. (1988). Predicting employee theft: A quantitative review of the validity of a standardized measure of dishonesty. *Journal of Business and Psychology*, 2, 327–343. doi: 10.1007/BF01013765
- Morey, L. (1981). *Statistical properties of the Wilkerson Pre-employment Audit.* Wilkerson & Associates, Tulsa, OK.
- Normand, J., Salyards, S. D., & Mahoney, J. J. (1990). An evaluation of preemployment drug testing. *Journal of Applied Psychology*, 75, 629-639. doi: 10.1037/0021-9010.75.6.629
- Ones, D. S., & Viswesvaran, C. (1998). Integrity testing in organizations. In R. W. Griffin, A. O'Leary, & J. Collins (Eds.), *Dysfunctional behavior in organizations:* Vol. 2, Nonviolent behaviors in organizations. Greenwich, CT, JAI Press.
- 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. doi: 10.1037/0021-9010.78.4.679
- Rudner, L. (1992). Pre-employment testing and employee productivity. *Public Personnel Management*, 21, 133-150.
- Ryan, A., & Sackett, P. (1987). Preemployment honesty testing: Fakability, reactions of test takers, and company image. *Journal of Business Psychology*, 1, 248-256.
- Sackett, P. (1994). Integrity testing for personnel selection. Current Directions in Psychological Science, 3, 73-76. doi: 10.1111/1467-8721.ep10770422
- Sackett, P., & Wanek, J. (1997). Integrity testing: An overview. Security Journal, 8, 11-18.
- Schmidt, F. L., & Hunter, J. E. (1998). The validity and utility of selection methods in personnel selection: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin*, 124, 262–274. doi: 10.1037/0033-2909.124.2.262
- Taylor, H., & Russell, J. (1939). The relationship of validity coefficients to the practical effectiveness of tests in selection. *Journal of Applied Psychology*, 23, 656-678. doi: 10.1037/h0057079
- Van Iddekinge, C., Roth, P., Raymark, P., & Odle-Dusseau, H. (2012). The criterion-related validity of integrity tests: An updated meta-analysis. *Journal of Applied Psychology*, 97, 499-530. doi: 10.1037/a0021196
- Vereecken, C., & Maes, L. (2006). Comparison of a computer-administered and paper and pencil administered questionnaire on health and lifestyle behaviors. *Journal of Adolescent*

Health, 38, 426-432. doi: 10.1016/j.jadohealth.2004.10.010

- Wanek, J. (1999). Integrity and honesty testing: What do we know? How do we use it? International Journal of Selection and Assessment, 7, 183-195. doi: 10.1111/1468-2389.00118
- Wright, D., Aquilino, W., & Supple, A. (2001). A comparison of computer-assisted and paperand-pencil self-administered questionnaires in a survey on smoking, alcohol, and drug use. *Public Opinion Quarterly*, 62, 331-353. doi: 10.1086/297849



	Primary Psychopathy	Secondary Psychopathy	Total Psychopathy	Social Symptomatology	
HARM Scale				<u> </u>	
Alcohol Use	.18	.18	.21	.26	
Interpersonal Conflict	.34	.26	.36	.42	
Unauthorized	16	21	21	28	
Absenteeism	.10	.21	.21	.20	
Missing Deadlines	.22	.26	.27	.19	
Drug Use	.22	.19	.24	.33	
Unauthorized Resource	10	20	22	22	
Usage	.10	.20	.22	.55	
Dishonesty	.33	.34	.39	.38	
Arrival Tardiness	.17	.21	.21	.24	
HARM Total Score	.36	.38	.42	.49	
Employee	20	24	26	41	
Misdemeanors	.29	.34	.50	.41	
Workplace	24	20	27	.44	
Antisociality	.94	.27			
Average Correlation	.25	.26	.30	.34	

Table 1. Criterion-related validity coefficients for the HARM (from Holden et al., 2003)

Note. Values are correlations and all are significant at the .05 level.

Variable	Condition	T-score	SD	t	р
Alcohol Use	Honest	49.86	9.90	6.55	< .001
	Deceptive	45.87	3.95		
Unauthorized					
Absenteeism	Honest	49.97	10.00	18.56	<.001
	Deceptive	39.18	2.45		
Missing Deadlines	Honest	50.00	9.99	10.16	< .001
	Deceptive	43.58	5.56		
Drug Use	Honest	50.01	10.07	3.88	< .001
	Deceptive	47.75	0.00		
Unauthorized					
Resource Usage	Honest	50.03	10.01	13.96	<.001
	Deceptive	41.67	4.81		
Dishonesty	Honest	50.00	9.99	17.19	<.001
	Deceptive	39.90	3.30		
Arrival Tardiness	Honest	51.99	9.99	3.32	.001
	Deceptive	49.95	4.31		
Interpersonal					
Conflict	Honest	50.02	10.02	9.98	<.001
	Deceptive	44.23	3.56		

Table 2. Study 1: Descriptive statistics and paired samples t-tests for HARM subscales.

Note. *N* = 300.

Variable	Condition	T-score	SD	t	р
Alcohol Use	Honest	46.59	4.69	3.83	< .001
	Deceptive	45.19	1.47		
Unauthorized					
Absenteeism	Honest	43.23	5.00	5.52	< .001
	Deceptive	41.02	1.05		
Missing Deadlines	Honest	47.05	3.39	3.23	.002
	Deceptive	46.15	1.58		
Drug Use	Honest	46.05	3.32	1.80	.07
	Deceptive	45.57	0.00		
Unauthorized					
Resource Usage	Honest	43.09	5.65	10.28	< .001
	Deceptive	37.88	3.13		
Dishonesty	Honest	48.06	7.25	9.16	< .001
	Deceptive	42.72	2.96		
Arrival Tardiness	Honest	47.10	6.26	8.53	< .001
	Deceptive	42.30	3.56		
Interpersonal					
Conflict	Honest	46.20	4.70	3.95	< .001
	Deceptive	44.61	2.36		

Table 3. Study 2: Descriptive statistics and paired samples t-tests for HARM subscales.

Note. N = 160.

		Н	Е	Х	А	С	0
Truthful	Alcohol Use	03	.06	05	10	11	.02
	Interpersonal						
	Conflict	05	.10	07	34**	07	.01
	Unauthorized						
	Absenteeism	- .11	.05	.01	21**	06	.12
	Missing Deadlines	01	.09	10	07	01	.03
	Drug Use	.09	.06	07	.04	03	.09
	Unauthorized						
	Resource Usage	16*	.11	09	18*	09	.03
	Dishonesty	22**	.02	05	19*	08	.06
	Arrival Tardiness	07	.05	11	17*	08	.10
Deceptive	Alcohol Use	08	.03	12	.06	13	09
-	Interpersonal						
	Conflict	16*	06	.003	16*	17*	11
	Unauthorized						
	Absenteeism	08	.08	04	08	09	07
	Missing Deadlines	03	.03	.01	04	.06	.02
	Drugs	а	а	а	а	а	а
	Unauthorized						
	Resource Usage	004	.05	04	02	07	06
	Dishonesty	16*	.04	02	21**	16*	.01
	Arrival Tardiness	.03	.06	18*	.01	22**	23**

Table 4. Correlations between HARM subscales and HEXACO personality traits.

Note. N = 160.

^a No correlations were calculated for Drug Use scores in the Deceptive condition because there was no variability.