INCREMENTAL VALIDITY OF ELEVATED UNDER REPORTING SCALES ON SELECTED MMPI-2-RF SUBSTANTIVE SCALES AND COLLATERAL MEASURES IN A CORRECTIONAL SAMPLE

A THESIS

SUBMITTED TO THE GRADUATE SCHOOL

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE DEGREE

MASTER OF ARTS

BY

KELSEY M. HURM

Dr. JOHNATHAN D. FORBEY – ADVISOR

BALL STATE UNIVERSITY

MUNICE, INDIANA

MAY 2017

Incremental Validity of Elevated Under Reporting Scales on Selected MMPI-2-RF Substantive Scales and Collateral Measures in a Correctional Sample

The Minnesota Multiphasic Personality Inventory family of assessments (e.g. MMPI-2; Butcher, Graham, Ben-Porath, Tellegen, & Dahlstrom, 2003; MMPI-2-RF; Ben-Porath & Tellegen, 2008) are the most widely used and researched standardized psychometric tests of adult personality and psychopathology (Camara, Nathan, & Puente, 2000). Psychologists and other mental health professionals use the MMPI-2 or MMPI-2-RF to help develop treatment regiments, assist with diagnosis, screen potential job candidates during the hiring process or as a component of therapeutic assessment (Butcher & Williams, 2009).

One of the reasons why the MMPI-2 and MMPI-2-RF are the most widely utilized tests of adult personality and psychopathology is because the measures include a number of validity scales that seek to directly measure the validity/interpretability of an individual's response style, referred to as protocol validity (Archer, Buffington-Vollum, Stredny, & Handel , 2006). These scales were developed because failure of a test-taker to respond accurately to the items, whether intentionally or unintentionally, can undermine the results and potentially cause the test results to be uninterpretable. The MMPI-2 and MMPI-2-RF are interpreted in various settings and can aid in determining a number of potential outcomes including potential hiring decisions in employment settings, parole and early release decisions in forensic settings, fitness to stand trial in judicial settings and potential diagnoses and medication in clinical settings. Misinterpretation of the assessment can prove to be costly in various settings (e.g., clinical settings where individuals are prescribed medication or forensic settings where individuals are psychologically assessed for competency to stand trial [Butcher & Williams, 2009]).

The numerous validity scales of the MMPI-2 and MMPI-2-RF are designed to detect biased response styles that an individual may employ when undergoing assessment. Ben-Porath (2012) described two such response styles, non-content based invalid responding (NCIR) and content-based invalid responding (CBIR), which an individual may engage in, either intentionally or unintentionally which can impact the interpretability of tests results. While it is important to note that elevated scales scores on NCIR or CBIR may invalidate a test result, the elevations do not necessarily suggest intentional misrepresentation, as there are many non-deliberate reasons one can elevate these scales (such as reading difficulties, confusion, etc.; [Ben-Porath 2012]). Regardless of intention, NCIR is the first potential biased response style that should be examined when determining protocol validity as it suggests that an individual taking the assessment is not paying attention to item content when answering the questions. Patterns of non-responsiveness can affect scores on substantive scales and validity scales designed to detect CBIR response patterns, thus the rest of the scales should only be interpreted if NCIR is not indicated by the relevant scales.

Once NCIR is ruled out, potential CBIR should be examined (Ben-Porath, 2012). Ben-Porath (2012) described CBIR as a response style where individuals are actively attending to the item content of the test and intentionally or unintentionally responding in a way that is not reflective of their current psychological functioning. Content-based invalid responding can be further divided into two subtypes: over-reporting (CBIR-OR) or under-reporting (CBIR-UR). A CBIR-OR response style consists of endorsing more symptoms, a large number of unusual symptoms, or endorsing symptoms at a higher severity than what the individual is actually experiencing. Individuals are motivated to engage in CBIR-OR in a variety of situations to achieve secondary gain, including disability evaluations, criminal responsibility cases and

psychiatric evaluations (Rogers, 2008). Rogers, Sewell, Martin and Vitacco's (2003) meta-analysis that examined CBIR-OR utilized 65 studies and found that generally CBIR-OR scales are able to detect and accurately classify over-reporters in coached and natural settings. However, the researchers noted that some CBIR-OR scales were more vulnerable to false-positives and were not likely to add incremental validity in the detection of over reporting.

Conversely, a CBIR-UR response style consists of an individual's tendency to present themselves in a light that suggests they are more psychologically healthy than they actually are (Ben-Porath, 2012). Individuals are motivated to engage in CBIR-UR in situations where they can achieve a secondary gain including custody cases, employment screening and other situations where they benefit from presenting themselves as more psychologically well-adjusted than they actually are (Rogers, 2008). Baer and Miller's (2002) meta-analysis that examined CBIR-UR on the MMPI-2 reviewed 22 studies that used coaching simulations or differential prevalence groups (i.e. the comparison of a sample of a specific population that is known to have a higher prevalence of CBIR-UR to a sample of a population with a standard prevalence of CBIR-UR) and found that those who engaged in CBIR-UR scored an average of 1.25 standard deviations below those who were given standard instructions on underreporting validity scales. Baer and Miller (2002) also examined incremental validity in 5 of the 22 studies included in the meta-analysis and found that L and K are reasonably accurate in detecting CBIR-UR.

While both types of CBIR are important to examine empirically, based on the two aforementioned meta-analytic studies (Baer & Miller, 2002; Rogers, Sewell, Martin & Vitacco, 2003), CRIR-UR appears to be a relatively understudied phenomenon in the MMPI literature compared to CBIR-OR as evidenced by the comparative numbers of reviewed studies in each study. This may reflect that fact that CBIR-UR is thought to be a relatively rarer test-taking

approach in various mental health settings (Baer & Miller, 2002; Forbey & Lee 2011; Forbey et al., 2013; McNulty et al., 2003).

While a rarer phenomenon, one of the most common research designs used in assessing the efficacy of CBIR-UR validity scales involves what is referred to in the literature as a "coaching" simulation (e.g., Baer & Sekimlak, 1997; Shores & Carstairs, 1998; Wygant Ben-Porath, Arbisi, Berry, Freeman, & Heilbronner, 2009). In such simulations participants are randomly assigned to one of two groups, one of which is given the standard instructions for taking the MMPI family of tests and the other is given specific instructions to intentionally under-report their psychological symptoms. In other words, the under-reporting group is told to report their psychological functioning in a way that makes them look psychologically "healthy" (e.g., to keep their psychological distress private). The scale scores of the two groups are then compared with the expectation that the participants who were coached or instructed to underreport their psychological symptoms will have higher scores on the underreporting validity scales and lower scores on the substantive scales of the MMPI family of tests.

While "coached" underreporting research has been fairly successful in pointing to the efficacy of the underreporting scales at identifying CBIR-UR, Baer, Wetter and Berry (1992) and Rogers (1997) determined that the ecological validity of coaching studies suffered because there is no evidence that suggests that experimental groups of under-reporters respond in the same way as those who under-report in a natural setting. Along these lines, Baer, Wetter and Berry (1992) suggested that to increase the ecological validity of coaching studies, it is important to give participants in the under-reporting group the necessary and most natural motivation to underreport. Studies since then have given prompts to participants that tell them to imagine one of a number of scenarios that would give them a more natural motivation to under-report. For

example, studies have given prompts telling participants that they are interviewing for a new job, or taking the assessment for a custody case, or offered monetary rewards for under-reporting (Baer, Wetter, Nichols, Greene, & Berry, 1995; Borum, & Stock, 1993).The result of these studies suggest that when participants are coached to under-report their symptoms while provided with sufficient motivation, there are significant elevations on the CBIR-UR scales, suggesting that these scales are effective in determining the difference between a standard profile and an underreporting profile. However, because a CBIR-UR response style in a natural setting is a comparatively rarer phenomenon than over-reporting (Forbey & Lee, 2011; Forbey et al., 2013; McNulty et al., 2003), the research assessing ecological validity of coaching studies has lagged behind, but the existing research suggests that the underreporting validity scales can accurately differentiate between under-reporting profiles and standard profiles (Baer & Miller 2002).

More recent research on the CBIR-UR validity scales (e.g., Forbey & Lee, 2011; Forbey et al., 2013) has stepped away from coached or instructed groups to focus more on CBIR-UR that occurs naturally. In one of the first such studies, McNulty et al. (2003) examined the prevalence of naturally occurring underreporting and over reporting response styles. McNulty and colleagues examined MMPI-2 profiles of 51,486 found that 7.4% of participants produced a CBIR-UR profile, whereas 2.1% of the participants produced a CBIR-OR profile. Unfortunately, compared to previous coaching research that examined the impact of CBIR-UR on MMPI-2 scale scores, the McNulty et al. study was limited to only reporting percentages of participants that engaged in this response style. However, more recent studies of naturally occurring underreporting have mirrored previous coaching research in examining the impact of such a response style on MMPI-2 and MMPI-2-RF scale scores. In addition, these studies have

expanded the former line of research by examining the impact of CBIR-UR on conjointly administered collateral measures.

The first such study conducted by Forbey and Lee (2011) examined the impact of CBIR-UR on mean scores on selected MMPI-2 scales and fourteen external collateral measures selected to reflect the constructs and content of several of the MMPI-2 scales. Specifically, the researchers compared scores of individuals who did not elevate any under-reporting validity scales to those who elevated at least one under-reporting validity scale. In their archival sample of 1,112 college students who were given standard test instructions, roughly 5% of participants naturally engaged in CBIR-UR. Results of mean scale score comparisons suggested that if a college student engaged in a CBIR-UR response style on the MMPI-2 (defined by the elevation of at least one CBIR-UR validity scale), their scores on the selected MMPI-2 substantive scales were significantly lower. Further, individuals who engaged in CBIR-UR also had significantly lower scores on external self-report measures. These findings suggest that if an individual naturally engages in an CBIR-UR response style on the MMPI-2, they will likely approach other self-report measures in the same manner. These results have clinical utility as they suggest individuals whose MMPI-2 profiles indicate a CBIR-UR response style, the individual has likely taken a CBIR-UR biased response approach on all measures they completed thus supplying caution to the interpretability and validity of all measures in the test battery.

Forbey, Lee, Ben-Porath, Arbisi, and Gartland (2013) replicated and extended the findings from Forbey and Lee (2011) with the MMPI-2-RF. Forbey et al. (2013) included data from three different samples: college students, individuals seeking outpatient treatment from a Veterans Affairs facility, and individuals undergoing intake at a correctional facility. The percentages of individuals who engaged in CBIR-UR in the college student sample, the

outpatient psychiatric sample and the correctional participant sample were 5%, 10%, and 27% respectively. The authors found that if individuals engaged in CBIR-UR (defined by at least one CBIR-UR validity scale elevation) on the MMPI-2-RF across the disparate settings they had significantly lower substantive scale scores on the MMPI-2-RF. Further, individuals who engaged in CBIR-UR also carried that that particular response style over to collateral measures taken conjointly or up to one week later as evidenced by significantly lower scores on external collateral measures of internalizing, externalizing and thought disorder.

Another recent line of MMPI-2-RF based CBIR-UR research has explored the incremental validity of the two CBIR-UR validity scales (i.e. Uncommon Virtues [L-r] and Adjustment Validity [K-r]) in non-clinical populations. This research has examined the scores on L-r and K-r in terms of their incremental validity to one another (i.e., do the combination of scales add to the scales individually in predicting CBIR-UR). In a two-part study, Sellbom and Bagby (2008) found L-r and K-r, when utilized together, exhibited utility in differentiating between groups given standard instructions, groups given instructions to underreport and a differential prevalence group of individuals undergoing evaluation for child custody cases. Specifically, they found that groups who were instructed to underreport and the differential prevalence group had significantly lower scores on a majority of the RC scales and higher scores on the underreporting validity scales compared to groups given standard instructions. These findings suggest that L-r and K-r should be utilized in conjunction when attempting to identify underreporting profiles instead of individually. However, one shortcoming of the Sellbom and Bagby (2008) study is that this study only examined the impact of L-r and K-r elevations on the RC scales and did not examine the incremental validity of the underreporting validity scales in terms of investigating the impact of CBIR-UR on other substantive scale scores including the

Higher Order scales of the MMPI-2-RF. Further, Sellbom and Bagby (2008) stated that they did not use a manipulation check to determine whether or not participants in the coaching condition complied with instructions to underreport which could have distorted the predictive validity statistics of L-r and K-r.

Crighton, Marek, Dragon and Ben-Porath (2016) addressed these shortcomings by examining the impact of underreporting scale elevations on substantive scale scores using a coaching simulation with a sample of 302 college students who were given a manipulation check questionnaire to determine whether they were compliant with the instructions they received. They found that those who complied with instructions to underreport produced significantly lower scores on the majority of substantive scales and higher scores on L-r and K-r compared to individuals given standard instructions. Further, they found that when L-r and K-r are used in conjunction with one another to detect underreporting profiles, they possess more predictive power than each scale used alone.

Both Sellbom and Bagby (2008) and Crighton et al. (2016) provided strong evidence for incremental validity of L-r and K-r, but these studies are not without limitations as they both lack a certain degree of ecological validity. Baer and Miller (2002) suggested that research on incremental validity of under-reporting scales should examine naturally occurring under-reporting by analyzing a known groups design instead of differential prevalence group, and still no such study exists. Further, both studies only examined the impact of CBIR-UR on validity and selected substantive scales of the MMPI-2-RF. Examining the incremental validity of the CBIR-UR scales on external measure scale scores along with the MMPI-2-RF substantive scale scores would be a valuable addition to the literature base because the MMPI-2-RF is usually one of several measures given in a clinical setting.

Overall, previous research has suggested that MMPI-2 and MMPI-2-RF CBIR-UR scale scores can detect underreporting response styles (Baer & Miller 2002; Baer & Sekimlak, 1997Baer, et al., 1995; Borum, & Stock, 1993; Ben-Porath, 2012; Shores & Carstairs, 1998; Tellegen & Ben-Porath, 2008/2011; Wygant et al., 2009). In addition, this research has indicated that MMPI-2 and MMPI-2-RF detected underreporting response styles lead to suppression of the substantive scale scores of these measures (Ben-Porath, 2012; Sellbom & Bagby 2008; Tellegen & Ben-Porath, 2008/2011) as well as scores on collateral measures (Forbey & Lee 2011; Forbey et al., 2013). Separately, research has suggested that the combination of L-r and K-r add significantly to one another in the detection of CBIR-UR and should be used in combination when attempting to determine protocol validity (Crighton et al., 2016; Sellbom & Bagby 2008). However, to date no study has examined the combination of these two disparate lines of research to examine incremental validity of CBIR-UR scales and its relative impact on substantive scales of the MMPI-2-RF and conjointly administered collateral measures. The current study sought to strengthen and extend research on the MMPI-2-RF CBIR-UR validity scales by connecting naturally occurring underreporting (Forbey & Lee 2011; Forbey et al., 2013) and incremental validity (Sellbom & Bagby 2011; Crighton et al., 2016). Specifically, the aim of the current study is examine the relation between underreporting validity scale elevations, substantive scale scores on the MMPI-2-RF and collateral measure scores in an archival dataset of MMPI-2-RF profiles from correctional inmates undergoing intake assessment (who have been repeatedly demonstrated to more frequently elevate CBIR-UR scales [McNulty et al, 2003; Forbey et al 2013)]. Further, the impact of both individual and multiple CBIR-UR scale elevations will be examined on both MMPI-2-RF scale scores as well as in scores of conjointly administered measures of psychological functioning.

Based on the results of previous research (Forbey & Lee, 2011; Forbey et al., 2013), it was hypothesized that if an underreporting response style was suggested by either of the MMPI-2-RF validity scales (i.e., L-r and K-r), the response style would be evident on selected MMPI-2-RF substantive scales as well as conjointly administered collateral self-report measures. Further, based on the results of Sellbom and Bagby (2008), it is hypothesized that individuals who have differential numbers of elevated validity scale scores (i.e., 0, 1 or 2) would have significantly different mean scores from one another across the majority of measures and MMPI-2-RF substantive scales and criterion measure scores. In other words, it was thought that individuals who had one validity scale elevated would have significantly lower scores than individuals with no validity scores elevated and that individuals with two validity scales elevated would have significantly lower scores than those individuals who had one or no validity scale elevations.

Method

Participants

The archival data utilized in the current study included a total of 632 male inmates from a large Midwestern U.S. intake correctional facility who volunteered to participate in a larger, ongoing study of computerized adaptive testing with the MMPI-2 (Forbey & Ben-Porath, 2007). The age of the participants ranged from 18 to 66 with a mean (SD) of 32.27 (9.29) years. In terms of ethnicity, 317 (50.2%) of the participants identified as Caucasian, 184 (29.1%) identified as African American and 131 (20.7%) identified as other/not reported.

Research suggests that elevations on NCBIR scales (i.e. VRIN-R, TRIN-R and CNS) can artificially elevate CBIR scales (Burchett et al., 2015). Therefore, in order to reduce error variance in the analyses, individuals who provided content nonresponsive MMPI-2-RF profiles, defined by the omission of responses on 15 items or more (Cannot Say [CNS/?]) and/or T scores

of 80 or greater on Variable Response Inconsistency-Revised (VRIN-R) or True Response Inconsistency-Revised (TRIN-R) were removed from the analysis. Further, individuals who engaged in CBIR-OR, defined by T scores of 100 or greater on the Infrequent Responses (F-r), Symptom Validity (FBS-r), or Response Bias (RBS) Scales, and/or 80 or more on the Infrequent Psychopathology Responses (F_p-r) or Infrequent Somatic Responses (Fs) scales, were also removed because the current study focused only on CBIR-UR and Within-Normal-Limits (WNL) profiles. Leaving in the CBIR-OR profiles would lead the comparison group to appear more psychologically disturbed than they actually are which would exaggerate any results found. All exclusionary criteria listed above are based upon the recommendations in the MMPI-2-RF manual (Ben-Porath & Tellegen, 2008/2011). This procedure excluded 58 participants (9.1%) from the analysis, leaving 569 participants remaining in the analysis.

After the NCIR and CBIR-OR exclusions, the final sample included 569 male correctional participants (age range: 18-66 years, M = 32.58 years, SD = 9.93 years). In terms of ethnicity, 286 (50.3%) of the participants the final sample identified as white, 162 (28.5%) of the participants identified as African American, and 121 (21.3%) of participants identified as other/not reported. No significant differences between included and excluded participants were indicated in terms of age and/or ethnicity.

Measures

Minnesota Multiphasic Personality Inventory–2–Restructured Form (Ben-Porath & Tellegen, 2008; Tellegen & Ben-Porath, 2008/2011)**.** Revised from the MMPI-2 (Butcher et al., 2001), the MMPI-2-RF is a 338-item, true/false, self-report inventory that assesses an individual's psychological functioning in a number of domains (i.e., personality, psychopathology, and social/ behavioral functioning). The MMPI-2-RF contains 9 validity scales

designed to assess the test taker's response style and 42 substantive scales assessing behavioral, emotional, social, and thought dysfunction. The MMPI-2-RF Technical Manual (Tellegen & Ben-Porath, 2008/2011) provides extensive evidence supporting reliability and validity for scales on this instrument (Tellegen & Ben-Porath, 2008/2011). For the current study, the MMPI-2-RF was rescored from a conventional administration of the MMPI-2. Tellegen and Ben-Porath (2008/2011) and Van der Heijden, Egger, and Derksen (2010) have reported that MMPI-2-RF scale scores generated from an MMPI-2 administration are interchangeable with those generated from the MMPI-2-RF booklet. The current study analyzed 12 of the 42 substantive scales. Specifically, Higher Order (HO) and Restructured Clinical (RC) scales were used because they measure constructs conceptually relevant to the eight collateral measures that were administered to participants. Scale descriptions and internal consistency statistics for this sample can be found in Table 1.

Collateral Measures

Eight collateral self-report measures, originally utilized for the purpose of examining the comparative validity of computer adaptive versions of the MMPI-2 to conventional administrations, were selected because they conceptually relate to constructs measured by the MMPI-2 and by extension, the MMPI-2-RF.The eight self-report collateral measures are listed below and were rationally organized into internalizing, externalizing and thought disorder categories to highlight the similar constructs measured by the collateral measures and the MMPI-2-RF Higher—Order scales.

Internalizing

Beck Depression Inventory (BDI; Beck, Ward, Mendelson, Mock, & Erbaugh, 1961). The BDI is a 21-item self-report inventory designed to measure depressive symptomatology.

Specifically, it is designed to measure characteristic attitudes and symptoms of depression. Items are rated on a 4-point Likert-type scale, with higher ratings generally indicating higher levels of psychological distress. The BDI is well validated in clinical and non-clinical populations and reliable with an internal consistency (α) of .86 within psychiatric participants and an internal consistency (α) of .81within nonclinical samples (Beck, Steer & Carbin, 1988). In the current study, estimated internal consistency (α) for the BDI was .93.

Screener for Somatoform Disorders (SSD; Janca et al., 1995). The SSD is a 12-item self-report inventory designed to measure diffuse somatic complaints related to somatoform disorders as defined by the International Classification of Diseases (ICD-10; World Health Organization, 1992) and the fourth edition of the American Psychiatric Association's (APA) Diagnostic and Statistical Manual of Mental Disorders (DSM-IV; APA, 1994). Scores on the SSD demonstrate good test-retest reliability (kappa = .76). Items on the SSD are based upon diagnostic criteria which suggests good content validity (Janca et al., 1995). In the current study, the internal consistency (α) was .84.

Externalizing

Barratt Impulsivity Scale–Version 10 (BIS; *Barratt, 1985*) The BIS is a 34-item self-report inventory of impulsivity which provides a total score, as well as scores on three dimensions of impulsivity: Non-planning (12 items), Motor (11 items), and Cognitive (11 items). Item responses are rated on a 4-point Likert-type scale ranging from 1(rarely/never) to 4 (almost always/always). Scores on the BIS are highly correlated with similar scores on self-report measures including the I7 Impulsiveness Questionnaire (Luengo, Carrillo-De-La-Pena, & Otero, 1991). The BIS-10 demonstrates relatively high internal consistency (α = .79) and test-retest

reliability (Spearman's ρ = .83; Barratt, 1985; Patton, Stanford & Barratt 1995). In the current study, only the total scale score was used, which had an estimated internal consistency (α) of .87.

Drug Abuse Screening Test (DAST; Skinner, 1982). The DAST is a 20-item self-report measure designed to measure an individual's self-reported use of prescription, over-the-counter, and other illicit drugs. The items are presented with dichotomous answer choices (yes or no). The DAST has been well validated as it demonstrates adequate discriminant validity and specificity in a clinical sample when using DSM drug abuse and dependence diagnoses (Gavin, Ross & Skinner, 2006). The DAST demonstrates a strong internal consistency (α) of .92 (Skinner, 1982). In the current study, the estimated internal consistency (α) was.93.

Michigan Alcohol Screening Test (MAST; Selzer, 1971). The MAST is a 24-item self-report measure designed to measure an individual's level of problematic alcohol use. The items are presented with dichotomous answer choices (yes or no). The MAST demonstrates adequate convergent validity as it correlates well with other measures of self-reported alcohol use, though the measure is prone to false positive indications of alcohol abuse (Gibbs, 1983). The MAST demonstrates a strong internal consistency in nonclinical populations (α = .84; Storgaard, Nielsen, & Gluud, 1994). In the current study, the internal consistency (α) was .80.

Trait Anger subscale of State-Trait Personality Inventory (STPI-TA; Spielberger, 1979). The STPI is an 80-item self-report measure designed to assess anxiety and anger. Each item is rated on a four-point Likert-type scale (1 = almost never, 4 = almost always). The measure is divided into 8 subscales comprised of 10-items each. The current dataset relied on a modified version of the assessment containing only 10 items from the trait anger subscale. Scores on the Trait Anger subscale correlated highly with other measures of trait hostility including the Buss-Durkee Hostility Inventory and Hostility Scale (Spielberger, Jacobs, Russell, & Crane, 1983).

Spielberger and Reheiser (2009) reported a strong internal consistency (α = .87) for the scores on trait anger subscale. In the current study, the estimated internal consistency (α) was .88.

Thought Disorder

Magical Ideation Scale (MIS; Eckbald & Chapman, 1983). The MIS is a 30-item self-report inventory designed to measure beliefs about unconventional causal relations between events which are commonly associated with thought disorders. Items are presented in a dichotomous manner with choices of true or false. The MIS has been well validated in nonclinical populations and demonstrates a strong internal consistency (α = .80; Eckald & Chapman, 1983). In the current study, the estimated internal consistency (α) was .81.

*Perceptual Aberration Scale (PAS; Chapman, Chapman & Raulin, 1978).*The PAS is a 35-item self-report inventory that examines physical and other perceptual distortions related to thought disorders. The items are presented with dichotomous answer choices (true or false). The PAS displays very high internal consistency for actively schizophrenic patient samples (α = .90) as well as a nonclinical samples (α = .88). Items on the PAS were constructed using abnormal experiences and perceptions that frequently appeared in the clinical literature for schizophrenics, suggesting content validity (Chapman, Chapman & Raulin, 1978). In the current study, internal consistency (α) was estimated to be .97.

Procedure

All participants completed a standard paper and pencil version of the MMPI-2 (later rescored as an MMPI-2-RF) upon intake to the correctional facility as part of standard institutional screening procedures. Though the MMPI-2 was administered at intake for all inmates, data was only used in analysis if the inmates agreed to volunteer their results to the study. Eight self-report based collateral measures with conceptually relevant personality and

psychopathology constructs were administered in a group format typically between 1 and 5 days after administration of the intake MMPI-2. All eight collateral measures were administered via paper and pencil. Data were then entered into a computer, deidentified and archived for analysis.

The NCIR and CBIR-OR profiles were first removed from the archival data then grouped into one of three different groups based upon how many underreporting validity scales they elevated. This resulted in three groups, 1) within normal limits (i.e., no CBIR-UR scales elevated), 2) one CBIR-UR scale elevated and 3) two CBIR-UR scales elevated. Elevations were defined using the criteria from the Manual for Administration, Scoring and Interpretation (Ben-Porath & Tellegen, 2008/2011) where the recommended cutoff T scores for L-r is 70 or greater and for K-r is 66 and greater. The "within normal limits" group (WNL) consisted of 406 participants. The second group consisted of 134 participants who elevated either L-r or K-r. The third group consisted of 29 participants who elevated both L-r and K-r. Collateral measure scores were converted to Z-scores based upon the sample statistics in order to facilitate the interpretation of scale score differences.

Results

A series of non-parametric one-way ANOVAs were used to analyze mean differences between the three groups on scores from the selected MMPI-2-RF substantive scales and eight collateral measures. Due to the violation of the assumption of homogeneity of variances in an overwhelming majority of the ANOVA analyses, Welch's corrections were used. All significant omnibus tests were followed up by pairwise comparisons between each of the groups on all scales and collateral measures. Cohen's d effect sizes with 0.2, 0.5 and 0.8 reflecting small, medium and large effects, respectively, for the differences between means in each group were reported for all post-hoc comparisons regardless of significance.

The first set of analyses explored H-O and RC scale score differences between the three underreporting groups. Table 2 contains the results of these analyses. To reduce Type I error, a Bonferroni correction was applied making the critical alpha for these analyses 0.004 (0.05/12). Results of the 12 ANOVA analyses comparing substantive scale scores between the underreporting scale elevation groups indicated that all omnibus tests were significant ($p < .001$). Post-hoc analyses revealed a significant downward stepwise pattern between the elevation groups on a majority of the selected MMPI-2-RF substantive scales. Specifically, the WNL group had the highest mean scale scores, the one elevated CBIR-UR scale group had significantly lower mean scale scores than the WNL group, and the two elevated CBIR-UR scales group had significantly lower scores than both the WNL and one elevated CBIR-UR scale groups. This pattern emerged on all selected MMPI-2-RF substantive scale except for RC1 and RC2. While mean scale scores on RC1 and RC2 followed the hypothesized pattern, the comparison between two validity scale elevations and WNL was the only significant result. Effect sizes (Cohen's d) for all MMPI-2-RF comparisons can be found in Table 4. Effect sizes ranged from 0.34 – 1.86 with the majority (i.e., 32 out of 36) of comparisons between groups yielding a medium to large effect size.

The second series of non-parametric ANOVA analyses explored the mean differences on the external collateral measures between the three CBIR-UR validity scale elevation groups. Table 3 contains the results of these analyses. To reduce Type I error, a Bonferroni correction was applied making the critical alpha value for these analyses 0.006 (0.05/8). Results of the eight non-parametric ANOVA analyses between the three groups indicated that all omnibus tests were significant ($p < .001$). Further post-hoc analyses revealed a significant downward step-wise pattern for one internalizing measure (i.e., the BDI). Specifically for BDI, the WNL group had

the highest mean Z-scores, the one elevated CBIR-UR scale group had significantly lower mean Z-scores than the WNL group ($p < .001$), and the two elevated CBIR-UR scales group had significantly lower scores than both the one elevation and WNL groups ($p < .010$). Post-hoc analyses for one of the internalizing measures (SSD), and three of the externalizing measures (BIS, MAST and STPI-TA), revealed that individuals who elevated one or two validity scales had significantly lower mean Z-scores compared to the WNL group ($p < .001$), but those who elevated two CBIR-UR validity scales did not have significantly lower mean Z-scores than those who elevated only one CBIR-UR validity scale. Lastly, post-hoc analyses for one of the externalizing measures (DAST) and two thought disorder measures (MIS, PAS) revealed that individuals who elevated two underreporting validity scales had significantly lower mean Z-scores than those who elevated one validity scale and those who did not elevate any ($p < .001$). However for these three criterion measures the one elevation group did not have significantly lower mean Z-scores than the WNL group. Effect sizes (Cohen's d) for all group comparisons on collateral measures are provided in Table 5. Effect sizes ranged from 0.05-1.06 with half (i.e., 12 out of 24) of the comparisons yielding a medium to large effect size.

Deleted:

Discussion

The current study examined the incremental validity of the under-reporting scales of the MMPI-2-RF (i.e., L-r and K-r) in terms of how they impact scores on selected MMPI-2-RF substantive scales and conjointly administered collateral measures in a sample of males undergoing intake at a correctional facility. It was hypothesized that if an underreporting response style was suggested by the MMPI-2-RF validity scales L-r and/or K-r, the response style would be evident on the scores of selected MMPI-2-RF substantive scales as well as carry over to the scores on eight conjointly administered collateral self-report measures. This

hypothesis was supported as results of the current study indicate that if a CBIR-UR style was suggested by the elevation of L-r and/or K-r, selected mean MMPI-2-RF scale scores and mean collateral measure Z-scores were significantly lower compared to those who had no validity scale elevations. This suggests that individuals who engaged in a CBI-UR response style on the MMPI-2-RF also approached the external collateral measures in a similar manner. More importantly, it was further hypothesized that individuals who had differential numbers of elevated CBIR-UR scales (i.e., 0, 1 or 2) would have significantly different mean scores from one another across the MMPI-2-RF substantive scale and criterion measure scores. This hypothesis was generally supported by the results for the MMPI-2-RF scales as this pattern emerged in ten out of twelve selected MMPI-2-RF scales. Although the hypothesized pattern emerged in only one of eight collateral measures, in all eight collateral measures those who elevated two validity scales had significantly lower scores than those who were within normal limits.

In terms of the first hypothesis, the results of the analyses replicated and expanded previous research (e.g., Forbey & Lee, 2011, Forbey et al., 2013), as individuals who naturally engaged in a CBIR-UR response style as evidenced by either one and/or two CBIR-UR scale elevations on the MMPI-2-RF not only had lower scale scores on all twelve of the selected MMPI-2-RF substantive scales but also on all eight external collateral measure scores (depending on group membership). This study expanded upon Forbey et al. (2013) and Forbey and Lee (2011) by examining the impact of a varying number of MMPI-2-RF underreporting validity scales elevations on MMPI-2-RF substantive scales and collateral measures. Specifically, instead of dividing the sample into two groups based upon at least one validity scale elevation, the current study further divided the CBIR-UR group to examine the potential impact

of one and/or two validity scale elevations on scale scores and collateral measure scores (i.e., incremental validity).

Regarding the second set of second set of hypotheses examining potential incremental validity of the CBIR-UR scales, results were mixed across the MMPI-2-RF and collateral measure scores. First, for the MMPI-2-RF substantive scale scores, results suggested that if both CBIR-UR scales were elevated, all three H-O scales and seven of nine RC scales displayed an incremental validity pattern (i.e., with the exception of RC1 and RC2). This finding supports and expands upon previous research on the impact of naturally occurring CBIR-UR (Forbey & Lee, 2011: Forbey et al., 2013) by incorporating CBIR-UR predictive/ incremental validity (Crighton et al., 2016; Sellbom & Bagby, 2008) research. Specifically, previous naturally occurring research found that in some cases having one or two scales wasn‘t enough for a significant difference in mean scores from those who did not have any validity scales elevated (Forbey & Lee, 2011; Forbey et al., 2013). In particular, those who elevated at least one under-reporting validity scale did not have significantly different mean scale scores on RC1 and RC2 compared to those who did not elevate any under-reporting scales (Forbey & Lee, 2011; Forbey et al., 2013). This finding is supported in multiple populations including college students and correctional participants. In comparison, Sellbom and Bagby (2008) found in multiple populations including college students and individuals undergoing custody hearings that L-r and K-r possess greater predictive power to differentiate CBIR-UR profiles from standard profiles when they are both elevated. Further, Crighton et al., (2016) replicated Sellbom and Bagby (2008) and found stronger evidence for incremental validity of L-r and K-r, however only in a sample of college students that were either instructed to under-report or were given standard instructions.

A majority of the second set of non-parametric ANOVA analyses comparing mean Z – scores between each of the three CBIR-UR groups did not display the hypothesized stepwise pattern of results in terms of incremental validity (i.e. two validity scale elevations would have lower mean collateral measure scale scores compared to one validity scale elevation; one validity scale elevation would have lower mean collateral measure scale scores compared to zero validity scale elevations). The only external collateral measure to display the hypothesized pattern (i.e. 0, 1 and 2 scale differences in mean scores) was the BDI. In a similar sample of correctional participants, Forbey et al. (2013) found that participants who elevated at least one underreporting validity scale had significantly lower mean Z-scores on all collateral measures except for the thought disorder measures. The results of the current study suggested that the lack of a significant difference between underreporting and comparison groups on collateral thought disorder measures in Forbey et al. (2013) is likely because only individuals who elevated both of the validity scales had significantly lower mean Z-scores than the comparison group. Because Forbey et al. (2013) defined underreporting as an elevation on at least one underreporting scale, the definition of underreporting was too broad to fully capture this specific pattern. Further, Forbey and Lee (2011) only found significant differences in 15 of 35 collateral measures between CBIR-UR and a comparison group. The authors noted significant power loss due to a small percentage of naturally occurring underreporting. While the current study also lacks power, the results of Forbey and Lee (2011), Forbey et al (2013) and the current study suggest that if an individual elevates both underreporting validity scales, they will likely have lower scale scores on MMPI-2-RF substantive scales and will approach all other measures in the battery in the same manner. While mean Z-scores followed the general pattern of decreasing as the number of underreporting validity scale elevations increased, the majority of the mean Z-score post-hoc

differences did not fit the hypothesis. However, the results do expand upon the incremental validity research (Crighton et al., 2016; Sellbom & Bagby 2011) because in all eight collateral measures two validity scale elevations produced significantly lower mean collateral measure scores compared to zero validity scale elevations. While one validity scale elevation did produce lower mean collateral measure scores, in many cases the differences were not significant. This finding suggests that L-r and K-r possess more predictive validity when both are elevated to determine CBIR-UR.

While the hypothesized pattern was not fully supported when examining the impact of varying number of validity scale elevations on mean collateral measure Z –scores, this study produced some interesting and unexpected findings. Only one of the collateral measure analyses produced the hypothesized downward stepwise pattern where one elevation produced significantly lower mean scores than the comparison group, and two elevations produced significantly lower mean scores compared to one elevation. However, in a majority of the analyses, elevations of both L-r and K-r led to lower mean MMPI-2-RF scale scores and mean collateral measure Z-scores than one elevation and the comparison group. Further, in all analyses two elevations produced significantly lower scores than the comparison group which suggests that L-r and K-r act in an additive manner working together to detect an underreporting response style. L-r and K-r likely measure different constructs of underreporting as evidenced by situations where one scale is elevated and the other is not. However decreased mean scale and collateral measure scores occur when both scales are elevated suggesting that when L-r and K-r are used together, the MMPI-2-RF contains a higher predictive ability to detect underreporting response styles. This claim supports previous research on the incremental validity of L-r and K-r

stating that these scales should be interpreted in conjunction with one another to accurately determine an underreporting response style (Crighton et al., 2016; Sellbom & Bagby 2008).

This study has many strengths including the use of naturally occurring invalid responding rather than simulated invalid responding, the use of underreporting validity scales in conjunction with one another, and the replication of previous research (e.g. Forbey & Lee, 2011; Forbey et al., 2013). However, this study is not without limitations. First, the statistical analyses likely lacked power. Only about 27% of the sample engaged in CBIR-UR which led to large differences in cell *n*s between each of the three groups. The large group differences led to a violation of homogeneity of variances which was corrected by using the Welch statistic. Due to the number of ANOVAs performed, a Bonferroni correction was applied to correct for Type I Error Rate, but in doing so, the probability of Type II Error increased slightly. Because effect size values are independent of sample size, it is important to consider the effect sizes of the comparisons in conjunction with significance values. Though some analyses were not statistically significant, the comparisons still yielded a medium to large effect size. For example, the post-hoc comparison between 1 and 2 elevations on RC1 was not statistically significant ($p=.11$) but yielded a medium effect size ($d = 0.54$). Medium to large effect sizes for the large majority of the comparisons suggest that the mean differences in scale scores and collateral measure scores have a degree of practical significance (i.e., clinical utility). This study, along with previous research provide strong evidence that if an individual engages in an underreporting response style on the MMPI-2-RF, they will likely approach all other self-report measures with the same response style, which means that the validity of the entire test battery should be in question (Forbey & Lee, 2011; Forbey et al., 2013). Further, the current study supports the interpretation of both L-r and K-r when determining underreporting response styles, which is

recommended in the *Manual for Administration, Scoring and Interpretation* (Ben-Porath & Tellegen 2008/2011) and supported by other research (Sellbom & Bagby 2008; Crighton et al., 2016).

Regardless of the limitations of this study, the findings suggest that as individuals elevate an increasing number of CBIR-UR validity scales, they tend to have significantly lower scores on MMPI-2-RF substantive scales. Generally the same pattern exists in the external collateral measures. However, in both MMPI-2-RF substantive scales and collateral measures, the two elevation group always had significantly lower mean scores than the comparison group. This pattern suggests that when participants who engage in CBIR-UR on the MMPI-2-RF as defined by L-r and K-r elevations, they will likely approach external measures similarly. Also, when the CBIR-UR scales are used in conjunction with one another, an underreporting response style can be consistently detected evidenced by significantly lower scale and collateral measure scores. However, the results indicate that one and two elevation groups differ significantly from the WNL group for half of the collateral measures. This study also expands upon previous research by Sellbom and Bagby (2008) and Crighton et al. (2016) by adding ecological validity to incremental validity studies of the CBIR-UR scales. This study demonstrates the additive properties of the underreporting validity scales by showing the general pattern of increasingly lower mean scale scores as the number of under-reporting validity scale elevations increase. In addition, it appears to be the first study conducted that specifically examines the impact of a varying number of underreporting validity scales on both MMPI-2-RF substantive scales and conjointly administered collateral measures. Future research should build off of these conclusions and include a larger sample size with smaller group differences from multiple different populations with higher prevalence of CBIR-UR to further explore its impact on

MMPI-2-RF substantive scales and external collateral measures. Future research should also examine the potential incremental validity of CBIR-OR scales in a similar fashion to determine how scale scores and collateral measure scores fluctuate with increasing numbers of elevated CBIR-OR validity scales. Currently this line of research has only examined the impact of validity scale elevations on scores of self-report measures. Examining the impact of MMPI-detected CBIR on scores of behavioral and cognitive measures would add stronger clinical implications for this line of research.

References

American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental health disorders* (4th ed.). Washington DC: Author.

Archer, R. P., Buffington-Vollum, J. K., Stredny, R. V., & Handel, R. W. (2006). A survey of psychological test use patterns among forensic psychologists. *Journal of personality assessment*, *87*(1), 84-94.

Baer, R. A., & Miller, J. (2002). Underreporting of psychopathology on the MMPI–2: A meta-analytic review. *Psychological Assessment*, *14*, 16–26.

Baer, R. A., & Sekirnjak, G. (1997). Detection of underreporting on the MMPI-2 in a clinical population: Effects of information about validity scales.*Journal of personality assessment*, *69*(3), 555-567.

Baer, R. A., Wetter, M. W., & Berry, D. T. (1992). Detection of underreporting of psychopathology on the MMPI: A meta-analysis. *Clinical Psychology Review*, *12*, 509-525.

Baer, R. A., Wetter, M. W., Nichols, D. S., Greene, R., & Berry, D. T. (1995). Sensitivity of MMPI-2 validity scales to underreporting of symptoms. *Psychological Assessment*, *7*, 419.

Bagby, R. M., Rogers, R., & Buis, T. (1994). Detecting malingered and defensive responding on the MMPI-2 in a forensic inpatient sample. *Journal of Personality Assessment*, *62*, 191-203.

Barratt, E. S. (1985). Impulsiveness subtraits: Arousal and information processing. *Motivation, emotion, and personality* (Vol. 5, pp. 137-146). New York, NY: North-Holland.

Beck, A. T., Steer, R. A., & Carbin, M. G. (1988). Psychometric properties of the Beck Depression Inventory: Twenty-five years of evaluation. *Clinical Psychology Review*, *8*, 77-100.

Beck, A. T., Ward, C., Mendelson, M., Mock, J., & Erbaugh, J. (1961). Depth of depression. *Archives of General Psychiatry*, *4*, 561-571.

Ben-Porath, Y. S., & Tellegen, A. (2008/2011). *MMPI-2-RF: Manual for administration, scoring and interpretation*. Minneapolis, MN: University of Minnesota Press.

Borum, R., & Stock, H. V. (1993). Detection of deception in law enforcement applicants. *Law and Human Behavior*, *17*, 157–166.

Burchett, D., Dragon, W. R., Smith Holbert, A. M., Tarescavage, A. M., Mattson, C. A., Handel, R. W., & Ben-Porath, Y. S. (2016). "false feigners": Examining the impact of non-content-based invalid responding on the 27innesota multiphasic personality inventory-2 restructured form content-based invalid responding indicators. *Psychological Assessment, 28*(5), 458-470

Butcher, J. N. (1999). *A beginner's guide to the MMPI-2*. American Psychological Association.

Butcher, J. N., Graham, J. R., Ben-Porath, Y. S., Tellegen, A., Dahlstrom, W. G., & Kaemmer, B. (2001). *MMPI–2 (Minnesota Multiphasic Personality Inventory–2): Manual for administration, scoring, and interpretation* (Rev. ed.). Minneapolis: University of Minnesota Press.

Butcher, J. N., & Han, K. (1995). Development of an MMPI-2 scale to assess the presentation of self in a superlative manner: The S Scale. In J. N. Butcher, C. D. Spielberger, J. N. Butcher, C. D. Spielberger (Eds.) *Advances in personality assessment, Vol. 10* (pp. 25-50). Hillsdale, NJ, England: Lawrence Erlbaum Associates, Inc.

Butcher, J. N., & Williams, C. L. (2009). "Personality assessment with the MMPI-2: Historical roots, international adaptations, and current challenges". *Applied Psychology: Health and Well-Being. 1* (1): 105–135.

Camara, W. J., Nathan, J. S., & Puente, A. E. (2000) Psychological test usage: Implications in professional psychology. *Professional Psychology: Research & Practice*, *31*, 141–154.

Chapman, L. J., Chapman, J. P., & Raulin, M. L. (1978). Body-image aberration in schizophrenia. *Journal of Abnormal Psychology*, *87*, 399-407.

Crighton, A. H., Marek, R. J., Dragon, W. R., & Ben-Porath, Y. S. (2016). Utility of the MMPI-2-RF Validity Scales in Detection of Simulated Underreporting Implications of Incorporating a Manipulation Check. *Assessment*, 1073191115627011.

Eckblad, M. L., & Chapman, L. J. (1983). Magical ideation as an indicator of schizotypy. *Journal of Consulting and Clinical and Psychology*, 51, 215-225.

Forbey, J. D., & Ben-Porath, Y. S. (2007). Computerized adaptive personality testing: A review and illustration with the MMPI-2 Computerized Adaptive Version. *Psychological Assessment*, *19*, 14-33.

Forbey, J.D., & Lee, T.T.C. (2011). An exploration of the impact of invalid MMPI-2 protocols on collateral self-report measures. *Journal of Personality Assessment*, *93*, 556-565.

Forbey, J.D., Lee, T.T.C, Ben-Porath, Y.S., Arbisi, P.A., & Gartland, D. (2013). Associations between MMPI-2-RF validity scale scores and extra-test measures of personality and psychopathology. *Assessment, 20*, 448-461.

Gavin, D. R., Ross, H. E., & Skinner, H. A. (1989). Diagnostic validity of the drug abuse screening test in the assessment of DSM-III drug disorders. *British journal of addiction*, *84(3)*, 301-307.

Gibbs, L. E. (1983). Validity and reliability of the Michigan Alcoholism Screening Test: A review. *Drug and alcohol dependence*, *12*(3), 279-285.

Graham, J. R. (1990). *MMPI-2: Assessing personality and psychopathology*. New York: Oxford University Press.

Janca, A., Burke, J. D., Isaac, M., Burke, K. C., Costa e Silva, J. A., Acuda, S. W., & Tacchini, G. (1995). The World Health Organization Somatoform Disorders Schedule: A preliminary report on design and reliability. *European Psychiatry*, *10*, 373-378.

Luengo, M. A., Carrillo-De-La-Pena, M. T., & Otero, J. M. (1991). The components of impulsiveness: a comparison of the I. 7 Impulsiveness Questionnaire and the Barratt Impulsiveness Scale. *Personality and Individual Differences*, *12*(7), 657-667.

McNulty, J. L., Forbey, J. D., Graham, J. R., Ben-Porath, Y. S., Black, M. S., Anderson, S. V., & Burlew, A. K. (2003). MMPI-2 validity scale characteristics in a correctional sample. *Assessment*, *10*, 288-298.

Meehl, P. E., & Hathaway, S. R. (1946). The K factor as a suppressor variable in the Minnesota Multiphasic Personality Inventory. *Journal of Applied Psychology*, *30*, 525.

Nichols, D., Greene, R., & Schmolck, P. (1989). Criteria for assessing inconsistent patterns of item endorsement on the MMPI: Rationale, development, and empirical trials. *Journal of Clinical Psychology*, *45*, 239–250.

Patton, J. H., Stanford, M. S., & Barratt, E. S. (1995). Factor structure of the Barratt impulsiveness scale. *Journal of Clinical Psychology*, *5*, 768-774.

Radloff, L. S. (1977). The CES-D scale: A self-report depression scale for research in the general population. *Applied Psychological Measurement, 1*, 385-401.

Rogers, R. (1997). Researching dissimulation. In R. Rogers (Ed.), *Clinical assessment of malingering and deception* (2nd ed., pp. 309–327). New York: Guilford Press

Rogers, R. (2008). *Clinical assessment of malingering and deception* (3rd ed.). New York, NY: Guilford.

Rogers, R., Sewell, K. W., Martin, M. A., & Vitacco, M. J. (2003). Detection of Feigned Mental Disorders A Meta-Analysis of the MMPI-2 and Malingering. *Assessment, 10*, 160-177.

Sellbom, M., & Bagby, R. M. (2008). Validity of the MMPI-2-RF (restructured form) Lr and Kr scales in detecting underreporting in clinical and nonclinical samples. *Psychological Assessment, 20*, 370-395.

Selzer, M. (1971). The Michigan Alcoholism Screening Test: The quest for a new diagnostic instrument. *American Journal of Psychiatry, 127*, 1653-1658

Shores, A. E., & Carstairs, J. R. (1998). Accuracy of the MMPI-2 computerized Minnesota Report in identifying fake-good and fake-bad response sets. *The Clinical Neuropsychologist, 12*(1), 101-106.

Skinner, H. A. (1982). The Drug Abuse Screening Test. *Addictive Behaviors, 7*, 363-371.

Spielberger, C. D. (1979). P*reliminary manual for the State-Trait Personality Inventory (STPI)*. Tampa: University of South Florida.

Spielberger, C. D., & Reheiser, E. C., (2009). Assessment of emotions: Anxiety, anger, depression, and curiosity. *Applied Psychology: Health and Well-being, 1*, 271-302.

Spielberger, C. D., Jacobs, G., Russell, S., & Crane, R. S. (1983). Assessment of anger: The state-trait anger scale. *Advances in personality assessment, 2*, 159-187.

Storgaard, H., Nielsen, S. D., & Gluud, C. (1994). The validity of the Michigan alcoholism screening test (MAST). *Alcohol and Alcoholism,29*, 493-503

Tellegen, A., & Ben-Porath, Y. S. (2008/2011). *MMPI-2-RF (Minnesota Multiphasic Personality Inventory-2 Restructured Form): Technical manual.* Minneapolis: University of Minnesota Press.

Tellegen, A., Ben-Porath, Y. S., McNulty J. L., Arbisi, P. A., Graham, J. R., & Kaemmer, B. (2003). *MMPI-2 Restructured Clinical (RC) Scales: Development, validation and interpretation*. Minneapolis: University of Minnesota Press.

Van der Heijden, P. T., Egger, J. I. M., & Derksen, J. J. L. (2010). Comparability of scores on the MMPI–2–RF scales generated with the MMPI–2 and MMPI–2–RF booklets. *Journal of Personality Assessment*, 92(3), 254-259.

World Health Organization. (1992). *The ICD-10 international classification of mental and behavioral disorders. Clinical descriptions and diagnostic guidelines*. Geneva, Switzerland: Author.

Wygant, D. B., Ben-Porath, Y. S., Arbisi, P. A., Berry, D. T., Freeman, D. B., & Heilbronner, R. L. (2009). Examination of the MMPI-2 Restructured Form (MMPI-2-RF) validity scales in civil forensic settings: Findings from simulation and known group samples. *Archives of Clinical Neuropsychology, 24*, 671-680.

Table 1

Higher-Order and Restructured Clinical Scale Descriptions and internal consistency

Scale and Description	α
Higher-Order (H-O) Scales	
Emotional/Internalizing Dysfunction (EID)	.92
41-item scale designed to assess a range of emotional and internal problems. Low scores indicate a below-average level of emotional difficulties.	
Thought Dysfunction (THD)	.81
26-item scale designed to assess broad range of difficulties associated with thought disorder. Elevated scores indicate considerable difficulties with thought dysfunction.	
Behavioral/Externalizing Dysfunction (BXD)	.83
23-item scale designed to assess a broad range of behavioral problems. High scores indicate a broad range of externalizing behaviors while low scores indicate a higher than average level of behavioral-constraint.	
Restructured Clinical (RC) Scales	
Demoralization (RCd)	.92
25-item scale designed to measure emotional discomfort associated with feeling discouraged, demoralized, pessimistic, and overwhelmed. High scores indicate extreme emotional discomfort.	
Somatic Complaints (RC1)	.82
27-item scale designed to measure preoccupation with bodily and health concerns. High scores indicate above-average complaints of weakness, fatigue or chronic pain.	
Low Positive Emotions (RC2)	.79
17-tiem scale designed to measure the lack of positive emotional experiences. Low scores correspond with a high level of psychological well-being and a wide range of positive emotional experiences.	
Cynicism (RC3)	.85
15-item scale designed to measure view of human nature. Low scores indicate the test taker believes others to be well-intentioned and trustworthy, whereas high scores indicate a high level of cynicism of other people's motivation.	
Antisocial Behavior (RC4)	.83
22-item scale designed to measure antisocial behaviors and family conflict. Low scores suggest the test-taker reported a below-average level of past antisocial behavior.	
Ideas of Persecution (RC6)	.79
17-item scale designed to measure persecutory ideation associated with psychotic disorders. High scores indicate an above-average believe of persecution and paranoia.	

Table 1 cont.

Dysfunctional Negative Emotions (RC7)	.87
24-item scale designed to measure the extent to which the test-taker endorses negative emotional experiences. Low scores indicate a below-average level of reported dysfunctional negative emotional experiences.	
Aberrant Experiences (RC8)	.74
18-item scale designed to measure unusual thought and perceptual experiences of disordered thinking. High scores are associated with symptoms of psychotic disorders and disorganized thinking.	
Hypomanic Activation (RC9)	.77
28-item scale designed to measure a variety of emotions, cognition, behaviors and attitudes consistent with mania or hypomania. High scores suggest a high level of engagement with the environment.	

Note: The n for internal consistencies is 594

Table 2

MMPI-2-RF Substantive Scale Scores by Group

MMPI-2-RF Scale	WNL		L-r or K-r Elevated		L-r and K-r Elevated		F	df	p ≤
	Mean	(SD)	Mean	(SD)	Mean	(SD)			
N	**406**		**134**		**29**				
Higher-Order									
EID	47.80	(10.46)	39.54	(6.20) [a]	35.10	(3.60) [a, b]	124.961	108.10	.001
THD	51.17	(9.34)	47.40	(9.27) [a]	43.07	(5.08) [a, b]	32.485	87.17	.001
BXD	64.47	(10.92)	53.94	(9.67) [a]	48.07	(9.16) [a, b]	83.596	75.47	.001
RC Scales									
RCd	51.64	(9.64)	43.20	(5.93)[a]	39.52	(3.50)[a, b]	119.898	105	.001
RC1	48.05	(10.14)	45.05	(8.41)	41.10	(6.19)[a]	17.570	82.79	.001
RC2	47.32	(10.01)	42.37	(6.76)	40.21	(6.01) [a]	29.724	82.77	.001
RC3	56.99	(11.18)	50.63	(11.08)[a]	39.90	(5.33) [a, b]	114.692	93.34	.001
RC4	66.69	(11.33)	56.07	(10.68)[a]	49.59	(10.18) [a, b]	74.243	74.21	.001
RC6	56.44	(11.74)	51.20	(10.60)[a]	45.86	(5.77) [a, b]	40.953	91.46	.001
RC7	46.84	(9.24)	39.99	(5.95) [a]	35.52	(2.86) [a, b]	134.521	119.62	.001
RC8	49.81	(8.33)	45.34	(8.16) [a]	41.72	(4.67) [a, b]	41.977	85.86	.001
RC9	50.82	(9.44)	43.53	(6.63) [a]	38.76	(5.69) [a, b]	81.343	82.79	.001

Note:MMPI-2-RF = Minnesota Multiphasic Personality Inventory-2-Restructured Form; EID = Emotional /Internalizing Dysfunction; THD = Thought Dysfunction; BXD = Behavioral /Externalizing Dysfunction; RC = Restructured Clinical; RCd = Demoralization; RC1 = Somatic Complaints; RC2 = Low Positive Emotions; RC3 = Cynicism; RC4 = Antisocial Behavior; RC6 = Ideas of Persecution; RC7 = Dysfunctional Negative Emotions; RC8 = Aberrant Experiences; RC9 = Hypomanic Activation; Personality Psychopathology; [a] = Significant difference from WNL; [b] = Significant difference from One Elevated Scale.

Table 3

Collateral Measure Z-Scores by Group

Collateral Measure	WNL Mean (SD)	L-r or K-r Elevated Mean (SD)	L-r and K-r Elevated Mean (SD)	F	df	p ≤
N	**319**	**130**	**29**			
Internalizing						
BDI	.01 (.80)	-.35 (.87)[a]	-.67 (.41)[a, b]	33.935	89.67	.001
SSD	-.02 (.88)	-.32 (.77)[a]	-.55 (.47)[a]	17.861	87.46	.001
Externalizing						
BIS	.15 (.88)	-.57 (.92)[a]	-.91 (.94)[a]	41.554	68.81	.001
DAST	.08 (1.03)	-.28 (.85)	-.44 (.75)[a, b]	11.821	78.35	.001
MAST	.10 (1.04)	-.30 (.79)[a]	-.46 (.61)[a]	16.545	83.59	.001
STPI-TA	.02 (.81)	-.45 (.71)[a]	-.67 (.51)[a]	35.609	82.25	.001
Thought Disorder						
MIS	-.05 (.85)	-.15 (.99)	-.73 (.36)[a, b]	37.120	100.21	.001
PAS	-.10 (.19)	-.09 (.23)	-.19 (.09)[a, b]	10.855	94.31	.001

Note: BDI = Beck Depression Inventory; SSD = Screener for Somatoform Disorder; BIS = Barratt Impulsivity Scale; DAST = Drug Abuse Screening Test; MAST = Michigan Alcohol Screening Test; STPI-TA= Trait Anger subscale of the State-Trait Personality Inventory; MIS = Magical Ideation Scale; PAS = Perceptual Aberration Scale; [a] = Significant difference from WNL; [b] = Significant difference from One Elevated Scale.

Table 4

Effect Sizes for Post-hoc group comparisons MMPI-2-RF Scales

Scale	WNL -1 elevation	WNL – 2 elevations	1 elevation – 2 elevations
H-O Scales			
BXD	1.02	1.63	0.62
THD	0.41	1.07	0.58
EID	0.96	1.62	0.88
RC Scales			
RCd	0.99	1.60	0.76
RC1	0.32	0.83	0.54
RC2	0.58	0.86	0.34
RC3	0.55	1.86	1.23
RC4	0.96	1.59	0.62
RC6	0.47	1.14	0.63
RC7	0.88	1.65	0.96
RC8	0.55	1.20	0.55
RC9	0.89	1.55	0.77

Note: All effect sizes were calculated using Cohen's d.

Table 5

Effect Sizes for Post-hoc group comparisons collateral measures

Measure	WNL – 1 elevation	WNL – 2 elevations	1 elevation – 2 elevations
Internalizing			
BDI	0.43	1.06	0.47
SSD	0.37	0.75	0.37
Externalizing			
BIS	0.80	1.16	0.37
DAST	0.38	0.58	0.20
MAST	0.42	0.66	0.24
STPI-A	0.62	1.01	0.35
Thought Disorder			
MIS	0.12	1.06	0.78
PAS	0.05	0.60	0.59

Note: All effect sizes were calculated using Cohen's d.