

2018

# MMPI-2-RF UNDERREPORTING VALIDITY SCALES IN FIREFIGHTER APPLICANTS: A CROSS-VALIDATION STUDY

Kullen Charles Balthrop

University of Kentucky, kullen.balthrop@uky.edu

Author ORCID Identifier:

 <https://orcid.org/0000-0003-0895-1575>

Digital Object Identifier: <https://doi.org/10.13023/etd.2018.419>

[Click here to let us know how access to this document benefits you.](#)

## Recommended Citation

Balthrop, Kullen Charles, "MMPI-2-RF UNDERREPORTING VALIDITY SCALES IN FIREFIGHTER APPLICANTS: A CROSS-VALIDATION STUDY" (2018). *Theses and Dissertations--Psychology*. 149.

[https://uknowledge.uky.edu/psychology\\_etds/149](https://uknowledge.uky.edu/psychology_etds/149)

This Master's Thesis is brought to you for free and open access by the Psychology at UKnowledge. It has been accepted for inclusion in Theses and Dissertations--Psychology by an authorized administrator of UKnowledge. For more information, please contact [UKnowledge@lsv.uky.edu](mailto:UKnowledge@lsv.uky.edu).

## **STUDENT AGREEMENT:**

I represent that my thesis or dissertation and abstract are my original work. Proper attribution has been given to all outside sources. I understand that I am solely responsible for obtaining any needed copyright permissions. I have obtained needed written permission statement(s) from the owner(s) of each third-party copyrighted matter to be included in my work, allowing electronic distribution (if such use is not permitted by the fair use doctrine) which will be submitted to UKnowledge as Additional File.

I hereby grant to The University of Kentucky and its agents the irrevocable, non-exclusive, and royalty-free license to archive and make accessible my work in whole or in part in all forms of media, now or hereafter known. I agree that the document mentioned above may be made available immediately for worldwide access unless an embargo applies.

I retain all other ownership rights to the copyright of my work. I also retain the right to use in future works (such as articles or books) all or part of my work. I understand that I am free to register the copyright to my work.

## **REVIEW, APPROVAL AND ACCEPTANCE**

The document mentioned above has been reviewed and accepted by the student's advisor, on behalf of the advisory committee, and by the Director of Graduate Studies (DGS), on behalf of the program; we verify that this is the final, approved version of the student's thesis including all changes required by the advisory committee. The undersigned agree to abide by the statements above.

Kullen Charles Balthrop, Student

Dr. David T. R. Berry, Major Professor

Dr. Mark Fillmore, Director of Graduate Studies

MMPI-2-RF UNDERREPORTING VALIDITY SCALES IN FIREFIGHTER  
APPLICANTS: A CROSS-VALIDATION STUDY

---

THESIS

---

A thesis submitted in partial fulfillment of the  
requirements for the degree of Master of Science in the  
College of Arts and Sciences at the University of  
Kentucky

By

Kullen Charles Balthrop

Lexington, Kentucky

Director: Dr. David T.R. Berry, Professor of Psychology

Lexington, Kentucky

2018

Copyright© Kullen Charles Balthrop 2018

## ABSTRACT OF THESIS

### MMPI-2-RF UNDERREPORTING VALIDITY SCALES IN FIREFIGHTER APPLICANTS: A CROSS-VALIDATION STUDY

The identification of potential underreporting in employment evaluations is important to consider when examining a measure's validity. This importance increases in personnel selection involving high-virtue positions (e.g., police officers and firefighters). The current study aimed to utilize an archival firefighter applicant sample to examine the construct validity of the Minnesota Multiphasic Personality Inventory-2-Restructured Form's (MMPI-2-RF) underreporting scales (L-r and K-r). Results were analyzed using a correlation matrix comprised of a modified version of the Multi-Trait Multi-Method Matrix (MTMM), as well as multiple regression and partial correlation. The present study provides additional support for the construct validity of the MMPI-2-RF's underreporting validity scales. Further research using outcome measures and alternate assessment methods would be able to provide further information on the efficacy of these scales.

**KEYWORDS:** Construct Validity, Personnel Selection, Underreporting, Assessment, Test Validity

*Kullen Balthrop*

*November 2<sup>nd</sup>, 2018*

MMPI-2-RF UNDERREPORTING VALIDITY SCALES IN FIREFIGHTER  
APPLICANTS: A CROSS-VALIDATION STUDY

By  
Kullen Charles Balthrop

Dr. David T.R. Berry  
Director of Thesis

Dr. Mark Fillmore  
Director of Graduate Studies

November 2<sup>nd</sup>, 2018

## TABLE OF CONTENTS

List of Tables.....	iv
Introduction.....	1
Methods.....	7
Results.....	12
Discussion.....	19
References.....	24
Vita.....	28

## LIST OF TABLES

Table 1, Validity Subscale Scores and elevations.....	12
Table 2, Validity Scales Multi-Trait Mono-Method Matrix.....	14

## **MMPI-2-RF Underreporting Validity Scales in Firefighter Applicants:**

### **A Cross-Validation Study**

Underreporting during psychological assessments, through either minimized or denied pathological personality traits, emotions, or behavioral problems, is a salient clinical issue, particularly in high-stakes assessments. Underreporting may artificially lower scores, potentially resulting in false negatives for psychopathology. Situations in which underreporting may be likely to occur include job applicants who wish to highlight their strengths while minimizing their weaknesses; individuals in a psychiatric inpatient setting who want to be released from the hospital; and parent litigants trying to improve their chances of being awarded child custody (Sellbom & Bagby, 2008). Psychological assessments aim to address this issue through various “fake good” or underreporting indices to estimate whether examinees’ responses are valid.

Malingering and underreporting research generally makes use of three types of study designs: simulation, differential prevalence, and known-groups. Simulation designs with college student samples are frequently used to assess an underreporting scale’s ability to discriminate between faking good and honest examinees. However, using college students as the experimental control, instead of groups who may be more prone to underreport, hinders the generalizability of their findings. Differential prevalence designs mitigate this somewhat, as this method involves participants who are likely naturally motivated to underreport. For example, applicants to high virtue jobs (e.g., police officer or firefighter) have been shown to have higher incidences of underreporting (Corey & Ben-Porath, 2014; Lowmaster & Morey, 2012). A differential prevalence design would obtain data from their application process, rather

than undergraduate students. This allows the results obtained to be more generalizable to the desired population. However, no precise information about the validity scales being measured can be identified with this design, as there is no clear underreporting group. Known-groups designs using clinically relevant samples as the experimental and control groups are the ideal way to investigate the validity of underreporting scales, but the most difficult. Known-groups designs allow comparisons in a naturalistic setting (i.e. in a preemployment scenario) between examinees shown to respond honestly versus those found to underreport. However, this method requires an underreporting criterion to determine examinees' respective groups. Known-groups research utilizing more than one underreporting index, and with a population that is generalizable, is extremely rare (Baer & Miller, 2002).

Effect sizes between honest and underreporting responders have been found to be much lower for underreporting scales ( $d = 1.05$ ) than for overreporting scales ( $d = 2.05$ ), despite both types of validity research using similar methodologies (Baer, Wetter, & Berry, 1992). It is possible that underreporting of psychopathology lacks unidimensionality that is present in overreporting research (Baer et al., 1992). Someone being instructed to act as if he or she has ADHD has a much clearer goal than someone who is told to underreport. In a similar vein, it is likely that test-takers asked to respond in a socially desirable manner may not respond the same way as those told to respond in a way that creates a positive impression. It is also possible underreporting takes different forms in different contexts, i.e. someone in a child custody case may underreport differently than someone applying for a high-virtue job. Analyses of studies using similar instructions in simulation designs, or who are researching specific underreporting populations, may yield effect sizes closer to those found in underreporting research.

The Minnesota Multiphasic Personality Inventory- 2 (MMPI-2; Butcher et al, 2001) and the contemporary MMPI-2 Restructured Form (MMPI-2-RF; Ben-Porath & Tellegen, 2008/2011) are widely used personality and psychopathology measures that have long included scales designed to detect underreporting. The MMPI-2 validity scales are well established. In particular, the underreporting validity scales L and K exhibit strong classification accuracies and large effect sizes,  $d = 1.19$  and  $d = 1.13$  respectively (Baer & Miller, 2002). The MMPI-2-RF was later developed as a shorter form of the test with the goal of providing updated psychometrics and greater efficiency than its predecessors, while measuring the same core clinical constructs (Ben-Porath & Tellegen, 2008/2011). The MMPI-2-RF validity scales are largely based on those from the MMPI-2, including the L-r (i.e., Uncommon Virtues) and K-r (i.e., Adjustment Validity) scales. However, cross-validation among different samples and contexts that might provoke underreporting is also necessary to ensure that the MMPI-2-RF is capable of detecting underreporting in the real world contexts the form will be used (e.g., public service personnel selection).

The underreporting scales on the MMPI-2-RF have been commonly validated through simulation studies (Crighton, Marek, Dragon, & Ben-Porath 2016; Rogers, 2008; Sellbom & Bagby, 2008). A simulation study involving a manipulation check reported large effect sizes between the standard compliant group (individuals in the standard group who complied with instructions) and the underreporting compliant group (individuals in the underreporting group who followed instructions) on the RF's L-r ( $g = -1.50$ ) and K-r ( $g = -1.34$ ) scales (Crighton et al., 2016). Similarly, another simulation study found large effect sizes between the honest and underreporting groups on the L-r and K-r scales (Sellbom & Bagby, 2008).

However, underreporting scales' validity may vary across different contexts. Among child custody litigants, the MMPI-2 and MMPI-2-RF validity scales were highly correlated (Archer, Hagan, Mason, Handel, & Archer, 2012; Kauffman, Stolberg, & Madero, 2015). One study examined the impact of demand characteristics on underreporting. Police officers were administered the MMPI-2-RF twice, once during the preemployment evaluation (high demand condition) and again after being admitted to candidate school and completing their training (low demand condition). At the second assessment, the police officers were told their scores would not be seen by the police department and would have no consequences. Underreporting scores significantly decreased and most of the clinical scale scores increased at the second assessment (Detrick & Chibnall, 2014), suggesting that underreporting occurred in the high demand context of applying for a police officer position. This study's findings corroborate prior research showing that law enforcement officer candidates had higher mean scores on the underreporting validity scales than average respondents (Tarescavage, Corey, Gupton, & Ben-Porath, 2015). Applicants to high-virtue jobs such as law enforcement may fake good to ensure that their profile matches the levels of virtue expected of employees in these fields. It is important then, that underreporting scales are empirically validated within different clinical contexts to ensure that their application is valid across all contexts in which they may be used.

However, previous research has not explicitly investigated whether the MMPI-2-RF's underreporting scales accurately represent underreporting as a construct. The current study examines the construct validity of the MMPI-2-RF's underreporting scales (L-r and K-r) involving fire fighter applicants. A common method used to establish construct validity for psychological tests is the Multi-Trait Multi-Method (MTMM; Campbell & Fiske, 1959), which assesses both convergent and discriminant

validity. The most typical way to validate a measure of a construct is through convergent validity – generally done by correlating the measure in question with another measure assessing the same construct. However, discriminant validity is also important. This ensures the measure being investigated is not highly correlated with those assessing a different construct (Campbell & Fiske, 1959).

In the present study, convergent validity of the MMPI-2-RF's underreporting scales was assessed using the underreporting scales taken from the Paulhus Deception Scales (PDS; Paulhus, 1998) and the Hilson Safety/Security Risk Inventory (HSRI; Inwald, 1995). The PDS includes two scales: Impression Management (IM) and Self-Deceptive Enhancement (SDE). The IM scale assesses engagement in behaviors often considered desirable, but which rarely occur (e.g., never stealing office supplies). Higher scores on this scale suggest that the respondent may be attempting to create an unrealistically positive impression of him or herself (Paulhus, 1998). The IM scale can also be used as a validity check for other measures, as it has been shown to successfully distinguish underreporters from honest respondents (Paulhus, Bruce, & Trapnell, 1995; Paulhus, 1991). The SDE scale assesses an unconscious positive bias similar to narcissism (Paulhus & John, 1998). Higher scores on this scale suggest a very strict and unyielding overconfidence. The third index used to assess convergent validity in the present study is the Defensive Response Scale (DEF) from the HSRI, intended to detect exceptionally high levels of defensive responding (i.e. not admitting minor faults, attempting to appear exceptionally virtuous; Inwald, 1995).

Discriminant validity of the MMPI-2-RF's underreporting scales was assessed using the MMPI-2-RF's Variable Response Inconsistency Index (VRIN-r) and the True Response Inconsistency Index (TRIN-r). Both scales measure non-content-based invalid responding. The VRIN-r scale is used to ensure that respondents are not

answering randomly, without regard to the content of the question. Similarly, the TRIN-r scale is used to ensure respondents are not answering in a fixed manner, such as all true or all false, regardless of the content of the question. Individuals who commit this type of responding are unconcerned with the content of the question being asked. In contrast, when content-based invalid responding occurs, the individual will change his or her answer based on the content of the question. Since the MMPI-2-RF's underreporting scales (L-r and K-r) measure content-based invalid responding, little correlation is expected between them and the non-content-based invalid response scales (VRIN-r and TRIN-r).

As noted above, previous research indicates that underreporting may be expected in candidate evaluations for employment in high-risk positions (Corey & Ben-Porath, 2014; Lowmaster & Morey, 2012). The current study examines the construct validity of the MMPI-2-RF's underreporting scales (L-r and K-r) in the context of personnel selection involving high virtue positions. This analysis may provide additional support for the construct validity of the MMPI-2-RF's underreporting scales.

The current study aims were:

Aim 1: Provide further convergent validity for the MMPI-2-RF's underreporting scales in the context of personnel selection involving high virtue positions.

Hypothesis 1: It was hypothesized that the MMPI-2-RF's L-r and K-r scales would have moderate correlations with the PDS underreporting scales (i.e., the SDE and IM scales) and the HSRI underreporting scale (i.e., the DEF scale).

Aim 2: Provide further discriminant validity for the MMPI-2-RF's underreporting scales in the context of personnel selection involving high virtue positions.

Hypothesis 2: It was hypothesized that the MMPI-2-RF's L-r and K-r scales would have no-to-low correlations with the MMPI-2-RF non-content-based responding scales (i.e., the TRIN-r and VRIN-r scales).

Aim 3: Investigate the amount of variance explained by the HSRI's Defensiveness scale and the PDS's Impression Management and Self-Deception scales on the MMPI-2-RF's underreporting scales.

Hypothesis 3: It was hypothesized the IM and DEF scales would predict the most variance for the L-r and K-r scales of the MMPI-2-RF.

## **Methods**

### **Participants**

Data were obtained from an archival sample of 209 applicants to the Lexington Fire Department (202 men, 7 women). Only those who gave written consent allowing the University of Kentucky to use their data for research were included in the present sample. The University of Kentucky Institutional Review Board approved the research project. Data from this archival sample have not been previously published. Examinees were excluded from analyses if their scores indicated Non-Content-Based Invalid Responding on the MMPI-2-RF (i.e., raw scores  $\geq 18$  on the Cannot Say scale and/or a T score  $\geq 80$  on the True Response Inconsistency-r scale or the Variable Response Inconsistency-r scale (Ben-Porath & Tellegen, 2008/2011). It is possible that excluding participants who elevate on the VRIN-r and TRIN-r scales could cause a restriction of range on these variables, potentially contaminating discriminant validity coefficients. However, no participants

from the selected data set met the exclusion criterion for these scales, reducing any risk of range-restriction on these scales.

## **Procedure**

Firefighter applicants first underwent pre-employment screening by the Lexington Fire Department via a series of interviews and background checks. Applicants who passed the pre-screening phase were given a conditional employment offer and then referred to the Jesse G. Harris Psychological Services Center for a psychological assessment. Next, applicants were administered a psychological assessment battery in a group setting including the MMPI-2, PDS, and HSRI. All participants were instructed to answer each measure honestly and to the best of their ability. All applicants then received a structured interview at the Harris Center conducted by a licensed clinical psychologist and a doctoral student. In this study, the MMPI-2 had been administered in the first, group setting, but the MMPI-2-RF was scored using the 338 items selected from the MMPI-2. Scores which have been derived from the MMPI-2 have shown to be interchangeable with scores taken from the direct administration of the MMPI-2-RF (Ben-Porath & Tellegen, 2008/2011).

## **Measures**

### **Underreporting Indices.**

*The Paulhus Deception Scale* (PDS; Paulhus, 1998). The PDS is a 40-item self-report measure of socially desirable responding, which is based on the Balanced Inventory of Desirable Responding (Paulhus, 1998). Items are rated on a 5-point scale. The PDS consists of two subscales; the Self-Deceptive Enhancement (SDE) subscale, an index of one's unconscious favorability towards the self, similar to narcissism, and the Impression Management (IM) subscale, an index of socially

desirable responding, through endorsement of uncommon desirable behaviors. Importantly, the IM and SDE subscales have a low intercorrelations (.20) and reflect different aspects of underreporting (Paulhus, 1998). The PDS has previously demonstrated adequate internal consistency (i.e.,  $\alpha > .80$ ) across several samples including general population ( $\alpha = .85$ ), college students ( $\alpha = .83$ ), prison entrants ( $\alpha = .86$ ), and military recruits ( $\alpha = .85$ ; Paulhus, 1998). The PDS SDE subscale has demonstrated convergent validity with the Edwards Social Desirability scale and the Personality Research Form desirability subscale, and the PDS IM subscale with the Eysenck's Personality Inventory's Lie scale and the Minnesota Multiphasic Personality Inventory's L scale (Paulhus, 1998). Raw scores are converted to T-scores, with T-scores  $> 70$  suggesting significant underreporting (Paulhus, 1998). As noted above, in the present study, the PDS IM and SDE subscales were used to assess the convergent validity of the MMPI-2-RF's underreporting scales.

***The Hilson Safety/Security Risk Inventory.*** The HSRI Inventory assesses an individual's tendency to engage in risk-taking and unsafe behaviors in the workplace (Inwald, 1995). The measure contains 178 statements that require either a true or false endorsement. The HSRI converts raw scores into standardized T-scores, with those  $> 69$  signaling significant elevation. This measure contains 3 higher order scales: Safety Risk ( $\alpha = .83$ ), Hostility/Anger Control ( $\alpha = .87$ ), and Preparation Concerns ( $\alpha = .62$ ; Inwald, 2008). The DEF subscale is an index of how guarded or defensive the test-taker's responses are. In the present study, this subscale was used to assess convergent validity of the MMPI-2-RF's underreporting scales.

***MMPI-2-RF underreporting indices.*** The L-r and K-r scales seek to measure the test-taker's tendencies to underreport. The L-r scale is designed to detect endorsement of unlikely virtuous behavior. It contains 14 items: 11 from the MMPI-

2's L scale, two from the Social Desirability (Sd) scale, and one from the Positive Malingering (Mp) scale. Three of the 14 items are keyed true (as opposed to all the L scale items being keyed false) making the L-r scale more robust to fixed responding. A L-r T-score of 80 or higher indicates an invalid protocol. The K-r scale contains 14 of the 30 items from the K scale. This scale assesses (exaggeration of?) the test-taker's psychological adjustment. A T-score of 70 or higher on this scale indicates an invalid protocol.

### **Discriminant Validity Indices.**

***MMPI-2-RF non-content-based validity indices.*** The VRIN-r and TRIN-r subscales seek to measure the test-taker's random and fixed response tendencies. Item pairs were restricted to positively correlated item pairs for VRIN-r and negatively correlated item pairs in the TRIN-r to reduce the overlap that existed between the scales' predecessors on the MMPI-2. This resulted in 53 item pairs for the VRIN-r and 26 for the TRIN-r. A T-score of 80 or more on either scale renders the protocol invalid. These indices are measuring a form of validity theoretically different than underreporting. Underreporting requires the test-takers to read each item and respond in a manner that will either increase their positive attributes or remove/minimize their negative attributes. Non-content-based responding does not require the test-taker to read the question at all, in fact, these indices are designed to detect individuals who answer without regard to the content of the question. Due to the theoretical and practical differences in these two validity types, the VRIN-r and TRIN-r scales were used to assess discriminant validity of the MMPI-2-RF's underreporting scales.

### **Statistical Analysis**

**Preliminary Analyses.** Preliminary analyses were conducted to examine the data distribution and adherence to statistical assumptions of Pearson's  $r$  and multiple regression. Tests for normality revealed none of the assumptions for Ordinary Least Squares (OLS) regression were violated. No missing data were present in the sample, and statistical analysis did not detect any outliers (data points more than three standard deviations from the mean). Assessment of the correct form of association was done using a scatter plot fitted with a Locally Weighted Scatterplot Smoother (LOESS) line. Each scatterplot represented the relationship between one of the dependent variables (L-r and K-r) and the independent variables being used (IM, SDE, and DEF). This led to a total of six scatterplots, all of which showed a positive linear relationship between the independent and dependent variables being measured. Due to the study's cross-sectional design and lack of nesting, it was deemed the residuals would be independent of each other. A Quantile-Quantile (Q-Q) plot suggested the normality of residuals assumption had been met. The assumption of no multicollinearity of the predictor variables was met, as none of the independent variables has an  $r \geq .8$ . A residual plot of each independent variable suggested that the assumption of heteroscedasticity had been met. Both conscious and unconscious aspects of underreporting were to be used as predictors, with each predictor targeting a different aspect of the underreporting construct. Based on these factors, the model should be correctly specified.

### **Validity Analyses.**

Analyses were conducted in SPSS (Version 23), with alpha set at  $p < .05$ , two-tailed, for all inferential tests. First, a Multi-Trait Mono-Method correlation matrix was used to assess convergent and discriminant construct validity among the scales of interest (IM, SDE, K-r, L-r, DEF, VRIN-r, and TRIN-r). Significance difference

testing was used to gain further insight into the correlations present in the matrix. Second, two-stage hierarchical multiple regression models were used to predict MMPI-2-RF underreporting scales (L-r and K-r) from the other comparable scales (PDS IM and SDE subscales and the HSRI DEF subscale) to investigate the amount of variance in the L-r and K-r scales that can be explained by the IM, SDE, and DEF scales. This provides potentially better understanding of which aspects of underreporting L-r and K-r are measuring. Changes in the  $R^2$  and beta weights were evaluated for statistical significance for each predictor. Partial correlations were calculated for each predictor on both the L-r and K-r scale.

## Results

The means and standard deviations of the underreporting and non-content-based responding subscales can be seen in Table 1. None of the means of these scales was above the cut-score which would signify significant levels of underreporting. On average, this sample did not underreport on their assessment.

Table 1. Validity Subscale Scores and Elevations

Scale	T-score	% Sig. Elevations
PDS Impression Management	61.33 (9.5)	46 (23%)
PDS Self-Deceptive Enhancement	61.60 (13.6)	56 (28%)
HSRI Defensiveness	47.95 (8.5)	2 (1%)
MMPI-2-RF L-r	52.58 (9.9)	2 (1%)
MMPI-2-RF K-r	62.5 (7.8)	27 (14%)
MMPI-2-RF VRIN-r	40.06 (6.2)	0 (0%)
MMPI-2-RF TRIN-r	53.56 (4.9)	0 (0%)

---

*Note.* Means and frequencies; standard deviations and percentages are provided in parentheses. L-r = Unlikely Virtuous Behavior, K-r = Psychological Adjustment, VRIN-r = Variable Response Inconsistency Index – revised, TRIN-r = True Response Inconsistency Index-revised.

For the PDS and IM scales, T-scores > 70 are significantly elevated and suggest underreporting. For the L-r scale, T-Scores > 80 are significantly elevated and suggest underreporting. For the K-r scale, T-Scores > 70 are significantly elevated and suggest underreporting. For the DF scale, T-scores > 69 are significantly elevated and suggest underreporting. For the VRIN-r scale, T-scores > 80 are significantly elevated and suggest random responding. For the TRIN-r scale, T-scores > 80 are significantly elevated and suggest fixed responding.

---

Despite this, greater than 25% of the subjects elevated on the PDS, indicating there was some level of underreporting present. Further, the K-r mean T-score shown is similar to the mean T-score of the underreporting group in previous simulation studies (Crighton et al., 2017; Sellbom & Bagby, 2008), as well as police officer applicants (Detrick & Chibnall, 2014). Even though the L-r mean T-score in this study is lower than mean scores from previous studies, the T-scores remain within one standard deviation of each other (Detrick & Chibnall, 2014; Sellbom & Bagby, 2008). A study investigating the MMPI-2-RF profile of child custody litigants, another group thought to be predisposed towards underreporting, had a similar mean T-score for the L-r scale (Archer et al., 2012). If the current study were a simulation study, or one incorporating outcome measures, the low L-r scores would be a problem – it would make it impossible to have an underreporting L-r comparison group. However, one

can still look at the relationship between underreporting scale scores even if a lower number underreport.

The Multi-Trait Mono-Method Matrix for the validity scales is presented in Table 2. For the underreporting scales, intercorrelations were generally small to medium (Cohen, Cohen, West, & Aiken, 2003). Strong correlations were observed between L-r and DEF as well as IM and DEF. Most of the correlations were of medium strength. Medium correlations were also observed between L-r and K-r, L-r and IM, L-r and SDE, K-r and IM, K-r and DEF, and SDE and DEF. Small correlations were observed between IM and SDE. However, the K-r and SDE exhibited a small and non-significant correlation.

In contrast, for the MMPI-2-RF non-content-based responding scales, correlations with the underreporting scales were generally small and non-significant, with one notable exception; the VRIN-r had a medium negative correlation with the K-r scale. The non-content based responding scales themselves had a small positive correlation with one another.

**Table 2.** Validity Scales Multi-Trait Mono-Method Matrix

	L-r	K-r	IM	SDE	DEF	VRIN-r	TRIN-r
L-r	-						
K-r	.35*	-					
IM	.49*** <sup>a</sup>	.30**	-				
SDE	.30**	.14	.24**	-			
DEF	.61*** <sup>a</sup>	.30**	.51**	.34**	-		
VRIN-r	-.13 <sup>a</sup>	-.43**	-.14	-.09	-.13	-	
TRIN-r	.05	-.05	.06	-.08	.05	.27**	-

---

*Note.* \*\* =  $p < .01$ ; a = significant difference between L-r and K-r was found. L-r = Unlikely Virtuous Behavior, K-r = Psychological Adjustment, IM = Impression Management, SDE = Self-Deceptive Enhancement, DEF = Defensiveness, VRIN-r = Variable Response Inconsistency Index – revised, TRIN-r = True Response Inconsistency Index-revised.

---

Significant difference testing revealed L-r had a higher correlation than K-r with the Impression Management and Defensiveness subscales, while K-r had a higher correlation than L-r with the VRIN-r subscale.

In order to investigate the ability of the other underreporting scales (SDE, IM, and DEF) to predict the MMPI-2-RF's underreporting scales (L-r and K-r) Four, two-stage hierarchical multiple regression analyses were conducted. In these analyses, L-r and K-r were the dependent variables. The independent variables consisted of IM and DEF paired together to represent the aspect of conscious underreporting, with SDE representing unconscious underreporting.

#### ***Unlikely virtue predicted by conscious and unconscious underreporting***

As this regression model is investigating the conscious (IM and DEF scales) and unconscious (SDE scales) underreporting scales ability to predict the L-r scale, the L-r scale was entered as the dependent variable. The SDE subscale was entered first to control for Self-Deceptive Enhancement (unconscious underreporting). The second stage had the IM and DEF (conscious underreporting) variables included. The two-stage hierarchical multiple regression revealed that at stage one, the SDE subscale contributed significantly to the regression model ( $R^2 = .30$ ,  $F(1, 199) = 19.09$ ,  $p < .01$ ) and accounted for 8.8% of the variation in the L-r subscale.

Introducing the Conscious Underreporting variables explained an additional 33.7% of the variation in the L-r subscale and this change in  $R^2$  was significant ( $R^2 = .65$ ,  $F(1,199) = 57.33$ ,  $p < .01$ ). In this model, the IM subscale and the DF subscale were the most significant predictors of the L-r subscale, accounting for 33.7% of its variation. Together, the 3 independent variables accounted for 42.5% of the variation in L-r. It was found that impression management ( $\beta = .24$ ,  $p < .001$ ) and defensiveness ( $\beta = .46$ ,  $p < .001$ ) significantly predicted endorsement of unlikely virtuous behavior. This indicates that both groups of the underreporting scales were able to account for a significant amount of variance in the L-r scale's scores. Further, impression management and defensiveness were able to predict endorsement of unlikely virtuous behavior, while self-deceptive enhancement did not.

#### ***Unlikely virtue predicted by conscious and unconscious underreporting 2***

In the second hierarchical regression analysis, L-r was once again the dependent variable. The IM and DF subscales were entered first to control for the component of underreporting pertaining to socially desirable responding. The SDE subscale was entered in the second stage. The second analysis revealed that at Stage one, the Conscious Underreporting subscales contributed significantly to the regression model ( $R^2 = .65$ ,  $F(1,199) = 70.96$ ,  $p < .01$ ) and accounted for 41.9% of the variation in L-r. Introducing the Unconscious Underreporting variables explained an additional .6% of the variance, though this change in  $R^2$  was not significant ( $R^2 = .65$ ,  $F(1,199) = 1.98$ ,  $p = .161$ ). As in the above model, the three independent variables accounted for 42.5% of the variation in L-r. As found in the above model, impression management ( $\beta = .24$ ,  $p < .001$ ) and defensiveness ( $\beta = .46$ ,  $p < .001$ ) significantly predicted endorsement of unlikely virtuous behavior. This shows the SDE subscale

did not account for a significant amount of variance in the L-r subscale, after the variance from the IM and DEF subscales was accounted for. As with the first model, impression management and defensiveness were able to predict endorsement of unlikely virtuous behavior, while self-deceptive enhancement could not. This shows the L-r subscale is more similar to other scales measuring conscious underreporting, with the unconscious underreporting index unable to account for either significant variance or predictive ability in this scale.

### ***Psychological adjustment predicted by conscious and unconscious underreporting***

As this regression model is investigating the conscious (IM and DEF scales) and unconscious (SDE scales) underreporting scales' ability to predict the K-r scale, the K-r scale was entered as the dependent variable. The SDE subscale was entered first to control for self-deceptive enhancement (unconscious underreporting). The second stage had the IM and DEF (conscious underreporting) variables included. The hierarchical multiple regression revealed that at Stage one, the SDE subscale contributed significantly to the regression model ( $R^2 = .14$ ,  $F(1, 199) = 4.22$ ,  $p < .05$ ) and accounted for 2.1% of the variation in the K-r subscale. Introducing the Conscious Underreporting variables explained an additional 9.8% of the variation in the K-r subscale and this change in  $R^2$  was significant ( $R^2 = .35$ ,  $F(1, 199) = 10.93$ ,  $p < .01$ ). In this model, the IM and DF subscales were the most significant predictors of the K-r subscale, accounting for 9.8% of its variation. Together, the 3 independent variables accounted for 11.9% of the variation in K-r. Both impression management ( $\beta = .20$ ,  $p < .05$ ) and defensiveness ( $\beta = .19$ ,  $p < .05$ ) significantly predicted psychological adjustment. This indicates that both groups of the underreporting scales were able to account for a significant amount of variance in the K-r scale's scores.

Further, impression management and defensiveness were able to predict endorsement of psychological adjustment, while self-deceptive enhancement could not.

### ***Psychological adjustment predicted by conscious and unconscious underreporting 2***

In this analysis, K-r was once again the dependent variable. The IM and DF subscales were entered first to control for the component of underreporting pertaining to socially desirable responding. The SDE subscale was entered in the second stage. The second analysis revealed that at Stage one, the Conscious Underreporting subscales contributed significantly to the regression model ( $R^2 = .34$ ,  $F(1,199) = 13.19$ ,  $p < .01$ ) and accounted for 11.8% of the variation in K-r. Introducing the Unconscious Underreporting variables explained an additional .1% of the variance, though this change in  $R^2$  was not significant ( $R^2 = .34$ ,  $F(1,199) = .226$ ,  $p = .635$ ). As in the above model, the three independent variables accounted for 11.9% of the variation in K-r. Thus, both impression management ( $\beta = .20$ ,  $p < .05$ ) and defensiveness ( $\beta = .19$ ,  $p < .05$ ) significantly predicted psychological adjustment. This shows the SDE subscale did not account for a significant amount of variance in the K-r subscale, after the variance from the IM and DEF subscales was accounted for. As with the first model, impression management and defensiveness were able to predict endorsement of psychological adjustment, while self-deceptive enhancement did not. This shows the K-r subscale is more similar to other scales measuring conscious underreporting, with the unconscious underreporting index unable to account for either significant variance or predictive ability in this scale. However, 88.1% of the variance in this scale remains unaccounted for. This indicates this scale is measuring something distinct from the conscious and unconscious aspects of underreporting the independent variables were measuring.

## Discussion

The present study provides additional support for the convergent and discriminant validity for the MMPI-2-RF underreporting scales. It also provides insight into how conscious and unconscious underreporting differ in the context of personnel selection.

Consistent with the first hypothesis, medium correlations between the different underreporting scales were found. Normally, moderate relationships between scales that are intended to index the same construct suggest such scales may not be measuring what they say they are. However, the MMPI-2-RF and PDS underreporting indices were created to measure somewhat different aspects of the same construct (Tellegen & Ben-Porath, 2008/2011; Paulhaus, 1998). Moderate relationships between the scales indicate some overlap, most likely at the broad overall construct of underreporting, but that differences among these scales exist.

Significance difference testing allows for further insight into the correlations found in the matrix. It showed that the L-r scale had a significantly higher correlation with the IM and DEF scales than the K-r scale. This indicates that L-r taps more of the impression management and guardedness aspects of underreporting than the K-r scale does (Inwald, 2008; Paulhus, 1998; Tellegen & Ben-Porath, 2008/2011). All three of these scales were designed to gain insight into the deliberate, conscious, aspect of underreporting. Individuals who would elevate on these scales would be more likely to have endorsed virtuous behavior that is highly unlikely to occur. A conscious underreporter will be more likely to respond in a guarded manner, in order to appear in a more positive light. They may also be engaging in a response style deliberately designed to present a certain type of impression on the employer. Each of these response styles is likely to overlap with the other. Someone trying to present as highly

moral or ethical may endorse items that represent unlikely virtuous behavior, for example. An individual who responds in a guarded manner may also endorse items designed to detect positive impression management.

K-r, however, was designed to measure psychological adjustment. This scale is not a measure of overt, conscious underreporting. In fact, it has been shown to align more with the SDE scale, of the PDS, an index of unconscious underreporting (Sellbom & Bagby, 2008). Given that L-r, IM, and DF are measuring more similar aspects of underreporting than K-r, it is logical that the L-r correlated more highly with the DEF and IM management scales than K-r did.

Consistent with the second hypothesis, there was no significant correlation between the L-r scale and the non-content-based responding scales. However, contrary to the hypothesis, K-r had a significant moderate correlation with the VRIN-r scale. This suggests the K-r scale may be more susceptible to score fluctuation caused by random responding than the MMPI-2-RF's other underreporting scale, the L-r scale, as well as the other underreporting scales investigated. This contrasts with the findings by Burchett et al. (2015). They found that one of the MMPI-2-RF's overreporting scales, FBS-r, and an underreporting scale, K-r, were unaffected by high scores on the TRIN-r and VRIN-r subscales. The findings by Burchett et al. (2015) would indicate a robustness against fixed and random responding, rather than a susceptibility. However, Tellegen & Ben-Porath (2008/2011) found a similar correlation between K-r and VRIN-r. Further research should be done in order to better understand why this correlation has appeared in some studies but not others.

Consistent with the third hypothesis, the IM and DEF scales accounted for the most variance in the L-r and K-r scales, over and above the SDE scale. For both the

L-r and K-r scales, SDE only accounted for a significant amount of variance when put in stage one of the analysis. This indicates that for both scales, whatever amount of variance SDE can provide is accounted for by the IM and DEF scales. SDE, much like K-r, is thought to measure a more unconscious aspect of underreporting than L-r, IM, and DF. In this study conscious underreporting may be the main form of underreporting, as it would allow individuals to appear better qualified for the high-virtue position to which they are applying. Considering the context of this assessment, the SDE scale – a measure of unconscious underreporting, should not account for much variance (Paulhus, 1998).

These findings suggest convergent validity between L-r and K-r and other measures of underreporting in a personnel selection context. When examining the two scales' utility in underreporting detection, results raise the possibility that L-r has stronger underreporting validity than K-r. It had higher correlations with the other measures of conscious underreporting (IM and DEF) than K-r, as well as more of its variance explained by those measures. Moreover, it did not exhibit a correlation with any of the non-content-based responding scales. This indicates the scores presented on this scale are not influenced by another construct. However, it is important to keep the underreporting context in mind. It is likely that this sample was biased towards a more deliberate and overt style of underreporting than a sample in another underreporting context might be. This probably favors the L-r scale, and another context could lead to a bias leading to its results showing K-r to be stronger.

### ***Limitations and future directions***

While this study was able to provide further support for the construct validity of the MMPI-2-RF's underreporting scales, limitations are present. The data were

collected as part of a personnel selection context. Further, these data were collected during the final stage of the application process. As such, only the most qualified applicants were assessed. It is possible applicants who may have been more likely to underreport had already been excluded from the applicant pool. This sample most likely represents the final stage of applicants to high risk public service positions, and not all applicants. However, this may not be uncommon, as most psychological assessment is likely done with the final stage of applicants. Despite not having access to the earlier stages of applicants, the ability of the MMPI-2-RF to distinguish underreporting at the final stage is crucial. Another issue is that these data were collected as part of a personnel selection process; thus no race or ethnicity data were collected. This was done to ensure race or ethnicity played no part when recommending personnel, but it does mean an additional limitation is present in this study. Due to the lack of race or ethnicity data, it is not possible to parse out if race or ethnicity would have any impact on the results of this study. It is also important to note greater than 90% of the applicants were male. However, it is likely this is representative of individuals applying to this position.

A statistical limitation of this study exists as well. Due to the data available, a true MTMM matrix (one containing alternate assessment methods) was not able to be used. An alternate method would have allowed for the ability to rule out method variance as a contributor to the correlations. Finally, no outcome measures were present in this study. This reduces the study's ability to understand how effective the MMPI-2-RF underreporting scales are at detecting underreporting in a personnel selection sample.

This study highlights the need for future research in this area. One question this study was unable to answer regarding the K-r scale is noteworthy. That is, are

elevations of this scale due to underreporting, or due to the candidates themselves (who have made it to the final round of candidate selection) being honest? Many of the final applicants at this stage do seem to have extremely high psychological adjustment and exhibit traits that would make them strong candidates. However, clinical judgements are not always accurate, making it important to have a form of assessment capable of rooting out those who are truly great candidates and those who are merely pretending to be. Since it seems the use of the MMPI-2-RF in personnel selection of high moral value positions (e.g., police officers and fire fighters) is not going to go away, this question should be a target for future research. Further, studies involving outcome measures would allow to better evaluate the effectiveness of the underreporting scales in discriminating between honest and feigning applicants. A study using outcome measures highlighting the effectiveness of the new hires would also allow researchers to look at the efficacy of a psychological assessment battery at selecting the best applicants for the job.

### ***Conclusion***

The present study provides support for the construct validity of the MMPI-2-RF underreporting scales. However, further research is needed to better establish the efficacy of the MMPI-2-RF's underreporting scales in the context of personnel selection, particularly using alternate assessment methods (e.g., a structured interview) and outcome measures (e.g., new hire effectiveness, re-administration of assessment battery to new hires).

## References

- Archer, E. M., Hagan, L. D., Mason, J., Handel, R., & Archer, R. P. (2012). MMPI-2-RF characteristics of child custody litigants. *Assessment, 19*, 14-20. doi: 10.1177/1073191110397469
- Baer, R. A., & Miller, J. (2002). Underreporting of psychopathology on the MMPI-2: A meta-analytic review. *Psychological Assessment, 7*, 419-423.
- Baer, R. A., Wetter, M. W., & Berry, D. T. R. (1992). Detection of underreporting of psychopathology on the MMPI: A meta-analysis. *Clinical Psychology Review, 12*, 509-525.
- Ben-Porath, Y. S., & Tellegen A. (2008/2011). *MPI-2-RF: Manual for administration, scoring, and interpretation*. Minneapolis: University of Minnesota Press.
- Butcher, J.N., Graham, J.R., Ben-Porath, Y.S., Tellegen, A., Dahlstrom, W.G., & Kaemmer, B. (2001). *Minnesota Multiphasic Personality Inventory-2: Manual for administration, scoring, and interpretation* (Rev. ed.). Minneapolis: University of Minnesota Press.
- Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin, 56*(2), 81-105.
- Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences- Third edition*. Mahwah, NJ: Lawrence Erlbaum Associates Publishers.
- Corey, D. M., & Ben-Porath, Y. S. (2014). *MMPI-2-RF (Minnesota Multiphasic Personality Inventory-2 Restructured Form): User's guide for the Police*

*Candidate Interpretive Report*. Minneapolis, MN: University of Minnesota Press.

Crichton, A. H., Marek, R. J., Dragon, W. R., & Ben-Porath, Y. S. (2017). Utility of the MMPI-2-RF validity scales in detection of simulated underreporting: Implications of incorporating a manipulation check. *Assessment, 24*(7), 853-864. doi: 10.1177/1073191115627011

Detrick, P., & Chibnall, J. T. (2008). Positive response distortion by police officer applicants: Association of Paulhus Deception Scales with MMPI-2 and Inwald Personality Inventory validity scales. *Assessment 15*(1), 87-96.

Detrick, P., & Chibnall, J. T. (2014). Underreporting on the MMPI-2-RF in a high-demand police officer selection context: An illustration. *Psychological Assessment, 26*, 1044-1049. doi: 10.1037/pas0000013

Inwald, R. (1995). *Hilson Safety/Security Risk Inventory (HSRI) technical manual*. New York: Hilson Research, Inc ((Reprinted (2008), Chigaco, IL: IPAT, Inc., a subsidiary of OPP Ltd.) (1995)).

Inwald, R. (2008). The Inwald Personality Inventory (IPI) and Hilson Research inventories: Development and rationale. *Aggression and Violent Behavior 13*, 298-327. doi: 10.1016/j.avb.2008.04.006

Kauffman, C. M., Stolberg, R., & Madero, J. (2015). An examination of the MMPI-2-RF (Restructured Form) with the MMPI-2 and MCMI-III of child custody litigants. *Journal of Child Custody, 12*(2), 129-151. doi: 10.1080/15379418.2015.1057354

- Lowmaster, S. E., & Morey, L. C. (2012). Predicting law enforcement officer job performance with the personality assessment inventory. *Journal of Personality Assessment, 94*, 254-261. doi: 10.1080/00223891.2011.648295
- 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., Bruce, M. N., & Trapnell, P. D. (1995). Effects of self-presentation strategies on personality profiles and their structure. *Personality and Social Psychology Bulletin, 21*, 100-108.
- Paulhus, D. L. (1998). *Paulhus Deception Scales: Manual of the Balanced Inventory of Desirable Responding (BIDR-7)*. Toronto, Canada: Multi-Health Systems.
- Paulhus, D. L., & John, O. P. (1998). Egoistic and moralistic biases in self-perception: The interplay of self-deceptive styles with basic traits and motives. *Journal of Personality, 66*, 1025-1060.
- Rogers., R. (2008). Detection strategies for malingering and defensiveness. In R. Rogers (Ed.), *Clinical assessment of malingering and deception* (3<sup>rd</sup> ed., pp 14-35). New York: Guilford Press.
- Sellbom, M., & Bagby, R. M. (2008). Validity of the MMPI-2-RF (restructured form) L-r and K-r scales in detecting underreporting in clinical and nonclinical samples. *Psychological Assessment, 20*, 370-376. doi: 10.1037/a0012952
- Tarescavage, A. M., Corey, D. M., Gupton, H. M., & Ben-Porath, Y. S. (2015). Criterion validity and practical utility of the Minnesota Multiphasic Personality Inventory-2-Restructured Form(MMPI-2-RF) in assessments of

police officer candidates. *Journal of personality Assessment*, 97(4), 1-13. doi:  
10.1080/00223891.2014.995800

## KULLEN BALTHROP VITA

1. Education and Degrees
  - a. M.Sc. from the University of Edinburgh
  - b. B.A. from Murray State University
2. Professional Positions
  - a. Teaching Assistant at the University of Kentucky
3. Scholastic and Professional Honors
  - a. Summa Cum Laude at Murray State University