

Initial Reliability and Validity for the Critical Hire® - Screen

Anthony W. Tatman

Des Moines, Iowa

&

Matthew T. Huss

Creighton University, Omaha, Nebraska

Author Note

Anthony W. Tatman, Critical Hire®, Des Moines, Iowa; Matthew T. Huss, Department of Psychology, Creighton University.

Correspondence concerning this article should be addressed to Anthony “Tony” Tatman, 500 Hickman Road, #813, Waukee, IA 50263-7713. E-mail: [tonytatman@criticalhire.com](mailto:tonytatman@criticalhire.com).

ORCID #: 0000-0002-0195-0193

Manuscript accepted for publication in the journal *Personnel Assessment and Decisions*.

### Abstract

Although the use of integrity testing during the application process has become a frequent practice in general business settings, their use has been rather non-existent in the field of corrections. This limited use may stem from a lack of awareness about integrity tests in corrections, as well as a lack of integrity measures that have been normed and validated for use with correctional applicants. This study outlines the development, reliability and validity for the Critical Hire®-Screen (CH-S), an overt integrity assessment measure developed for probation, parole, and other correctional officer job applicants. Four separate studies were conducted and provide evidence for the internal consistency, test-retest reliability, and convergent validity for the CH-S. Results provide initial support for its use as a tool correctional agencies could use to measure various traits of integrity at a pre-offer phase in the hiring process.

### Initial Reliability and Validity for the Critical Hire® - Screen

Integrity is a critical characteristic for any employee. As Warren Buffet stated, “When looking for people to hire, you look for three qualities: integrity, intelligence, and energy. And if you don’t have the first, the other two will kill you” (Schwantes, 2018). The use of assessments that measure job applicant integrity has grown exponentially since the Employee Polygraph Protection Act of 1988, prohibiting the use of polygraphs for all but a select few employment settings. Initially developed as a surrogate for the polygraph, and designed to measure dishonesty, integrity tests have expanded in type (e.g., overt vs personality-based tests; Sackett, Burris, & Callahan, 1989), and scope, evaluating the risks for poor job performance and a wide variety of counterproductive work behaviors (CWBs) such as theft, drug use, disruptive behaviors, and violence in the workplace (Berry, Sackett & Wiemann, 2007). Integrity tests have since been identified as the most widely used type of assessment tool for predicting CWBs among job applicants and employees (Fine, Horowitz, Weigler, & Basis, 2010). This popularity and widespread use stems largely from the extensive empirical evidence supporting integrity test’s reliability and validity in predicting job performance and CWBs (Berry, et. al., 2007; Fine, 2013; Fine et al., 2010; Jones, Cunningham, & Dages, 2010; Marcus, Ashton, & Lee, 2013; Ones, Viswesvaran, & Schmidt, 1993; Schmidt, Oh, Shaffer, 2016; Schmidt & Hunter, 1998; Wanek, 1999). For example, Ones et al.’s (1993) comprehensive meta-analysis explored the predictive validity of both overt and personality-based integrity measures, as well as possible moderators impacting validity estimates. Ones et al. revealed that both overt and personality-based integrity measures significantly predicted overall job performance and CWB for low, medium and high-complexity jobs, as defined by Hunter, Schmidt, and Judiesch (1990). Integrity tests have also been found to be valid predictors of a variety of specific CWBs, such as theft,

tardiness, property damage, rule-breaking, violence, and absenteeism (Nicol & Paunonen, 2002; Ones et al., 1993; Ones, Viswesvaran, & Schmidt, 2003). Research has further revealed that integrity tests add significant incremental validity to measures of general mental ability in predicting job performance (Schmidt & Hunter, 1998; Schmidt et. al., 2016), which are well known in the field for having high levels of validity for predicting job performance when compared to other personnel measures (Hunter, 1986; Hunter & Hunter, 1984, Hunter & Schmidt, 1996; Ree & Earles, 1992; Schmidt, 2002; Schmidt, Schaffer, & Oh, 2008).

Counterproductive work behaviors, poor job performance and deficiencies in integrity are problematic for any organization, but can be exceptionally problematic for correctional agencies where its employees are expected to be models of character and integrity, and entrusted to uphold the law, protect the public, and serve as agents of change for their clients. With this expectation, correctional officers have been given considerable power, authority, autonomy, and public trust, making an officer's integrity a vital and essential characteristic to the health of their agency, the community, and the law enforcement profession as a whole. However, despite the essential function integrity plays in correctional officers, the empirical support for the use of integrity testing in the field of law enforcement (Jones et al., 2010), and the widespread use of integrity testing in other industries, it has been our experience that pre-offer integrity testing remains a less known and utilized process for many correctional agencies. This limited knowledge and use may be due to a level of comfortability and reliance on traditional hiring methods, such as criminal background and reference checks, and/or limited availability of validated integrity tests for correctional applicants.

Most selection processes for correctional officers follows a compensatory model and involves an initial screening to ensure the applicant meets minimum qualifications (e.g., age and

education level). Screening of minimum qualifications are then typically followed by interviews, as well as criminal and employment background checks. However valuable parts of any hiring process, criminal and employment background checks have their limitations. Both checks monitor a candidate's history, which may or may not raise red flags to concerns regarding the applicant's integrity and the degree to which they may engage in future CWBs. For example, most employment background checks only give an individual's start and/or end date, or only provide positive feedback, despite substandard or worse job performance, out of fear of defamation and/or lawsuits. Similarly, criminal background checks have limitations of being difficult to compile and track if the person has lived in multiple states or out of the country. Criminal background checks are also limited by only capturing charges at certain levels of severity, frequently excluding misdemeanors and "lesser" crimes. Many CWBs also occur without being caught, or even reported to authorities if caught, further limiting the utility of criminal background checks. For example, a study from the University of Cincinnati found that, while 64% of small businesses they studied experienced employee theft, only 16% of those businesses reported the incident to police despite the amount stolen averaging \$20,000.00 (Brooks, 2014).

In some correctional agencies interviews and background checks are followed by post-conditional offer, preemployment psychological evaluations (PPEs). Similar to the screening processes listed above, although there are numerous benefits and value in PPEs (e.g., identifying antisocial personality characteristics or significant mental health symptoms), there are also notable limitations. PPEs are expensive and require highly trained and professionally licensed personnel to administer. Additionally, Courts have ruled (*Griffin v Steeltek, Inc.*, 1997; *Karraker v Rent-A-Center*, 2003) that many of the assessment measures Psychologists use in PPEs (e.g.,

Minnesota Multiphasic Personality Inventory-2) are medical in nature, and would therefore violate the Americans with Disabilities Act (ADA) and Equal Employment Opportunity Commission (EEOC) guidelines if administered before a conditional offer for employment was given. Therefore, PPEs make up the final step in the hiring process. Identifying an applicant's propensity for antisocial thinking, attitudes supportive of theft or bribery, or a tendency to violate rules or policies, for example, would naturally be valuable information to obtain through a PPE, but financially costly to uncover at this final stage in the hiring process. Integrity testing, on the other hand, can be administered either at the pre- or post-offer phase in the hiring process, and when used as intended, do not violate the ADA or EEOC guidelines since they are not considered medical tests, and are not designed, or used, to identify mental disorders (Berry et al., 2007; Stabile, 2002). In addition to obtaining valuable information earlier in the hiring process, incorporating pre-offer integrity tests prior to the post-offer PPE has also been found to reduce costs for hiring agencies. Corey (2008) reported that the applicant disqualification rate for law enforcement officers after post-offer PPEs averaged 25% across the United States. However, when PPEs were preceded by a valid and reliable pre-offer integrity assessment the disqualification rate following a PPE fell to 5%. This 20% reduction in rejection rate has significant cost saving implications for hiring agencies by reducing costs associated with repeated and costly advertising, screening of applications, interviewing, and PPEs.

Pre-offer integrity testing may also help hiring agencies save money through potentially reducing turnover and terminations. For example, after implementing integrity tests into their application and hiring process one business saw a 50% reduction in terminations historically caused by employee misconduct such as theft, illegal drug use, and violence over a five-year period (Brown, Jones, Terris, & Steffy, 1987). This reduction in turnover through terminations

has considerable cost savings. The U.S. Department of Labor estimates that replacing a poor performer could cost 30% of that employee's potential first-year earnings (Fatemi, 2016). With the average annual salary for probation and correctional officers ranging between \$43,540 and \$56,630 (United States Department of Labor, Bureau of Labor Statistics, 2017), the cost to replace a poor performer, either through involuntary or voluntary departures, could fall between \$13,062 and \$16,989 for each officer. Unfortunately for hiring agencies, these figures are low estimates and do not include additional, direct costs such as legal fees and settlements costs, or indirect costs such as lost productivity, strained moral, and fractured public trust often associated with turnover and terminations.

#### Introducing the Critical Hire® – Screen

Seeing a need for a pre-offer, integrity assessment for correctional applicants, the first author conducted an extensive review for published integrity assessments with norms specific to correctional applicants. No such instruments were found. As a result, the Critical Hire® – Screen (CH-S; Tatman, 2018) was created. The CH-S is a pre-offer integrity test developed on, and validated with, correctional employees and applicants. Since its development, the CH-S has been used throughout various Midwest departments of correctional services as part of a comprehensive interview and hiring process. However, up to this point, reliability and validity data have not been published making this tool rather unknown in the larger field of personnel assessment and corrections. Therefore, the purpose of this paper is to introduce the CH-S and share initial reliability and validity data.

Development of the CH-S began with the first author developing test questions that elicited information on various CWBs pertinent in the field of corrections: illegal substance use, substance use in the work place, theft, disregard for laws, negative views toward the court and

police, problems with authority, attitudes supporting law violations, propensities for violating rules and policies, propensities to manipulate others for personal gain, and blaming victims for crimes committed against them. Fifty-two items were generated. These items were then administered to a convenience sample of correctional administrators, college students, and non-correctional participants (N = 346). Following a series of independent t-tests, items were eliminated that showed significant differences based on racial status, age, and gender. Specifically, items were eliminated that showed disproportionately elevated scores for minorities (vs Caucasians), individuals 40 years of age or older (as opposed to 39 or younger), and female participants. Using principal component analysis with Varimax rotation, the remaining 29 items were factor analyzed which revealed the presence of five distinct factors subsequently labeled Substances, Theft, Authority, Rules and Deception, and Responsibility. These five integrity factors make up the Personal Opinions and Beliefs section of the CH-S. The first author also developed 13 test questions that identify potential socially desirable responding. Sample questions consisted of “I always admit when I make a mistake” and “I am always nice to others, even to people who are not nice to me.” The same process mentioned above was used to eliminate items measuring social desirability that showed significant differences based on minority status, age, and gender. The resulting pool of items were further reduced by retaining only those items that were endorsed by 20% or less of the developmental sample. The resulting items made up the CH-S’s Impression Management Scale (IMS; N = 4). Lastly, with consultation with corrections and law enforcement administrators, the first author also developed test questions inquiring about past work and legal history, aptly labeled Employment and Legal History (N = 22 items). The Employment and Legal History section contains direct admission questions, inquiring into the applicant’s past work and legal experiences. Example questions



contained in the Employment and Legal History section include “Have you ever quite a job to avoid being terminated?” and “Have you ever been on probation, parole, or under some other form of Court mandated supervision?” Further descriptions and example items for the IMS and five integrity scales making up the Personal Opinions and Beliefs section of the CH-S can be found in the Appendix. The entire CH-S is set at a 5<sup>th</sup> grade reading level (Flesch-Kincaid Grade Level).

### Studies and Research Questions

This paper presents four separate studies, and asked seven different research questions, to measure the reliability and validity of the CH-S.

#### Study 1:

Research Question 1: Do the five CH-S scales and IMS have adequate internal consistency?

Research Question 2: Does the IMS have adequate convergent validity?

Research Question 3: Are there significant differences based on race, gender, and age for the five CH-S scales and IMS?

#### Study 2:

Research Question 4: Do the five CH-S scales and IMS have adequate internal consistency when administered to a sample of correctional applicants?

#### Study 3:

Research Question 5: Do the five CH-S scales and IMS have adequate internal consistency when administered to a sample of correctional applicants in Study 3?

Research Question 6: Does the CH-S have adequate convergent validity?

Study 4:

Research Question 7: Does the CH-S have adequate test-retest reliability?

### Study 1 - Method

#### Participants

Participants in study 1 consisted of a convenience sample of correctional administrators and supervisors within a large metropolitan, community-based corrections agency, and college students attending a Midwestern university. In order to help ensure valid information, participants were allowed to complete the surveys anonymously. Due to the anonymous nature of the data collected, the exact percentage of these groups of participants is unknown. Participants consisted of a convenience sample of 308 adult males ( $N=114$ ) and females ( $N = 193$ ). One participant did not identify gender. Racial breakdown consisted of 244 (79.2%) Caucasian, 32 (10.4%) Asian, 15 (4.9%) Hispanic, and 8 (2.6%) African American participants. Four (1.3%) participants identified country of origin without identifying racial status, while 5 (1.6%) did not specify racial status. The mean age was 29.45 ( $SD = 15.99$ ), with a range of 18-77.

#### Measures

In addition to the CH-S, participants completed a modified version of the Balanced Inventory of Desirable Responding-16 (BIDR-16; Hart, Ritchie, Hepper, & Gebauer, 2015). The BIDR-16 is a 16-item measure of socially desirable responding that was developed with the use of confirmatory factor analysis from the 40-item Balanced Inventory of Desirable Responding (BIDR; Paulhus, 1991, 1998). For this study the BIDR-16 was completed using a True-False response format in order to match the response format used for the IMS, as opposed to its original, Likert scale format. Although the BIDR-16 uses a Likert scale response format,

dichotomous response methods were authorized for the original BIDR (Paulhus, 1994). Given that the BIDR-16 item content originated from the BIDR, the present authors felt using a dichotomous scoring method for the BIDR-16 fell within the original test administration standards and was appropriate to use in this study. The BIDR-16 consists of 16 items generating two sub-scales: Self-Deceptive Enhancement (SDE; overly positive responding or enhancing one's strengths) and Impression Management (IM; bias toward pleasing others or minimizing one's faults). Hart et al. found that the BIDR-16 showed significant correlations with the BIDR (IM = .84,  $p < .001$ ; SDE = .87,  $p < .001$ ) and Marlow Crown Social Desirability Scale-Short (Strahan & Gerbasi, 1972; IM=.53,  $p < .001$ ; SDE = .32,  $p < .001$ ), and showed adequate 2-week test-retest reliability (IM = .74,  $p < .001$ ; SDE = .79,  $p < .001$ ).

## Procedures

Participants were invited to complete a paper-pencil survey which included the CH-S and BIDR-16. Other than the college students, who received course credit for their participation, participants were not compensated for their participation. Although it is unknown how many individuals declined to participate, 308 surveys were ultimately completed.

## Study 1 - Results

*Research Question 1 - Do the five CH-S scales and IMS have adequate internal consistency?*

Cronbach Alpha coefficients revealed that the five CH-S scales have moderate internal consistency for this sample: Substances ( $\alpha = .79$ ), Theft ( $\alpha = .76$ ), Authority ( $\alpha = .61$ ), Rules & Deception ( $\alpha = .70$ ), and Responsibility ( $\alpha = .72$ ). The IMS, however, showed relatively low internal consistency ( $\alpha = .50$ ).

*Research Question 2 - Does the IMS have adequate convergent validity?*

Results revealed that the IMS produced significant Pearson Correlation coefficients with the BIDR-16 IM (.416,  $p < .000$ ), SDE (.380,  $p < .000$ ), and full BIDR-16 (.472,  $p < .000$ ) scales. This finding suggests that, although the IMS showed low internal consistency with this sample, there is evidence of convergent validity based on these significant relationships with the BIDR-16, a similar measure of social desirability.

*Research Question 3 - Are there significant differences based on race, gender, and age for the five CH-S scales and IMS?*

A series of independent samples t-tests were conducted to investigate the degree to which the five CH-S scales and IMS showed significant differences based on race. Due to the relatively small number of minority participants in the various racial groups found in this sample, we chose to combine all minorities into a single “minority” group for this analysis. Results revealed no significant differences between racial minorities and Caucasian participants on the five CH-S scales or IMS: IMS ( $t = .99$ ;  $p = .32$ ), Substances ( $t = -1.56$ ;  $p = .12$ ), Theft ( $t = -1.60$ ;  $p = .11$ ), Authority ( $t = -1.66$ ;  $p = .10$ ), Rules & Deception ( $t = -1.63$ ;  $p = .10$ ), and Responsibility ( $t = -.64$ ;  $p = .52$ ).

A series of independent samples t-tests were also conducted to investigate the degree to which the five CH-S scales and IMS differed based on gender. Results revealed no significant differences between Males and Females on the IMS, Theft, Authority, or Rules & Deception scales (Table 1). Males, however, were found to score significantly higher than Females on the Substances and Responsibility scales. However, the effect size (Cohen’s  $d$ ) for each of these differences is relatively small, suggesting that, although there is a statistical difference, the magnitude of this difference in real world applications is rather insignificant. Common Language Effect Sizes (CLES; McGraw & Wong, 1992) were also calculated to measure the degree to

which these significant differences had real world implications. The CLES is a variation to Cohen's  $d$ , and described by McGraw and Wong as being easier to interpret and use by the general public who may not be well versed in statistics. CLES scores simply represent the probability that a score sampled at random from one distribution (i.e., Males) would be greater than a score sampled from the other distribution (i.e., Females). Table 1 shows that the respective CLES scores were just above chance, further suggesting that the effect size, or real word impact, for these differences have little significance in practical, real world applications.

A series of independent samples t-tests were also conducted to measure whether the five CH-S integrity scales and IMS scores significantly differed based on age. The Age Discrimination in Employment Act (ADEA) prohibits discrimination against people who are ages 40 or older. Therefore, to be consistent with this ADEA guideline, the following t-test comparisons were based on participants who were 40 or older ( $N = 77$ ) against participants who were 39 and younger ( $N = 222$ ). Results from these analyses revealed that the two age groups did not significantly differ on the IMS (Table 2). However, significant differences were found for all five of the CH-S scales, with individuals under 40 scoring significantly higher (i.e., representing a higher level of support for theft, substance use, and rule violations, for example) than participants 40 years of age or older. This finding is consistent with the Age-Crime Curve theorized by Hirschi and Gottfredson (1983), and supported by Rocque, Posick, and Hoyle (2016) which proposes that as a person ages their propensity for deviant attitudes and behavior diminish.

## Study 2 – Method

### Purpose, Measure, Participants, and Procedures

Study 2 was conducted to answer Research Question 4: “Do the five CH-S scales and IMS have adequate internal consistency with a sample of correctional applicants?” The sample used in this study consisted of 218 adults applying for positions as residential officers, probation and parole officers, and certified police officers within three Midwest departments of community-based corrections. Potential participants were provided digital copies of the agencies application and a link to complete the CH-S by their prospective employer with instructions that the CH-S would be used in the agency’s hiring process. Consistent with EEOC guidelines, age, gender, and racial status were not solicited during this initial application phase, and therefore not available for this study. Zero participants declined to complete the CH-S.

### Study 2 – Results

Cronbach Alpha coefficients were calculated to measure the CH-S’s internal consistency with this sample of applicants. Results indicate that the CH-S had moderate to high internal consistency with this sample: Substances ( $\alpha = .90$ ), Theft ( $\alpha = .73$ ), Authority ( $\alpha = .78$ ), Rules & Deception ( $\alpha = .76$ ), Responsibility ( $\alpha = .71$ ), and IMS ( $\alpha = .74$ ). It is noteworthy that the IMS had a moderate alpha coefficient in this sample, as compared to the low alpha found in Study 1. Therefore, these results provide added support for the internal consistency of the CH-S’s five scales, while introducing higher levels of internal consistency for the IMS compared to results obtained in Study 1.

### Study 3 - Method

#### Purpose

The purpose of Study 3 was to answer Research Question 5: “Does the CH-S have adequate convergent validity?” Convergent and discriminant validity are subtypes of construct validity. While convergent validity measures the degree to which constructs or measures that

should be alike are alike, discriminant validity measures the degree to which constructs or measures that should be different are different. The purpose of this study was limited in its scope to only measuring convergent validity, leaving subsequent research to be conducted on the CH-S's discriminant validity.

### Measures

Participants in this study completed on-line versions of the CH-S and Step One Survey II™ (SOS; Profiles International, Inc., 2004). The SOS is a proprietary, overt integrity assessment designed for general preemployment selection purposes. The SOS does not contain norms specific to correctional officers or other law enforcement personnel; however, was chosen for this analysis in order to measure the degree to which the general concepts between, and contextual framework of, the CH-S correlate with those of the SOS. The SOS is made up of 132 questions which include historical, direct admission questions, as well as questions measuring attitudes held in 4 primary domains of work conduct: Integrity (i.e., theft of money, time and property;  $\alpha = .83$ ), Substance Abuse (i.e., personal use or distribution of illegal substances;  $\alpha = .84$ ), Reliability (i.e., the degree to which they will follow rules or procedures;  $\alpha = .77$ ), and Work Ethic (i.e., attitudes toward work and supervisors;  $\alpha = .76$ ) (Profiles International, Inc.). The SOS also contains a Distortion scale designed to measure socially desirable responding. However, no psychometrics were provided on the Distortion scale in the SOS Construction and Validation Report (Profiles International, Inc.).

### Procedures

Applicants were provided digital copies of the respective agencies' employment application along with links to complete the on-line versions of the CH-S and SOS by their prospective employer. The CH-S and SOS were completed during the initial phases of the

applicant's application process. Pearson correlation coefficients were conducted to measure inter-scale correlations between the CH-S and SOS.

### Participants

Participants in study 3 consisted of 159 adults (Males =61; Females = 98) applying for positions as residential, probation and parole officers within two Midwest departments of community-based corrections. The SOS's Interview Report references the examinee's gender in the report's narrative, allowing this study's authors to know the participant's gender. Otherwise, consistent with EEOC guidelines other demographic information such as age and racial status were not solicited when these measures were completed, and therefore not obtained for this study.

### Study 3 – Results & Discussion

*Research Question 5: Do the five CH-S scales and IMS have adequate internal consistency when administered to a sample of correctional applicants in Study 3?*

Cronbach alpha coefficients revealed that the CH-S had adequate internal consistency with this sample of job applicants: Substances ( $\alpha = .93$ ), Theft ( $\alpha = .82$ ), Authority ( $\alpha = .77$ ), Rules & Deception ( $\alpha = .82$ ), and Responsibility ( $\alpha = .82$ ), and IMS ( $\alpha = .73$ ). These results provide added support for the reliability of the five CH-S scales and IMS.

*Research Question 6: Does the CH-S have adequate convergent validity?*

Pearson Correlation coefficients results revealed significant correlations between the CH-S scales and SOS factors (Table 3). These results suggest that the CH-S and SOS are measuring similar constructs of counterproductive work behavior, and provide evidence for the convergent validity for the five CH-S scales.



Consistent with Research Question 3, post-hoc t-tests were also calculated to measure the degree to which the CH-S scales differed between Male and Female participants. Results revealed no significant differences between genders in this sample: IMS ( $t = .11$ ;  $p = .92$ ), Substances ( $t = .41$ ;  $p = .68$ ), Theft ( $t = .74$ ;  $p = .46$ ), Authority ( $t = -.004$ ;  $p = .99$ ), Rules & Deception ( $t = 1.79$ ;  $p = .08$ ), and Responsibility ( $t = 1.19$ ;  $p = .23$ ).

#### Study 4 - Method

##### Purpose, Measure, Participants, and Procedures

Study 4 was conducted to answer Research Question 7 “Does the CH-S have adequate test-retest reliability?” Probation, parole, and residential officer applicants from three Midwest departments of community-based corrections completed the CH-S as part of their application process. One hundred applicants completed the CH-S on multiple occasions as they repeatedly applied for positions. In order to measure the test-retest reliability over different retest frequencies the sample was divided into Short and Long frequency groups (Table 4). Applicants in the Short frequency group ( $N = 42$  applicants) had a retest frequency between 1 and 30 days (Mean = 12 days, Median = 12 days). Applicants in the Long frequency group ( $N = 57$  applicants) had a retest frequency between 31 and 541 days (Mean = 130, Median = 79 days).

#### Study 4 – Results and Discussion

##### *Research Question 7: Does the CH-S have adequate test-retest reliability?*

Pearson Correlation coefficients between CH-S administrations are provided in Table 4. Results from Study 4 provide evidence that the CH-S has good test-retest reliability over a relatively short duration (i.e., 1 month), and that this reliability is relatively preserved for over 1 year, supporting the stability of CH-S scores over time.

#### Summary, Discussion & Limitations

The results generated from the four studies conducted in this paper provide support for the internal consistency, test-retest reliability, and convergent validity of the Critical Hire® – Screen (CH-S). Particularly noteworthy, data gathered from Studies 1 and 3 provide tentative evidence that CH-S test scores do not show group differences in terms of gender, age, and racial status that would infringe on EEOC and ADEA guidelines. CH-S scores also appear to be stable over time, with results from Study 4 suggesting that the five CH-S scales and IMS maintain their reliability for approximately 1 ½ years. Although the five CH-S integrity scales showed internal consistency throughout the various studies in this paper, mixed results were, however, on the internal consistency of the CH-S's IMS. While the initial analysis in Study 1 found the IMS to have low internal consistency, Studies 2 and 3 found adequate Cronbach alpha coefficients. This difference may be due to the different sample populations used between Study 1 and Studies 2 and 3, with Study 1 being comprised of a considerable number of college students, while Studies 2 and 3 did not. Although Studies 2 and 3 provide compelling evidence for the internal consistency of the IMS, further analysis is recommended to help determine if this mixed result is a function of the sample or scale development.

These findings have relevance for agencies hiring correctional personnel. Correctional officers are tasked with an enormous responsibility to protect the public, enforce Court ordered sentencing obligations, and reduce recidivism through rehabilitation. As a result, correctional officers have been placed into positions of considerable trust, authority, and power by their hiring agencies, the Courts, and the public. Therefore, it is critical that individuals entering these high trust, high power positions possess unparalleled character and integrity. Results shared in this paper provide supporting evidence for the reliability and validity of the CH-S as a tool hiring agencies could use to measure traits of integrity at a pre-offer phase in the hiring process. Results

from these studies further suggest that the CH-S is appropriate for use with applicants ages 18 to 77, males and females, and with applicants of African American, Asian, Caucasian, and Hispanic racial identification.

#### Limitations and Future Research

Readers should note that, although the present findings are promising, these remain initial findings and additional research is encouraged to explore and expand on results presented in this paper. Additional research is recommended before conclusive generalizations can be made about the applicability of these findings to the hiring practices of correctional officers. Although it was not the purpose or goal of this paper, a limitation of the present findings is the lack of data measuring the five CH-S scales in relation to job performance measures, supervisory ratings, client satisfaction surveys, and/or other outcome measures of counterproductive work behaviors. Additional research exploring the predictive validity of the CH-S is recommended. Additional research was also recommended by a reviewer of this manuscript, and supported by these authors, that would explore the association between items in the Employment and Legal History section and the five CH-S integrity scales. The Employment and Legal History section was developed to provide users with information that does not require psychometric inferences to interpret, as does the five CH-S integrity scales or IMS, but rather provides the user with a binary (“Yes” or “No”) answer to each historical question (e.g., “Have you ever been terminated?”). Research exploring the degree to which the Employment and Legal History section items are associated with the five CH-S integrity scales, and how the combination of items may have operational or criterion validity in predicting poor job performance and CWBs, could be valuable and is recommended. Readers should also note that the various studies conducted in this paper demonstrated significant differences in regard to statistical power that were less than optimal.

The range of statistical power observed, for example, fell between .08 and 1.0. Therefore, we would expect, and encourage readers to consider, the current findings to be conservative estimates.

## References

- Berry, C. M., Sackett, P. R., & Wiemann, S. (2007). A review of recent developments in integrity test research. *Personnel Psychology*, 60(2), 271-301. doi: 10.1111/j.1744-6570.2007.00074.x
- Brooks, C. (2014, February 19). Employee theft: Why most small businesses don't report it. *Business News Daily*. Retrieved from <http://www.businessnewsdaily.com/5945-why-small-business-employee-theft-goes-unreported.html>.
- Brown, T. S., Jones, W. J., Terris, W., & Steffy, B. D. (1987). The impact of pre-employment integrity testing on employee turnover and inventory shrinkage losses. *Journal of Business and Psychology*, 2(2), 136-149. doi.org/10.1007/BF01014208
- Cohen, J. (1992). A power primer. *Psychological Bulletin*, 112(1), 155-159. doi: 10.1037/0033-2909.112.1.155
- Corey, D. (2008, October). Bifurcation Implications. *Annual Conference of the Society of Policy and Criminal Psychology*. Walnut Creek, California.
- Fatemi, F. (2016, September 28). The true cost of a bad hire – It's more than you think. *Forbes*. Retrieve from <https://www.forbes.com/sites/falonfatemi/2016/09/28/the-true-cost-of-a-bad-hire-its-more-than-you-think/#38e97b494aa4>
- Fine, S. (2013). A look at cross-cultural integrity testing in three banks. *Personnel Review*, 42(3), 266-280. doi: 10.1108/00483481311320408
- 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. doi: 10.1016/j.hrmr.2009.03.010

Griffin v Steeltek, Inc., 964 F. Supp. 317 (N.D. Okla. 1997).

Hart, C. M., Ritchie, T. D., Hepper, E. G., & Gebauer, J. E. (2015). The balanced inventory of desirable responding short form (BIDR-16). *SAGE Open*, 1(4), 1-9. doi: 10.1177/2158244015621113

Hirschi, T., & Gottfredson, M. (1983). Age and the explanation of crime. *The American Journal of Sociology*, 89(3), 552-584. doi: 10.1086/227905

Hunter, J. E. (1986). Cognitive ability, cognitive aptitude, job knowledge, and job performance. *Journal of Vocational Behavior*, 29(3), 340-362. doi: 10.1016/0001-8791(86)90013-8.

Hunter, J. E., & Hunter, R. F. (1984). Validity and utility of alternative predictors of job performance. *Psychological Bulletin*, 96(1), 72-98. doi: 10.1037/0033-2909.96.1.72

Hunter, J. E., & Schmidt, F. L. (1996). Intelligence and job performance: Economic and social implications. *Psychology, Public Policy, and Law*, 2(3/4), 447-472. doi: 10.1037/1076-8971.2.3-4.447.

Hunter, J. E., Schmidt, F. L., & Judiesch, M. K. (1990). Individual differences in output variability as a function of job complexity. *Journal of Applied Psychology*, 75(1), 28-42. doi: 10.1037/0021-9010.75.1.28

Jones, J. W., Cunningham, M. R., & Dages, K. D. (2010). Pre-offer police integrity testing: Scientific foundations and professional issues. In P. A. Weiss (Ed.), *Personality Assessment in Police Psychology: A 21<sup>st</sup> Century Perspective* (pp. 159-187). Springfield, IL: Charles C Thomas Publisher.

Karraker v Rent-A-Center, Inc., 239 F. Supp. 2d 828 (C.D. Ill. 2003).

- Marcus, B., Ashton, M. C., & Lee, K. (2013) A note on the incremental validity of integrity tests beyond standard personality inventories for the criterion of counterproductive behaviour. *Canadian Journal of Administrative Sciences*, 30(1), 18-25. doi 10.1002/cjas.1235
- McGraw, K. O., & Wong, S. P. (1992) A common language effect size statistic. *Psychological Bulletin*, 111(2), 361-365. doi: 10.1037/0033-2909.111.2.361
- Nicol, A. M., & Paunonen, S. V. (2002). Overt honesty measures predicting admissions: An index of validity or reliability. *Psychological Reports*, 90(1), 105-115.
- Ones, D. S., Viswesvaran, C., & Schmidt, F. (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(4), 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(SPEC. 1), S19-S38. doi: 10.1002/per.487
- Paulhus, D. L. (1991). Measurement and control of response bias. In J. P. Robinson, P. R. Shaver, & L. S. Wrightsman (Eds.), *Measures of personality and social psychological attitudes* (pp. 17-59). San Diego, CA: Academic Press.
- Paulhus, D. L. (1994). *Balanced Inventory of Desirable Responding: Reference manual for BIDR Version 6*. Unpublished manuscript, University of British Columbia, Vancouver, Canada.
- Paulhus, D. L. (1998). *Manual for the Paulhus Deception Scales: BIDR Version 7*. Toronto, Ontario, Canada: Multi-Health Systems.
- Profiles International, Inc. (2004). *Construction and Validation Report for the Step One Survey II™*. Waco, Texas: Self-published.

- Ree, M. J., & Earles, J. A. (1992). Intelligence is the best predictor of job performance. *Current Directions in Psychological Science*, 1(3), 86-89. doi: 10.1111/1467-8721.ep10768746
- Rocque, M., Posick, C., & Hoyle, J. (2016). Age and Crime. In W. G. Jennings (Ed.), *The encyclopedia of crime and punishment* (pp. 1-8). Chichester, West Sussex: John Wiley & Sons, Inc.
- Sacket, P. R., Burris, L. R., & Callahan, C. (1989). Integrity testing for personnel selection: An update. *Personnel Psychology*, 42(3), 491-529. doi: 10.1111/j.1744-6570.1989.tb00666.x
- Schmidt, F. L. (2002). The role of general cognitive ability and job performance: Why there cannot be a debate. *Human Performance*, 15(1-2), 187-201. doi: 10.1207/S15327043HUP1501&02\_12.
- 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(2), 262-274. doi: 10.1037/0033-2909.124.2.262
- Schmidt, F. L., Oh, I.-S., Shaffer, J. A. (2016). The Validity and Utility of Selection Methods in Personnel Psychology: Practical and Theoretical Implications of 100 Years of Research Findings. Retrieved from <https://www.testingtalent.net/wp-content/uploads/2017/04/2016-100-Yrs-Working-Paper-on-Selection-Methods-Schmidt-Mar-17.pdf>
- Schmidt, F. L., Schaffer, J. A., & Oh, I.-S. (2008). Increased accuracy of range restriction corrections: Implications for the role of personality and general mental ability in job and training performance. *Personnel Psychology*, 61(4), 827-868. doi: 10.1111/j.1744-6570.2008.00132.x



- Schwantes, M. (2018, January 18). Warren Buffett says if you hire people on intelligence but they lack this other trait, don't bother. *Inc.* Retrieved from <https://www.inc.com/marcel-schwantes/warren-buffet-says-you-should-hire-people-based-on-these-3-traits-but-only-1-truly-matters.html>
- Stabile, S. J. (2002). The use of personality tests as a hiring tool: Is the benefit worth the cost? *University of Pennsylvania Journal of Labor and Employment Law*, 4(2), 279-314. Retrieved from <http://scholarship.law.upenn.edu/jbl/vol4/iss2/1>
- Strahan, R., & Gerbasi, K. C. (1972). Short, homogenous versions of the Marlow-Crown Social Desirability Scale. *Journal of Clinical Psychology*, 28(2), 191-193. doi: 10.1002/1097-4679(197204)28:2<191::AID-JCLP2270280220>3.0CO;2-G
- Tatman, A. (2018). Critical Hire® – Screen Professional Manual. Des Moines, Iowa: Self-published.
- United States Department of Labor, Bureau of Labor Statistics. (2018). *Occupational Employment Statistics: Occupational Employment and Wages, May 2017*. Retrieved from <https://www.bls.gov/oes/current/oes211092.htm>
- Wanek, J. E. (1999). Integrity and honesty testing: What do we know? How do we use it? *International Journal of Selection & Assessment*, 7(4), 183-195. doi: 10.1111/1468-2389.00118

Table 1

*Gender Means, Standard Deviations, t-tests, Cohen's d, and Common Language Effect Sizes**(CLES)*

	Gender	Mean	SD	t	p	d	d Quantified*	CLES
IMS	Male	10.87	1.13	.62	.54			
	Female	10.79	1.24					
Substances	Male	11.27	3.51	3.02	.003	.36	Moderate	.60
	Female	9.99	3.55					
Theft	Male	12.54	2.92	1.40	.16			
	Female	12.00	3.46					
Authority	Male	11.27	2.50	.01	.99			
	Female	11.27	2.46					
Rules & Deception	Male	15.54	3.71	1.60	.11			
	Female	14.87	3.41					
Responsibility	Male	4.41	1.46	4.24	.000	.48	Moderate	.63
	Female	3.79	1.09					

\* (Cohen, 1992)

Table 2

*Age Means, Standard Deviations, t-tests, Cohen's d, and Common Language Effect Sizes (CLES)*

	Age	Mean	SD	t	p	d	d Quantified*	CLES
IMS	Under 40	10.79	1.18	-.54	.587			
	40 +	10.87	1.25					
Substances	Under 40	11.21	3.60	7.56	.000	.93	Large	.74
	40 +	8.31	2.58					
Theft	Under 40	12.91	3.22	6.44	.000	.89	Large	.74
	40 +	10.29	2.60					
Authority	Under 40	11.55	2.33	3.35	.001	.42	Moderate	.61
	40 +	10.49	2.67					
Rules & Deception	Under 40	15.91	3.43	6.79	.000	.94	Large	.74
	40 +	12.93	2.87					
Responsibility	Under 40	4.13	1.29	2.46	.014	.32	Moderate	.58
	40 +	3.72	1.24					

\* (Cohen, 1992)

Table 3

*Correlations Between Critical Hire® - Screen and Step One Survey II™ factors*

Step One Survey™	Critical Hire® - Screen				
	Substances	Theft	Authority	Rules & Deception	Responsibility
Integrity	-.34*	-.42*	-.18**	-.34*	-.21***
Substance Abuse	-.41*	-.32*	-.47*	-.37*	-.40*
Reliability	-.30*	-.39*	-.34*	-.44*	-.30*
Work Ethic	-.43*	-.42*	-.43*	-.50*	-.35*

*Note.* Negative correlations were expected.\*  $p < .001$ . \*\*  $p < .05$ . \*\*\*  $p < .01$ .

Table 4

*Test Re-Test Correlations*

	Short Frequency Group (1-30 days)	Long Frequency Group (31 - 541 days)	Full Sample
IMS	.85	.82	.84
Substances	.91	.71	.80
Theft	.88	.62	.76
Authority	.96	.64	.77
Rules & Deception	.83	.68	.75
Responsibility	.77	.74	.76

*Note:* All correlation coefficients were significant above a  $p$  value of .000.

## Appendix

*Critical Hire®-Screen scales, descriptions, and definitions*

Scale	Scale description and definition
IMS	A 4-item scale that uses a dichotomous (True-False) question format designed to help measure a job applicant's propensity to endorse test items in a highly virtuous or socially desirable manner. Example item: "I have never been mad at a co-worker."
Substances	A 7-item scale that uses a 5-point Likert scale question format designed to help measure a job applicant's attitudes and beliefs around alcohol and illegal drug use in the workplace. Example item: "Using illegal drugs at work is ok if it helps you do your job."
Theft	A 5-item scale that uses a 5-point Likert scale question format designed to help measure a job applicant's attitudes and beliefs around stealing. Example item: "It is ok to steal from work from time to time - everyone does it."
Authority	A 4-item scale that uses a 5-point Likert scale question format designed to help measure a job applicant's attitudes and beliefs around authority and management. Example item: "Supervisors mainly work to make themselves look good."
Rules & Deception	A 7-item scale that uses a 5-point Likert scale question format designed to help measure a job applicant's attitudes and beliefs around rule violations and use of manipulation for personal gain. Example item: "Sometimes you have to cheat to get ahead."
Responsibility	A 2-item scale that uses a 5-point Likert scale question format designed to help measure a job applicant's attitudes and beliefs around blaming victims for crimes committed against them. Example item: "People who get robbed likely left themselves open to it so deserve what they get."