An Exploration of the Impact of Invalid MMPI–2 Protocols on Collateral Self-Report Measure Scores

Johnathan D. Forbey & Tayla T. C. Lee


To link to this article: https://doi.org/10.1080/00223891.2011.608757

Published online: 14 Oct 2011.

Submit your article to this journal

Article views: 628

View related articles

Citing articles: 12 View citing articles
An Exploration of the Impact of Invalid MMPI–2 Protocols on Collateral Self-Report Measure Scores

JOHNATHAN D. FORBEY1 AND TAYLA T. C. LEE2

1Department of Psychological Science, Ball State University
2Department of Psychology, Kent State University

Although a number of studies have examined the impact of invalid MMPI–2 (Butcher et al., 2001) response styles on MMPI–2 scale scores, limited research has specifically explored the effects that such response styles might have on conjointly administered collateral self-report measures. This study explored the potential impact of 2 invalidating response styles detected by the Validity scales of the MMPI–2, overreporting and underreporting, on scores of collateral self-report measures administered conjointly with the MMPI–2. The final group of participants included in analyses was 1,112 college students from a Midwestern university who completed all measures as part of a larger study. Results of t-test analyses suggested that if either over- or underreporting was indicated by the MMPI–2 Validity scales, the scores of most conjointly administered collateral measures were also significantly impacted. Overall, it appeared that test-takers who were identified as either over- or underreporting relied on such a response style across measures. Limitations and suggestions for future study are discussed.

Individuals who take psychological inventories across numerous domains and settings might not always respond with a reflective or representative self-report of their current level of psychological functioning. Possible reasons for misrepresenting psychological functioning usually revolve around some form of secondary gain, such as reducing legal ramifications in criminal trials, achieving favorable decisions in custody or disability evaluations, increasing chances of employment in employee screening, or receiving another desired outcome in other assessment situations (Rogers, 2008). Depending on the desired outcome of the assessment, individuals might employ, either deliberately or unintentionally, one of two common approaches to misrepresent their psychological functioning: overreporting or underreporting of psychological symptoms. Nichols, Greene, and Schmolck (1989) termed such approaches to completing psychological instruments content-responsive faking.

In content-responsive faking, an individual alters (either consciously or unconsciously) his or her responses to items to achieve some desired outcome (i.e., secondary gain). Nichols et al. (1989) defined content-responsive faking in the overreporting direction as an individual’s attempt to present one’s self in a less favorable, more pathological fashion than is reflective of actual psychological functioning. On the other hand, content-responsive faking in the underreporting direction is described as an attempt to present one’s self in an overly favorable, less pathological fashion not reflective of actual psychological functioning.

Unlike some measures of social, behavioral, or psychological functioning, the Minnesota Multiphasic Personality Inventory–2 (MMPI–2; Butcher et al., 2001) contains a number of Validity scales designed to address concerns related to potential patterns of distorted responding on the part of the test taker. In terms of overreporting, the MMPI–2 contains three primary scales, Infrequency (F), Back Infrequency (FB), and Infrequency Psychopathology (FP), that are designed to detect this general response style related to the self-report of psychological symptoms (Butcher et al., 2001). Several studies have explored the ability of these scales to detect overreporting in various combinations across settings. For example, in coaching and simulation studies utilizing college students (Bagby, Rogers, & Buis, 1994), inmates (Iverson, Franzen, & Hammon, 1995), and veterans seeking compensation for posttraumatic stress disorder (PTSD; Arbisi, Ben-Porath, & McNulty, 2006), participants who were instructed to exaggerate psychological symptoms typically obtained significantly higher scores on the overreporting Validity scales of the MMPI–2 compared to individuals who completed the test under standard instructions. Overall, a meta-analysis conducted by Rogers, Sewell, Martin, and Vitacco (2003) indicated the overreporting scales adequately detected exaggerated responding on the MMPI–2.

With regard to underreporting, the MMPI–2 contains three scales—Lie (L), K (Correction), and Superlative Self-Presentation (S)—designed to detect this general response style (Butcher et al., 2001). Several studies have explored the ability of these scales in different combinations to detect underreporting across numerous settings and using varied methodologies. As with overreporting, the vast majority of these studies rely on simulators as the primary content-responsive faking group. For example, coaching and simulation studies utilizing college students (Baer, Wetter, Nichols, Greene, & Berry, 1995; Graham, Watts, & Timbrook, 1991) and inpatients (Bagby et al., 1997) found significantly higher scores on L and K for underreporters. Archer, Handel, and Couvadelli (2004) conducted a study utilizing inpatients identified as underreporting based on a “known groups” design, in which a group of patients who produced within normal limits profiles (i.e., T scores less than 65 on the eight basic Clinical scales) shortly after admission were classified as underreporters. This study demonstrated that scores on the S scale significantly added to scores on L and K in the identification of defensiveness in a psychiatric population.
Finally, a meta-analysis by Baer and Miller (2002) suggested that the MMPI–2 Validity scales used to detect underreporting adequately identified this response style.

Previous studies have also explored the impact that response styles detected by the MMPI–2 over- and underreporting Validity scales have on MMPI–2 substantive scale scores across a variety of populations and settings. In simulation studies, results have indicated that scores on MMPI–2 substantive scales (e.g., Clinical, Content, and Restructured Clinical) were significantly higher for simulated overreporters compared to individuals who took the test under standard instructions, suggestive of higher levels of psychological, social, and behavioral discord when compared to others (Graham et al., 1991; Sellbom, Ben-Porath, Graham, Arbisi, & Bagby, 2005; Storm & Graham, 2000). For underreporting response styles detected by MMPI–2 Validity scales, coaching and simulation studies utilizing college students (Baer et al., 1995; Graham et al.) and inpatients (Bagby et al., 1997) found significantly lower scores on several of the MMPI–2 clinical scales when compared to individuals who completed the instrument under standard instructions. Overall, this research suggests that score elevations on MMPI–2 Validity scales designed to detect over- or underreporting appear to influence MMPI–2 substantive scale scores in the expected direction. Individuals with score elevations on MMPI–2 overreporting scales tend to have significantly higher scores on the vast majority of MMPI–2 substantive scales, whereas individuals with score elevations on the underreporting Validity scales typically have significantly lower scores on MMPI–2 substantive scales.

In contrast to research examining the effect of detected response styles on scores of MMPI–2 substantive scales, relatively little attention has been given to the impact that such detected response styles might have on scale scores of other, jointlly administered self-report measures of social, behavioral, or psychological functioning. In the only known study to examine the effect of MMPI–2 detected response styles on collateral measures, Garcia, Franklin, and Chambliss (2010) examined the MMPI–2 overreporting scales in a group of veterans diagnosed with PTSD. The authors reported that scores on the MMPI–2 clinical scales were significantly higher for individuals with high scores on the F scale, in concurrence with previous research (e.g., Arbisi et al., 2006). Garcia et al. also demonstrated that scores on non-MMPI–2, collateral measures assessing symptoms of depression and PTSD administered at the same time as the MMPI–2 were significantly higher in the overreporting group, suggesting a more generalized style of responding across measures. However, this study had several limitations. First, Garcia and colleagues did not exclude participants who had a T score greater than 80 on the MMPI–2 True Response Inconsistency (TRIN) scale, which can lead to artificial elevations on F (Graham, 2006). Additionally, the authors employed a liberal cut score to identify overreporters in their sample, using an F scale T score greater than 65 when cut scores of 90 or greater are recommended by Butcher et al. (2001) for the population examined (i.e., clinical outpatients).

This study was designed to further examine the effect that content-responsive faking, specifically over- and underreported response styles, detected by MMPI–2 Validity scales, might have on jointly administered self-report measures of social, behavioral, and psychological functioning. Using a criterion groups design, we examined scale score differences between subgroups of college students who had “naturally” engaged in an over- or underreporting response style and individuals who produced interpretable MMPI–2 profiles, as determined by scores on the MMPI–2 Validity scales. Based on previous research on the effect of content-responsive faking response styles on MMPI–2 substantive scale scores (e.g., Sellbom et al., 2005), as well as the recent research conducted by Garcia et al. (2010), we hypothesized that individuals engaging in such a response style approach would have readily identifiable differences not only on MMPI–2 substantive scales, but also on conjointly administered collateral measures. Specifically, we hypothesized that individuals identified as overreporting by MMPI–2 Validity scale scores (i.e., elevations on F, FB, or Fp) would exhibit scores on both MMPI–2 substantive scales and conjointly administered collateral measures indicative of higher psychopathology or negative functioning than individuals who produced interpretable MMPI–2 profiles. Further, we hypothesized that individuals who were identified as underreporting by their scores on MMPI–2 Validity scales (i.e., elevations on L, K, or S) would exhibit scores indicative of positive functioning or less psychopathology on MMPI–2 substantive scales and collateral measures when compared to individuals with interpretable MMPI–2 profiles. However, as previous research has indicated that not all substantive scales of the MMPI–2 exhibit consistent differences in scores as a result of under- or overreporting approaches to completing the test, no specific predictions were made as to which MMPI–2 scales or collateral self-report measures would exhibit differential scores between the groups.

**METHOD**

**Participants**

The initial group of participants included 1,197 undergraduate students (419 men, 778 women) from a Midwestern U.S. university enrolled in introductory psychology classes. To reduce error variance in analyses, individuals who provided content nonresponsive MMPI–2 profiles, defined by omission of responses on 30 items or more (Cannot Say [CNS/?]) or T scores of 80 or more on Variable Response Inconsistency (VRIN) or TRIN, were removed from this study. A total of 85 individuals produced content nonresponsive MMPI–2 profiles. No statistically significant differences between included and excluded participants were evident in terms of age or gender. Statistically significant differences in ethnicity or racial group were demonstrated between those who produced content nonresponsive MMPI–2 profiles and those who did not, $\chi^2(2, N = 1,197) = 17.911$, $p < .001$, $\Phi = .12$. A slightly higher proportion of individuals reporting they were African American or of another, unidentified ethnicity produced content nonresponsive MMPI–2 results, although the effect size of this difference was small (Cohen, 1988).

After exclusions, the final sample of 1,112 participants included 389 men and 723 women. Participants included in the final sample ranged in age from 18 to 44 ($M = 19.21, SD = 1.72$) and were Caucasian ($n = 994$, 89.4%), African American ($n = 67$, 6.0%), or of another or unidentified ethnicity ($n = 51$, 4.6%).

**Measures**

**Minnesota Multiphasic Personality Inventory–2.** Revised from the original Minnesota Multiphasic Personality Inventory–2. (Revised by G. O. M. McNeil, 1980). The MMPI–2, 1990. Published by the American Psychiatric Publishing, Inc. Copyright © 1990 by the American Psychiatric Association. All rights reserved. MMPI–2 is a registered trademark of the American Psychiatric Association. MMPI–2 is a registered trademark of the Minnesota Multiphasic Personality Institute.
The Belief in Personal Control Scale (Berrenberg, 1987) is a 24-item measure of aggression rated on a 5-point Likert scale ranging from 1 (never) to 5 (very frequently) that assesses an individual’s acceptance of authority. Rigby (1987) examined the convergent validity of the ABI scores and demonstrated significant associations between ABI scores and peer ratings of college-aged participants’ attitudes toward authority. In this study, internal consistency was acceptable across genders (α = .73).

The Authority Behavior Inventory (ABI; Rigby, 1987) is a 24-item measure rated on a 4-point Likert scale ranging from 1 (not at all true of me) to 4 (totally agree) assessing distorted negative perceptions common in individuals suffering from depression. The CTI provides a total score, as well as scores on three dimensions: View of Self (10 items), View of World (10 items), and View of Future (10 items). Lovibond and Lovibond (1995) provided initial support for the convergent and discriminant validity of the DASS scale scores in a sample of college students through an examination of correlations with collateral measures of stress and anxiety. In this study, internal consistencies (α) for combined genders were acceptable and ranged from .80 to .93 for total and subscale scores.

The Cognitive Triad Inventory (CTI; Beckham, Leber, Watkins, Boyer, & Cook, 1986) is a 36-item instrument rated on a 7-point Likert scale ranging from 1 (totally agree) to 7 (totally disagree) assessing distorted negative perceptions common in individuals suffering from depression. The CTI provides a total score, as well as scores on three dimensions: View of Self (10 items), View of World (10 items), and View of Future (10 items). Internal consistencies for the four scale scores were .91 and .90, respectively.

The Hypomanic Personality Scale (HPS; Eckblad & Chapman, 1986) consists of 48 true–false items designed to assess overactive and gregarious behavior associated with hypomanic episodes. Eckblad and Chapman (1986) provided evidence of the convergent and discriminant validity of HPS scores in a group of college students with a variety of measures of emotional, behavioral, and thought dysfunction. Internal consistencies were acceptable across genders in this study were acceptable to good, ranging from .83 to .98.
consistency of HPS scores in this study for combined genders was good (α = .87).

The Impostor Phenomenon Scale (IPS; Clance, 1985) consists of 20 items rated on a 5-point Likert scale ranging from 1 (not at all true) to 5 (very true) designed to assess an individual’s internalized doubt about his or her intellectual and other abilities. Chrisman, Pieper, Clance, Holland, and Glickhauf-Hughes (1995) reported that IPS scores demonstrated both convergent and discriminant validity with another measure of the impostor phenomenon, as well as measures conceptually related to the impostor phenomenon (e.g., depression, anxiety, etc.) collateral measures, in a group of college students. In this study, internal consistency of IPS scores was good for combined genders (α = .90).

The Impulsive Nonconformity Scale (INS; Chapman et al., 1984) consists of 51 true–false items designed to assess impulsive and antisocial behaviors during premorbid adjustment of some individuals experiencing psychosis. Chapman and colleagues (1984) provided support for the convergent validity of INS scores in college students and inpatients utilizing measures of psychotic symptoms and sensation seeking. In this study, internal consistency across genders for scores on the INS was acceptable (α = .86).

The Interpersonal Dependency Inventory (IDI; Hirschfield et al., 1977) consists of 48 items rated on a 4-point Likert scale ranging from 1 (not characteristic of me) to 4 (very characteristic of me) designed to provide scores on three dimensions: Emotional Reliance on Others (18 items), Lack of Self-Confidence (16 items), and Assertion of Autonomy (14 items). Hirschfield and colleagues (1977) provided evidence of the internal consistency as well as the concurrent validity of IDI dimension scores with conceptually relevant measures of depression, anxiety, and neuroticism across college students, psychiatric patients, and individuals within the general population. In this study, internal consistency (α) for combined genders for scores across the three dimensions was acceptable, ranging from .78 to .83.

The Persecutory Ideation Questionnaire (PIQ; McKay, Langdon, & Coltheart, 2006) consists of 10 items rated on a 5-point Likert scale ranging from 0 (very untrue) to 4 (very true) designed to assess cognitions that are persecutory in nature. McKay and colleagues (2006) demonstrated convergent validity for scores on the PIQ in both clinical and nonclinical settings with a measure of paranoia. In this study, internal consistency of scores for combined genders was good (α = .90).

The Social Anhedonia Scale–Revised (SAS–R; Mishlove & Chapman, 1985) consists of 40 true–false items designed to assess characteristics of social isolation common in individuals judged to be exhibiting thinking and behaviors often predating the onset of schizophrenia. Using various social and academic collateral measures, Mishlove and Chapman (1985) demonstrated the convergent validity of scores on the SAS–R in a sample of college students. In this study, internal consistency of SAS–R scores for combined genders was good (α = .86).

The Self-Efficacy Scale (Sherer et al., 1982) consists of 30 items rated on a 5-point Likert scale ranging from 1 (disagree strongly) to 5 (agree strongly) designed to assess two dimensions of personal expectations of mastery and success: General Efficacy (17 items) and Social Efficacy (6 items). Initial research by Sherer and colleagues (1982) demonstrated the factor structure of the measure in a college sample and provided support for the convergent and discriminant validity of the two dimension scores in a sample of Veteran’s Administration recovery patients utilizing historical educational and employment data. Internal consistencies of scores on the two subscales across genders in this study was acceptable, with αs of .81 and .67, respectively.

The Shyness and Sociability Scale (Cheek & Buss, 1981) consists of 14 items rated on a 5-point Likert scale ranging from 0 (never avoid) to 4 (always avoid) designed to assess two dimensions of social interaction: Shyness (9 items) and Sociability (5 items). Cheek and Buss (1981) provided initial support for the convergent and discriminant validity of scores on the two dimensions in a sample of college students demonstrating a differential pattern of associations with measures of shyness and sociability. In this study, the internal consistency across genders was acceptable, with αs for scores on the two subscales of .67 and .81, respectively.

Procedure

In exchange for credit in their general psychology courses, participants completed computerized administrations of the MMPI–2 and one of two packets of collateral measures during a single session (as part of a larger study). The MMPI–2 was presented either prior to or after the collateral measures, and collateral measures were administered via a Latin square design. All participants were assessed in accordance with procedures approved by the institutional review board of the university where the data were collected, and were free to withdraw their participation at any time.

After removing content nonresponsive profiles (as described earlier), participants were assigned to one of three distinct groups: content-responsive faking-overreporting (CRF-OR), content-responsive faking-underreporting (CRF-UR), and within normal limits reporting (WNL). Classification was based on their T scores on standard MMPI–2 Validity scales. The MMPI–2 Validity scale T score cutoffs used for group assignment in this study were consistent with those suggested for use in nonclinical samples by Graham (2006) and the MMPI–2 Manual for Administration, Scoring, and Interpretation (Butcher et al., 2001). Participants were assigned to the CRF-OR group if they had T scores of 80 or greater on the Infrequency (F) scale, 90 or greater on the Back Infrequency (FB) scale, or 100 or more on the Infrequency Psychopathology (FP) scale. Participants were assigned to the CRF-UR group if they had MMPI–2 T scores of 80 or greater on the Lie (L) scale, a Correction (K) scale score of 65 or greater, or a Superlative Self-Presentation (S) scale score of 70 or greater. In addition, no individuals in the CRF-OR group could have elevations on the underreporting scales, and vice versa. Finally, the WNL group consisted of individuals who had no elevations on any of the overreporting or underreporting scales.

Using this group assignment procedure, of the 1,112 included participants, 105 were assigned to CRF-OR, 44 were assigned to CRF-UR, and 962 were assigned to WNL. One participant had score elevations on at least one of the scales used to define CRF-OR and one of the scales used to define CRF-UR. Due to this pattern of scores, this person’s data were excluded from further analyses. Chi-square analyses indicated that administering the MMPI–2 either after or prior to the collateral measures did not have an impact on the proportion of individuals identified as CRF-OR or CRF-UR. Results of comparative analyses also suggested no statistically significant
differences between participants assigned to the three response style groups were evident in terms of reported age and race or ethnicity. Statistically significant differences of a small effect size were demonstrated between these groups for gender, \( \chi^2(2, N = 1,111) = 11.74, p < .003, \Phi \|\| = \pm .10 \), with men more likely to be assigned to the WNL group than women.

**Data Analyses**

To examine potential differences in MMPI–2 substantive scales among the three response style groups that were created, a series of \( t \) tests were computed comparing mean T scores on MMPI–2 substantive scales. To reduce potential Type I error, a Bonferroni correction was applied and the critical alpha for these analyses was \( .001 (.05/38) \). Cohen’s (1988) \( d \) effect sizes, with .3, .5, and .8 reflecting small, medium, and large effects, respectively, were computed comparing mean scores on the collateral measure analyses were converted to a \( z \) score metric to facilitate the reader’s interpretation. As with the MMPI–2 substantive scale analyses, to reduce potential Type I error, a Bonferroni correction was applied, setting the critical alpha for these analyses at \( .001 (.05/38) \) and effect sizes \( d \); Cohen, 1988) were reported for all analyses, regardless of significance.

To examine for potential differences between collateral measure scores between response style groups, a series of \( t \) tests were computed comparing mean scores on the collateral measures. As each of the 15 collateral measures has unique scale score ranges, means, and standard deviations, the descriptive statistics reported for the collateral measure analyses were converted to a \( z \) score metric to facilitate the reader’s interpretation. As with the MMPI–2 substantive scale analyses, to reduce potential Type I error, a Bonferroni correction was applied, setting the critical alpha for these analyses at \( .001 (.05/38) \) and effect sizes \( d \); Cohen, 1988) were reported for all analyses, regardless of significance.

**RESULTS**

The first set of analyses explored the differences on substantive scale scores between the WNL and CRF-OR response style group, as well as for the WNL and CRF-UR response style groups. Table 1 contains the results of these analyses. After a Bonferroni correction, \( t \) test results suggested scores on all of the 38 examined substantive scales of the MMPI–2 were significantly different between the CRF-OR and WNL groups. Examination of mean scores for the two groups indicated the CRF-OR group scored significantly higher than the WNL group on all MMPI–2 substantive scales. Median effect sizes for the CRF-OR and WNL comparisons were 1.29 (range = .83–2.04) for the Clinical scales, 1.24 (range = .69–1.71) for the Content scales, 1.32 (range = .84–1.60) for the Restructured Clinical scales, and .95 (range = .59–1.70) for the PSY–5 scales. For the comparison of mean MMPI–2 substantive scores of CRF-UR and WNL response style groups, 33 of the 38 examined scales demonstrated statistically significant differences. Inspection of mean scores for the two groups indicated that, on those scales that demonstrated statistically significant differences, the CRF-UR group had lower scores than the WNL group, with the exception of Clinical Scale 3 (Hy). Median effect sizes for the CRF-UR and WNL comparisons were –.81 (range = .71 to –1.40) for the Clinical scales, –1.09 (range = –.55 to –1.49) for the Content scales, –.98 (range = –.26 to –1.40) for the RC scales, and –.52 (range = –.28 to –1.06) for the PSY–5 scales. It should be noted that for the CRF-UR and WNL effect size analyses, nonsignificant results were included in median calculations.

Table 2 reports the differences between scores on the collateral measures for the CRF-OR and WNL response style groups, as well as CRF-UR and WNL response style groups. Within the table, total scores and subscale scores for collateral measures (where appropriate) are displayed and the measures are divided into three broad categorizations (i.e., externalizing, internalizing, and social) based on the general content or the construct examined by each collateral measure.

Overall, results indicated, after applying a Bonferroni correction, the CRF-OR response style group demonstrated significantly different mean scores from the WNL group across all eight of the externalizing collateral measures (median effect size \( |d| = .84 \), 13 of the 16 internalizing collateral measures (median effect size \( |d| = –.86 \), and 7 of the 11 social collateral measures (median effect size \( |d| = –.49 \)). The aforementioned median effect sizes include all results (i.e., significant and nonsignificant) in their calculations. For the scales that demonstrated a statistically significant difference, inspection of the mean scores indicated the CRF-OR group reported increased negative functioning on the criterion measures when compared to the WNL group.

With respect to the CRF-UR group, when compared to the WNL group and after applying a Bonferroni correction, six of the eight externalizing collateral measures (median effect size \( |d| = .86 \)), 5 of the 16 internalizing collateral measures (median effect size \( |d| = .64 \)), and 4 of the 11 social collateral measures (median effect size \( |d| = .61 \)) demonstrated significant differences in mean scores between groups. The aforementioned median effect sizes include all results (i.e., significant and nonsignificant) in their calculations. Inspection of mean scale scores suggested that the CRF-UR group reported higher positive functioning on the criterion measures than the WNL group.

**DISCUSSION**

The purpose of this study was to examine the impact of invalid response styles (over- and underreporting) on MMPI–2 substantive scale scores, as well as scores on concurrently administered collateral self-report measures. As hypothesized, results suggested that an invalid, over- or underreporting response style detected by MMPI–2 Validity scale scores had a significant effect on the majority of scores on the MMPI–2 substantive scales and self-report collateral measures when compared to individuals who were responding in a nonbiased manner. Specifically, compared to individuals identified as having WNL Validity scale profiles, individuals who were identified as overreporting symptoms on the MMPI–2 were significantly more likely to report a greater number of psychological, behavioral, and social difficulties, as well as fewer aspects of positive functioning, as reflected by scores on both MMPI–2 substantive scales and concurrently administered collateral self-report measures. Additionally, although to a lesser effect size than for overreporting, individuals identified by MMPI–2 Validity scale scores as underreporting were more likely to report significantly fewer psychological, behavioral, and social difficulties, as well as more indications of positive functioning, as indicated by their scores on the same measures when compared to individuals who had interpretable MMPI–2 profiles.

In terms of the effect of content-responsive faking on the MMPI–2 substantive scales, for overreporting all 38 examined scales demonstrated significantly higher scores in the
overreporting group compared to the WNL group. Although the effect sizes for this sample were large for all but two of the scale score comparisons (PSY–5 scales AGGR and DISC reached a medium effect size), they were somewhat lower overall compared to the effect sizes reported by Sellbom and colleagues (2005) in a reanalysis of data utilizing college student, psychiatric, and medical samples. However, the data examined by Sellbom et al. involved simulation studies of overreporting across settings, whereas this study relied on a naturally occurring group of individuals who were identified as overreporting by the MMPI–2 F, FB, and Fp scales. Further, the current effect sizes for the Clinical scales are similar to those reported by Garcia et al. (2010), who also examined naturally occurring groups of veterans diagnosed with PTSD who were overreporting or responding honestly to the instrument.

With respect to underreporting, scores on 33 of the 38 MMPI–2 substantive scales were significantly different in the underreporter group when compared to those of the WNL group. Of those scales that demonstrated a significant and expected difference in scores, the effect sizes in this study are similar to those reported by Sellbom et al. (2005) in their reanalysis of an undergraduate sample who were instructed to simulate defensive
responding (i.e., underreporting). With the exception of Clinical Scale 3 (Hy), all of the differences were in the expected direction. Post-hoc analyses demonstrated this unexpected direction of the effect for Clinical Scale 3 was potentially a result of scale heterogeneity. Indeed, on the Harris–Lingoes subscales for Clinical Scale 3, significantly higher scores were demonstrated for those subscales that reflect comparatively positive functioning as opposed to those subscales that reflect negative psychological functioning.

Finally, as with previous research (e.g., Graham et al., 1991; Lim & Butcher, 1996), several scales demonstrated no significant differences between underreporters and WNL reporters. However, an examination of this study and other studies (e.g., Sellbom et al., 2005) suggest no discernable patterns in terms
of which MMPI–2 scales reliably differ. For example, similar to this study, previous research reported no differences for college males who were instructed to underreport versus standard instructions on Clinical Scale 2 (e.g., Graham et al.; Lim & Butcher), whereas others have found significant differences on this scale (e.g., Sellbom et al.). Further, both Sellbom and colleagues and this study found significant differences for Clinical Scale 1, whereas both Graham et al. and Lim and Butcher did not. Thus it appears this issue needs to be explored further in future studies.

Most germane to the purpose of this study, results indicated that individuals who were identified as engaging in a content-responsive faking response style by MMPI–2 Validity scales demonstrated numerous differences on conjointly administered, self-report collateral measures. With respect to overreporting, 27 of 35 total scale or subscale scores of conjointly administered collateral measures reflected significantly higher levels of dysfunction for the overreporters group compared to WNL responders. These findings suggest that if an individual is identified on the MMPI–2 as engaging in an overreporting response style, scores on any conjointly administered, self-report, collateral measure(s) are also likely to reflect an attempt to present in an overly negative or psychologically dysfunctional manner.

Conversely, 15 of the 35 total scale or subscale scores on conjointly administered, self-report, collateral measures reflected significantly lower levels of dysfunction for the underreporters compared to the WNL group. Post-hoc exploration suggested that the reduction in the number of collateral scale score differences for the underreporting group when compared to the overreporting group might be reflective of the comparative limited power for this subsample. In keeping with Graham’s (2006) hypothesis that underreporting is harder to detect, for some analyses, the underreporting group contained as few as 15 participants, greatly reducing the ability of tests to detect a statistically significant result when compared to the relatively larger overreporting group sizes (e.g., there were data from 47 participants in the smallest overreporting group). Supporting the hypothesis of reduced power, an examination of the findings by group size revealed that 8 of 15 (53.3%) collateral measure scores differed for the underreporting group when the sample size was increased, whereas only 7 of 20 (35%) differed in the smallest group. Further, of the 20 scales that did not reflect a statistically significant difference in scores, 12 reached a medium effect size, 7 of which were in the smallest underreporting group.

Despite the limitations of the underreporting analyses, the results of the collateral measure analyses for the underreporting group suggest that, although not to an effect size as strong as demonstrated for the overreporting group, if an individual is identified on the MMPI–2 as engaging in an underreporting response style, scores on any conjointly administered, self-report, collateral measure(s) are also likely not reflective of their current psychological functioning. Interpretatively, we would suggest that an individual detected on the MMPI–2 as underreporting also has scores on conjointly administered collateral measures that are reflective of the individual’s attempt to present in an overly positive, psychologically healthy manner.

Taken as a whole, the results of the analyses of self-report, collateral measure score differences suggested that individuals who engage in content-responsive faking on the MMPI–2 (as identified by MMPI–2 Validity scales) are significantly more likely to produce scores on collateral measures that are different than those who engage in WNL responding on the Validity scales. Generally, for those collateral measures that examine negative functioning, when compared to individuals assumed to be responding honestly, our results indicated that individuals identified as overreporting tended to have higher scores on collateral measures, whereas individuals in the underreporting group had suppressed scores on collateral measures. Further, the opposite pattern was demonstrated for collateral measures that assessed aspects of positive functioning. Given the counterbalancing of the MMPI–2 and collateral measures, as well as the Latin square procedure followed in the administration of the collateral measures, this response set appears reflective of a generalized test-taking approach, rather than an artifact of administration order.

This study has several limitations. The sample used to examine the research questions in this study consisted of college student participants who likely had high levels of adaptive psychosocial functioning at the outset of the study, leading to four primary limitations. First, the use of a college sample leads to questionable generalizability to “real-world” settings where the MMPI–2 is used (e.g., clinical settings). However, given that this study utilizes empirically established cutoffs for non-clinical settings utilized to identify over- and underreporters, a similar pattern of responding would be expected across populations when empirically derived cut scores for the validity scales are utilized.

Second, as described earlier, use of college students led to a restricted ability to identify differences between underreporting and WNL groups as the mean level of functioning was most likely “positive” at the outset of the study. However, given that 15 of 35 (42.9%) collateral measure scores demonstrated significant differences between the groups, it is evident that engaging in an underreporting response style has some effect on collateral measure scores in this population.

Third, although each content-responsive faking group was identified by excluding individuals who were identified as engaging in content nonresponsiveness (i.e., CNS raw score > 30, VRIN T score > 80, or TRIN T score > 80), there was no readily identifiable motivation (i.e., apparent secondary gain) for participants to engage in a content-responsive faking test-taking approach (e.g., underreporting), despite being clearly identified as such by the MMPI–2 Validity scales. Compared to individuals in real-world settings who might have a bona fide rationale for utilizing a content-responsive faking approach, it is conceivable that some college students, who are naive to the research process, might believe that their research results might be subjected to individual scrutiny. Therefore, despite assurances of anonymity and confidentiality, they might feel compelled to engage in a manner that presents them in an overly negative (perhaps to seek attention) or overly positive light (perhaps to avoid further scrutiny). Regardless of their rationale, the use of such a “naturally” occurring sample of individuals engaging in content-responsive faking to examine the impact of such approaches on not only MMPI–2 substantive scales, but also conjointly administered collateral self-report measures, represents a significant methodological advancement when compared to previous research, especially in comparison to those studies relying solely on simulation data.

A fourth and final limitation of this study relates to the lack of a “gold standard” for assigning individuals to the content-responsive faking groups. Although the MMPI–2 Validity scale
scores utilized to identify each group were reflective of scores suggested by Graham (2006), as well as Butcher et al. (2001), it might be possible that those individuals who were believed to be engaging in an over- or underresponse style, or those in the WNL group, were misidentified, due in part to utilizing an approach that relied solely on cut scores for any individual scale (as opposed to the traditional configural or combination approach that is typically followed in clinical practice). Thus additional studies are needed to further explore the effect of detected response styles on collateral measures in additional populations of test takers (e.g., forensic, inpatient, outpatient) and using different methods of identifying response style groups.

To our knowledge, this study is the first conducted specifically to examine the impact of both forms of content nonresponsive faking, as identified by the MMPI–2 Validity scale scores, on collateral self-report measure scores. The results of this study suggest that individuals identified as “naturally” engaging in content-responsive faking approaches (i.e., over- and underreporting) have significantly different scores on both MMPI–2 substantive scales, and more important, conjointly administered collateral self-report measure scores. For research, these results suggest that excluding both MMPI–2 and collateral measure responses for individuals detected by MMPI–2 Validity scales as engaging in content-responsive faking from analyses will reduce the potential for error variance and spurious results. Clinically, these findings suggest that if either test-taking approach is identified by a mental health professional interpreting the MMPI–2, these findings suggest that if either test-taking approach is identified by the MMPI–2 Validity scale scores, as engaging in content-responsive faking from analyses will reduce the potential for error variance and spurious results. Clinically, these findings suggest that if either test-taking approach is identified by a mental health professional interpreting the MMPI–2, scores on any additional, conjointly administered self-report instruments should be viewed cautiously in terms of their validity. Overall, it appears that content-responsive faking, as identified by the MMPI–2 Validity scales, reflects a more generalized approach to self-presentation on objective personality measures that should be considered when multiple self-report measures are administered to an individual during a single assessment session.

ACKNOWLEDGMENT

Portions of this article were presented at the Annual Symposium on Recent MMPI–2, MMPI–2–RF, MMPI–A Research in St. Petersburg Beach, FL in March 2010.

REFERENCES


