3-2019

Examination of Differential Validity of MMPI-2-RF Scores by Gender and Ethnicity in Predicting Future Suicidal and Violent Behaviors in a Forensic Sample

Anthony M. Tarescavage
John Carroll University, atarescavage@jcu.edu

David M. Glassmire
Patton State Hospital

Danielle Burchett
California State University, Monterey Bay

Martin Sellbom
University of Otago

BRIEF REPORT

Examination of Differential Validity of MMPI-2-RF Scores by Gender and Ethnicity in Predicting Future Suicidal and Violent Behaviors in a Forensic Sample

Megan R. Whitman and Anthony M. Tarescavage
John Carroll University

David M. Glassmire
Patton State Hospital, Patton, California

Danielle Burchett
California State University, Monterey Bay

Martin Sellbom
University of Otago

Given the diversity of individuals who undergo psychological assessment, examining whether cultural bias exists in psychological assessment instruments (i.e., differential validity) is crucial. This issue occurs when a measure systematically over- or underpredicts a criterion across demographic groups or is associated with the criterion unequally across the groups. We tested the differential validity of a widely used psychological test, the Minnesota Multiphasic Personality Inventory (MMPI)-2-Restructured Form (MMPI-2-RF), as a function of gender (male, female) and ethnicity (Caucasian, African American, and Hispanic/Latino American) in large samples of forensic psychiatric inpatients. Regression models were estimated in a multigroup framework. The analyses yielded negligible to small statistical evidence of differential validity in MMPI-2-RF scores predicting the number of future suicidal behaviors and violent behaviors in the samples. This evidence supports use of the MMPI-2-RF as a generally unbiased instrument for predicting key criteria across genders and ethnicities in a forensic psychiatric population.

Public Significance Statement
This study supports use of the MMPI-2-RF in forensic settings as a generally unbiased instrument for predicting suicidal and violent behaviors across genders (male, female) and ethnicities (Caucasian, African American, and Hispanic/Latino American).

Keywords: MMPI-2-RF, test bias, risk assessment, forensic assessment

Suicide and violence pose risks to millions of individuals in the United States (Crosby, Han, Ortega, Parks, & Gfroerer, 2011; Smith et al., 2017). Predicting these behaviors is especially important in forensic populations, where suicidal and violent behaviors commonly occur (Metzner, 2002; Wolff, Blitz, Siegel, & Bachman, 2007). For these reasons, psycholo-

gists in forensic settings often complete risk assessments to determine the likelihood of suicidal or violent acts (Correia, 2000; Skeem & Monahan, 2011). The utility of psychological assessments to inform such predictions is of interest to forensic assessment researchers (Douglas, Ogloff, Nicholls, & Grant, 1999; Tarescavage, Glassmire, & Burchett, 2016, 2018).

Risk assessments of suicide and violence have limitations, as they typically include clinical interviews and evaluations of existing records (Borum, 1996), both of which are inherently subjective. Moreover, even when clinicians take steps to be objective in their assessments, implicit biases persist (Sue et al., 2007). These biases are of concern in forensic evaluations. For example, the Specialty Guidelines for Forensic Psychology, developed by the American Psychology-Law Society and American Academy of Forensic Psychology (2013), indicate that forensic psychologists should practice with impartiality and fairness, considering the impact of such biases. To increase the objectivity of risk assessments, clinicians can opt to supplement their evaluations with objective psychological assessment instruments. However, these instruments can also evidence bias. Test bias occurs when an assessment's scale scores demonstrate differential validity across demographic groups.

Notably, group mean differences across scale scores do not indicate test bias, as such differences can reflect genuine demographic variation. Rather, test bias is detected through analyses of differential predictive validity, which can manifest in two ways (Kaplan & Saccuzzo, 2018). First, a test can systematically over- or underpredict an outcome across demographic groups, such as gender or ethnicity (intercept bias). Second, a test can differ in the strength of associations across demographic groups (slope bias). Psychological assessment instruments should be evaluated for such biases, especially when being used to make potentially life-altering decisions about suicide and violence risk, as in the forensic field.

One instrument that has been used in forensic settings to predict suicide (Glassmire, Tarescavage, Burchett, Martinez, & Gomez, 2016; Tarescavage et al., 2018) and violence (Tarescavage et al., 2016) is the Minnesota Multiphasic Personality Inventory (MMPI)-2-Restructured Form (MMPI-2-RF; Ben-Porath & Tellegen, 2008/2011). This test is an updated version of one of the most widely used measures in forensic psychology (Archer, Buffington-Vollum, Stredny, & Handel, 2006), the MMPI-2 (Butcher et al., 2001). The MMPI-2-RF is a 338-item assessment of personality and psychopathology. Its items aggregate onto nine validity scales and 42 scales that measure substantive clinical content. The latter scales are arranged in a hierarchical structure, with three higher-order scales, nine restructured clinical scales, 23 specific problems scales, two interest scales, and five personality psychopathology-five scales. A substantial amount of evidence has accumulated for both reliability and validity of MMPI-2-RF scores across many populations, including those in forensic settings (see, e.g., Sellbom, in press, for a review).

Past research has examined the potential for differential validity of scale scores from the MMPI-2 (Butcher et al., 2001), the precursor to the MMPI-2-RF. Arbisi, Ben-Porath, and McNulty (2002) evaluated the MMPI-2 scale scores for racial bias in predicting concurrent psychological diagnoses among African American and Caucasian psychiatric inpatients. Step-down hierarchal regression analyses indicated intercept bias for some scales and criteria, but effect sizes were typically negligible to small. Notably, the bias generally resulted in the systematic underprediction of psychopathology in African American patients and overprediction of psychopathology in Caucasian patients.

Monnot, Quirk, Hoerger, and Brewer (2009) reported somewhat contrasting results. Their study used the Structured Clinical Interview for *DSM–III–R* Axis I Disorders patient version as a concurrent criterion for the prediction of diagnostic status in a sample of African American and Caucasian men seeking substance abuse treatment at a VA hospital. Monnot et al. (2009) found intercept and slope bias in most MMPI-2 scales, but the effect sizes were small. The scale scores both under- and overpredicted criteria for African Americans, but in nearly every case of slope bias, the MMPI-2 scale was less strongly associated with concurrent diagnoses among African American patients relative to Caucasian patients (to a small degree; Monnot et al., 2009).

To date, only one study has evaluated MMPI-2-RF scores for differential predictive validity. Marek, Ben-Porath, Sellbom, McNulty, and Heinberg (2015) examined scores for differential validity across gender, ethnicity, and age in a sample of African American, Caucasian, and Hispanic/Latino bariatric surgery candidates. They used concurrent interview record review information as external criteria. Similar to past MMPI-2 research, Marek et al. (2015) used step-down hierarchical multiple regression analyses yielding 12 (of 40) significant gender bias analyses. The analyses had negligible to small effect sizes. Of the 40 ethnicity analyses, 10 were statistically significant and these consistently underpredicted criteria among African Americans and overpredicted criteria in Caucasians to a small extent (Marek et al., 2015).

Current Study

Although research supports the utility of the MMPI-2-RF in suicide and violence risk assessments of forensic patients (Glassmire et al., 2016; Tarescavage et al., 2016, 2018), no study has examined the MMPI-2-RF scale scores for differential predictive validity across genders or ethnicities in a forensic sample. Moreover, past differential validity research using the MMPI-2 and the MMPI-2-RF did not include Hispanic/Latino American participants, and it has primarily utilized concurrent criteria. Past research has also been limited to step-down hierarchal multiple regression of differential validity, but Reynolds and Ramsay (2003) outlined the necessity of using diverse statistical techniques in evaluations for test bias to avoid unfavorable outcomes, such as discriminatory grouping or treatment.

For the reasons just noted, the purpose of the current study was to examine MMPI-2-RF scale scores for differential validity across genders (male, female) and ethnicities (Caucasian, African American, and Hispanic/Latino American) in the prediction of future suicidal and violent behaviors in a forensic psychiatric inpatient sample using a multigroup regression framework. Similar to past research with the MMPI instruments, we expected MMPI-2-RF scale scores to evidence differential validity to no more than a small extent.

Method

Participants

The current study used two subsamples of forensic psychiatric inpatients from a large state hospital in the Southwestern United

States. Prior research has used these samples to examine predictive validity of future suicidal (Glassmire et al., 2016; Tarescavage et al., 2018) and violent behaviors (Tarescavage et al., 2016). We chose these subsamples for the present study so that the differential validity analyses could not only address our research questions but also directly inform interpretation of the just mentioned published predictive validity analyses.

For the future suicidal behavior analyses, the initial sample was composed of 1,100 forensic psychiatric inpatients. Individuals were excluded if they produced invalid MMPI-2-RF protocols according to standard interpretive guidelines (CNS >18, VRIN-r > 80, TRIN-r > 80, F-r = 120, Fp-r > 100; Ben-Porath & Tellegen, 2008/2011). The final sample consisted of 751 patients (549 males, 202 females). The most common commitment types were *not guilty by reason of insanity* (NGRI; 54%), *mentally disordered offender* (postconviction parolees who were judicially determined to be dangerous by reason of mental disorder; 23%), and *incompetent to stand trial* (18%). The average age was 40.5 years (SD = 11.3). The average education level was 12.3 years (SD = 2.5). The sample was primarily comprised of Caucasian (56.9%), African American (24.3%), and Hispanic/Latino American (14.4%) patients. The most common *DSM–IV–TR* (American Psychiatric Association, 2000) Axis I disorders were schizophrenia (34.4%), schizoaffective disorder (24.4%), bipolar disorder (11.7%), and psychotic disorder not otherwise specified (8.9%). The total patient hospitalization time after testing ranged from less than 1 month to 19.6 years (M = 3.2 years, SD = 11.3 years). Additional information on the sample has been reported elsewhere (Tarescavage et al., 2018).

For the future violence analyses, the initial sample included 395 forensic psychiatric inpatients acquitted NGRI. Here too, individuals were excluded if they produced invalid MMPI-2-RF protocols, yielding a final sample of 303 patients (233 males, 70 females). The average age was 41.1 years. The sample was primarily comprised of Caucasian (56.6%), African American (21.9%), or Hispanic/Latino American (15.6%) patients. The most common primary *DSM–IV–TR* (American Psychiatric Association, 2000) Axis I disorders indicated in the medical records on the date of MMPI-2-RF testing were schizophrenia (36.0%), schizoaffective disorder (22.9%), bipolar disorder (13.8%), and psychotic disorder not otherwise specified (9.4%). The average hospitalization time after testing ranged from less than 1 month to 16.75 years (M = 3.8 years, SD = 3.9 years). Additional information on the sample has been reported elsewhere (Tarescavage et al., 2016). Descriptive statistics for the MMPI-2-RF scales across groups are presented in Supplementary Table 1. This study was approved by the California Department of State Hospitals institutional review board.

Measures

MMPI-2-RF. The MMPI-2 or the MMPI-2-RF was administered to patients in connection with their psychiatric care at the hospital, and MMPI-2 forms were rescored as the MMPI-2-RF. Past research using a forensic sample has demonstrated comparability of MMPI-2-RF scores from the MMPI-2 and MMPI-2-RF booklets (Tarescavage, Alosco, Ben-Porath, Wood, & Luna-Jones, 2015). The MMPI-2-RF was described in detail earlier.

Special incident reports (SIRs). The study criteria were prospective count variables derived from SIRs filed by hospital staff. Per hospital policy, SIRs were completed every time a patient engaged in self-harm or violent behavior at the hospital and the specific type of behavior was coded for each SIR-reportable behavior. Counts for future suicidal behavior were derived from SIRs that were completed when a patient threatened or attempted suicide after MMPI-2/-RF testing (BR = 10.7%; M = .4, SD = 1.8). Counts for future violent behavior were derived from SIRs that were completed when a patient was physically or verbally abusive to another patient or staff member after MMPI-2/-RF administration (BR = 33.7%; M = 2.1, SD = 6.0).

Analysis Plan

We ran nonlinear regression models to identify manifest MMPI-2-RF scales that evidenced slope or intercept bias across genders and ethnicities in predicting manifest counts of future suicidal or violent behaviors. These count data were parameterized in Mplus 7.3 (Muthén & Muthén, 1998–2015) using maximum likelihood with robust standard errors and a mean-adjusted chi-square test statistic that are robust to non-normality (MLM). This estimator is intended for use with non-normal criteria in analyses with complete data (Muthén & Muthén, 1998–2015), as was the case in the current samples.

After controlling for hospitalization time by regressing the criteria onto this variable (Little, 2013), we estimated models separately across genders and ethnicities. To examine for slope bias, we constrained the slopes to be equal across demographic groups, then compared the fit of this model with a freely estimated model. To examine for intercept bias, we constrained the intercepts to be equal across demographic groups, then compared the intercept-constrained model with the freely estimated model using the Satorra-Bentler chi-square test (Satorra & Bentler, 1994). We used this chi-square statistic to calculate omega (ω') prime to quantify effect sizes (.10 small, .30 medium, and .50 large; Cohen, 1988).

Of note, in cases where slope bias was evidenced by a statistically significant p value or by an effect size ≥ .10, both the intercepts and slopes were constrained and compared with the slope-constrained model because slope bias inherently produces differences in intercepts. In these cases, we assessed decrement in statistical fit across the models using Satorra-Bentler adjusted chi-square difference tests (Satorra & Bentler, 2001). For all analyses, Satorra-Bentler chi-squares associated with p > .05 indicated nonsignificant decrement in fit when intercepts and/or slopes were constrained to be equal. In other words, significant Satorra-Bentler chi-square values are interpreted as evidence of slope and/or intercept bias.

In cases yielding statistically significant differences in slopes or intercepts across the ethnicities, a series of three Satorra-Bentler chi-square tests was used to identify statistically significant differences across the three ethnic groups (one test for each possible pairing). Follow-up testing was unnecessary for gender, which had two categories in the current study.

Results

Study results are presented in complete detail in Supplementary Tables 2 through 5 and summarized in Table 1. These include

Table 1
Summary of Statistically Significant Analyses

Scale	Demographic	Criterion	Bias	ω'	Effect size	Interpretation
MSF	Gender	Suicide	Slope	.09	Negligible	Men > Women
ANP†	Ethnicity	Suicide	Slope	.10	Small	African American > Hispanic/Latino American
ACT	Ethnicity	Suicide	Slope	.09	Negligible	Caucasian > Hispanic/Latino American
NEGE-r†	Ethnicity	Suicide	Slope	.10	Small	Caucasian > Hispanic/Latino American
MSF	Ethnicity	Suicide	Intercept	.10	Small	African American > Caucasian
RC3	Gender	Violence	Slope	.11	Small	Men > Women
SUI	Gender	Violence	Slope	.12	Small	Men > Women
BRF	Gender	Violence	Slope	.13	Small	Men > Women
DSF	Gender	Violence	Slope	.13	Small	Men > Women
INTR-r	Gender	Violence	Intercept	.13	Small	Men > Women
RC7	Ethnicity	Violence	Slope	.18	Small	African American > Caucasian = Hispanic/Latino American
GIC	Ethnicity	Violence	Slope	.22	Small	African American > Caucasian = Hispanic/Latino American
NUC	Ethnicity	Violence	Slope	.19	Small	African American > Caucasian = Hispanic/Latino American
COG	Ethnicity	Violence	Slope	.15	Small	African American > Caucasian = Hispanic/Latino American
SUI	Ethnicity	Violence	Slope	.25	Small	African American > Caucasian = Hispanic/Latino American
AXY	Ethnicity	Violence	Slope	.18	Small	African American > Caucasian = Hispanic/Latino American
ANP†	Ethnicity	Violence	Slope	.15	Small	African American > Hispanic/Latino American
AGG†	Ethnicity	Violence	Slope	.17	Small	African American > Hispanic/Latino American

Note. Omega prime is used to quantify effect sizes (.10 small, .30 medium, and .50 large; Cohen, 1988). MSF = multiple specific fears; ANP = anger proneness; ACT = activation; NEGE-r = negative emotionality/neuroticism-revised; RC3 = cynicism; SUI = suicidal/death ideation; BRF = behavior restricting fears; DSF = disaffiliativeness; INTR-r = introversion/low positive emotionality-revised; RC7 = dysfunctional negative emotions; GIC = gastro-intestinal complaints; NUC = neurological complaints; COG = cognitive complaints; AXY = anxiety; AGG = aggression. For cases of intercept bias, > (the overall regression line systematically under-predicted the criterion among the prior sample and over-predicted the criterion among the latter sample). For cases of slope bias, > (the associations between the scale score and the criterion were stronger among the prior sample than among the latter sample). Equal sign (=) indicates not significantly different.
† Scales previously found to be statistically significant predictors of the relevant criterion by Tarescavage, Glassmire, and Burchett (2018, 2016).

examinations of slope and intercept bias across the 40 substantive scales of the MMPI-2-RF for the two study criteria (future suicidal behavior and future violent behavior), yielding 320 analyses. Overall, 18 of these analyses yielded statistically significant findings (5.6%), which is consistent with our chance level. Moreover, the effect sizes for the statistically significant results ranged from negligible to small. More specific aspects of these findings are described next.

Supplementary Table 2 presents data from the analyses for differential validity of MMPI-2-RF scales across genders in predicting future suicidal behavior. The multiple specific fears (MSF) scale was the only MMPI-2-RF scale whose scores evidenced statistically significant test bias in predicting this criterion. The MSF scale scores evidenced slope bias to a negligible extent (ω = .09). Overall, there was no evidence that the MMPI-2-RF scales predicted future suicidal behavior unequally across genders outside this one exception.

Data from the analyses for differential validity of MMPI-2-RF scale scores across ethnicities in predicting future suicidal behavior are presented in Supplementary Table 3. Overall, three analyses yielded statistically significant evidence of slope bias, and one evidenced intercept bias. The negative emotionality/neuroticism-revised (NEGE-r) scale evidenced slope bias to a small degree (ω' = .10) across ethnicities in predicting future suicidal behavior. The scale's association with the criterion was slightly stronger among Caucasians than Hispanic/Latino Americans in the current sample. The two other scales evidencing slope bias to a negligible (ω' = .09) and a small (ω' = .10) extent were the activation (ACT) and anger proneness (ANP) scales, respectively. Associations between the ACT scale followed the same trend across ethnicities as the NEGE-r scale, but the ANP scale scores had slightly stronger associations with the criterion among African Americans than Hispanic/Latino Americans in the current sample. The multiple specific fears (MSF) scale scores yielded the only evidence of intercept bias in predicting future suicidal behaviors across ethnicities, and the associated effect size was small (ω' = .10). The overall regression line for the associations between future suicidal behavior and this scale systematically underpredicted these behaviors among the African American sample and overpredicted these behaviors among the Caucasian sample.

Data from the analyses for differential validity of MMPI-2-RF scales across genders in predicting future violent behavior are presented in Supplementary Table 4. Four analyses produced statistically significant evidence of slope bias, each to a small extent, and one scale evidenced intercept bias to a small extent. The disaffiliativeness (DSF) scale was one of the four scales that evidenced slope bias. To a small extent (ω' = .13), the DSF scale was more strongly associated with the criterion among men than among women in the current sample. The other three scales that evidenced slope bias were the RC3 (cynicism), suicidal/death ideation (SUI), and behavior-restricting fears (BRF) scales, and each followed the same trend, having slightly stronger associations among men than women in the current sample. The Introversion/Low Positive Emotionality-Revised (INTR-r) Scale scores yielded the only statistically significant evidence of intercept bias in predicting future violent behaviors across genders. The overall regression line systematically underpredicted violent behaviors among the male sample and overpredicted the criterion among the female sample.

In Supplementary Table 5, data from the analyses for differential validity of MMPI-2-RF scales across ethnicities in predicting future violent behavior are presented. Overall, eight scales evi-

denced slope bias, each to a small extent. The aggression (AGG) scale scores evidenced slope bias in predicting violence across ethnicities to a small extent (ω' = .17). The scale scores were more strongly associated with future violence among the African American sample than among the Hispanic/Latino American sample. The ANP scale scores were also more strongly associated with the criterion among African Americans than Hispanic/Latino Americans in the current sample, whereas the other six scales yielding statistically significant evidence of slope bias in predicting the criterion (the RC7; dysfunctional negative emotions), gastrointestinal complaints, neurological complaints, cognitive complaints, suicidal/death ideation, and anxiety scales) were more strongly associated among African Americans than among both Caucasians and Hispanic/Latino Americans in the current sample. No scale scores evidenced intercept bias in predicting future violent behaviors across ethnicities.

Discussion

The purpose of the current study was to examine MMPI-2-RF scale scores for differential validity across genders (male and female) and ethnicities (Caucasian, African American, and Hispanic/Latino American) in predicting suicidal and violent behaviors among forensic psychiatric inpatients. After conducting 320 analyses, we found 18 cases of differential predictive validity across the MMPI-2-RF scales. Of these, two cases evidenced negligible effect sizes whereas the other 16 evidenced small effect sizes. Differential validity typically manifested in the form of slope bias, not intercept bias, which indicates that the strength of associations differed across demographic groups in most statistically significant analyses. Several aspects of these findings warrant further discussion.

Notably, the most common form of test bias evidenced in the current study differs from the only prior examination of the MMPI-2-RF substantive scale scores for differential predictive validity (Marek et al., 2015). Whereas in our study slope bias constituted all but two instances of differential predictive validity, Marek et al. (2015) found mostly intercept bias, not slope bias, in a sample of bariatric surgery candidates. This difference in trends may have been related to differences in populations (forensic vs. medical) and/or differences in types of criteria (prospective vs. retrospective). Future differential validity research is needed to determine if these trends persist.

Past research by Tarescavage et al. (2016, 2018) examined MMPI-2-RF scale scores to determine which were associated with suicidal and violent behaviors (scales that were meaningful predictors in the past studies are denoted in the supplementary tables). Of note, in the current study 14 cases of test bias were found in scales that were not previously found to be associated with the relevant outcome (future suicidal or violent behaviors) by this research, which utilized the same samples. Thus, the differences in associations between the scales and the relevant outcomes for these 14 cases of test bias (from 18 total) may not be pertinent to for practical use of the test.

Past research by Tarescavage et al. (2018) found that many internalizing and externalizing constructs measured by MMPI-2-RF scales were meaningfully associated with future suicidal behaviors. They found that scales in the emotional/internalizing dysfunction domain, including the ANP and NEGE-r scales, evidenced particularly robust associations with the criterion. The current study found both the ANP and NEGE-r scales evidenced slope bias in the prediction of the criterion across ethnicities to small extents (ω' = .10 for each). Each scale had scores that were more strongly associated with the criterion among the Caucasian group than the Hispanic/Latino American group. However, given the negligible to small effect sizes, the current study suggests that these scales can be interpreted similarly across ethnicities.

Past research by Tarescavage et al. (2016) also found that the AGG and ANP scales were the best predictors of institutional violence. Interestingly, neither of these scale scores evidenced differential validity across genders, but both evidenced slope bias in the prediction of the criterion across ethnicities to a small extent (ω' = .17 and .15, respectively). Each of these scales' scores were more strongly associated with the criterion among African Americans than among Hispanic/Latinos in the current sample. Notwithstanding this trend, here too the results of this study suggest that MMPI-2-RF scores can generally be interpreted similarly across demographic groups due to uniformly small effect sizes among scales demonstrating differential validity in the current sample.

The primary limitation of this study was insufficient sample sizes to investigate differential predictive validity across other demographic groups, such as Asian Americans, Native Americans, Alaska Natives, and multiracial Americans. It is particularly important that future studies explore this issue among Native Americans given the high rates of suicide in this population (Alcántara & Gone, 2007). Future research should examine the MMPI-2-RF scales for differential validity across more diverse samples as well as other cultures in countries where this test is used (e.g., European vs. indigenous descent in various commonwealth countries), as well as in translated versions. Similarly, future investigations of differential validity of MMPI-2-RF test scores among individuals who do not identify as male or female are of interest. Future research with other constructs of interest in a forensic setting (e.g., differential utility of validity scales in malingering) are also indicated.

Notwithstanding these limitations, this study indicates that the MMPI-2-RF does not meaningfully evidence differential predictive validity in forensic populations of men and women, as well as individuals from Caucasian, African American, and Hispanic/Latino American ethnicities in the prediction of future suicidal or violent behaviors. This research extends support of the MMPI-2-RF as an objective and generally unbiased component of suicide and violence risk assessment in forensic settings.

References

Alcántara, C., & Gone, J. P. (2007). Reviewing suicide in Native American communities: Situating risk and protective factors within a transactional-ecological framework. *Death Studies, 31,* 457–477. http://dx.doi.org/10.1080/07481180701244587

American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders* (4th ed., text revision). Washington, DC: Author.

American Psychology-Law Society and American Academy of Forensic Psychology. (2013). Specialty guidelines for forensic psychology. *American Psychologist, 68,* 7–19. http://dx.doi.org/10.1037/a0029889

Arbisi, P. A., Ben-Porath, Y. S., & McNulty, J. (2002). A comparison of MMPI-2 validity in African American and Caucasian psychiatric inpatients. *Psychological Assessment, 14,* 3–15. http://dx.doi.org/10.1037/1040-3590.14.1.3

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,* 84–94. http://dx.doi.org/10.1207/s15327752jpa8701_07

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

Borum, R. (1996). Improving the clinical practice of violence risk assessment. Technology, guidelines, and training. *American Psychologist, 51,* 945–956. http://dx.doi.org/10.1037/0003-066X.51.9.945

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

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Correia, K. M. (2000). Suicide assessment in a prison environment: A proposed protocol. *Criminal Justice and Behavior, 27,* 581–599. http://dx.doi.org/10.1177/0093854800027005003

Crosby, A. E., Han, B., Ortega, L. A. G., Parks, S. E., & Gfroerer, J. (2011). Suicidal thoughts and behaviors among adults aged ≥ 18 years—United States, 2008–2009. *Centers for Disease Control and Prevention: Surveillance Summaries, 60,* 1–22.

Douglas, K. S., Ogloff, J. R. P., Nicholls, T. L., & Grant, I. (1999). Assessing risk for violence among psychiatric patients: The HCR-20 violence risk assessment scheme and the Psychopathy Checklist: Screening version. *Journal of Consulting and Clinical Psychology, 67,* 917–930. http://dx.doi.org/10.1037/0022-006X.67.6.917

Glassmire, D. M., Tarescavage, A. M., Burchett, D., Martinez, J., & Gomez, A. (2016). Clinical utility of the MMPI-2-RF SUI items and scale in a forensic inpatient setting: Association with interview self-report and future suicidal behaviors. *Psychological Assessment, 28,* 1502–1509. http://dx.doi.org/10.1037/pas0000220

Kaplan, R. M., & Saccuzzo, D. P. (2018). *Psychological testing: Principles, applications, and issues*. Boston, MA: Cengage Learning.

Little, T. D. (2013). *Longitudinal structural equation modeling*. New York, NY: Guilford Press.

Marek, R. J., Ben-Porath, Y. S., Sellbom, M., McNulty, J. L., & Heinberg, L. J. (2015). Validity of Minnesota Multiphasic Personality Inventory-2-Restructured Form (MMPI-2-RF) scores as a function of gender, ethnicity, and age of bariatric surgery candidates. *Surgery for Obesity and Related Diseases, 11,* 627–634. http://dx.doi.org/10.1016/j.soard.2014.10.005

Metzner, J. L. (2002). Class action litigation in correctional psychiatry. *The Journal of the American Academy of Psychiatry and the Law, 30,* 19–29. Retrieved from http://jaapl.org/content/jaapl/30/1/19.full.pdf

Monnot, M. J., Quirk, S. W., Hoerger, M., & Brewer, L. (2009). Racial bias in personality assessment: Using the MMPI-2 to predict psychiatric diagnoses of African American and Caucasian chemical dependency inpatients. *Psychological Assessment, 21,* 137–151. http://dx.doi.org/10.1037/a0015316

Muthén, L. K., & Muthén, B. O. (1998–2015). Mplus 7.3 [Computer software]. Los Angeles, CA: Author.

Reynolds, C. R., & Ramsay, M. C. (2003). Bias in psychological assessment: An empirical review and recommendations. In I. B. Weiner, J. R. Graham, & J. A. Naglieri (Eds.), *Handbook of psychology* (Vol. 10, pp. 67–93). Hoboken, NJ: Wiley.

Satorra, A., & Bentler, P. M. (1994). Corrections to test statistics and standard errors in covariance structure analyses. In A. von Eye & C. C. Clogg (Eds.), *Latent variable analysis: Applications for developmental research* (pp. 399–419). Thousand Oaks, CA: Sage.

Satorra, A., & Bentler, P. M. (2001). A scaled difference chi-square test statistic for moment structure analysis. *Psychometrika, 66,* 507–514. http://dx.doi.org/10.1007/BF02296192

Sellbom, M. (in press). The MMPI-2 Restructured Form (MMPI-2-RF): Assessment of personality and psychopathology in the 21st Century. *Annual Review of Clinical Psychology*.

Skeem, J. L., & Monahan, J. (2011). Current directions in violence risk assessment. *Current Directions in Psychological Science, 20,* 38–42. http://dx.doi.org/10.1177/0963721410397271

Smith, S. G., Basile, K. C., Gilbert, L. K., Merrick, M. T., Patel, N., Walling, M., & Jain, A. (2017). *The national intimate partner violence survey (NISVS): 2010–2012 state report*. Atlanta, GA: National Center for Injury Prevention and Control, Centers for Disease Control and Prevention. Retrieved from https://www.cdc.gov/violenceprevention/pdf/NISVS-StateReportBook.pdf

Sue, D. W., Capodilupo, C. M., Torino, G. C., Bucceri, J. M., Holder, A. M. B., Nadal, K. L., & Esquilin, M. (2007). Racial microaggressions in everyday life: Implications for clinical practice. *American Psychologist, 62,* 271–286. http://dx.doi.org/10.1037/0003-066X.62.4.271

Tarescavage, A. M., Alosco, M. L., Ben-Porath, Y. S., Wood, A., & Luna-Jones, L. (2015). Minnesota Multiphasic Personality Inventory-2-Restructured Form (MMPI-2-RF) scores generated from the MMPI-2 and MMPI-2-RF test booklets: Internal structure comparability in a sample of criminal defendants. *Psychological Assessment, 22,* 188–197. http://dx.doi.org/10.1177/1073191114537347

Tarescavage, A. M., Glassmire, D. M., & Burchett, D. (2016). Introduction of a conceptual model for integrating the MMPI-2-RF into HCR-20^{V3} violence risk assessments and associations between the MMPI-2-RF and institutional violence. *Law and Human Behavior, 40,* 626–637. http://dx.doi.org/10.1037/lhb0000207

Tarescavage, A. M., Glassmire, D. M., & Burchett, D. (2018). Minnesota Multiphasic Personality Inventory–2–restructured form markers of future suicidal behavior in a forensic psychiatric hospital. *Psychological Assessment, 30,* 170–178. http://dx.doi.org/10.1037/pas0000463

Wolff, N., Blitz, C. L., Siegel, J., & Bachman, R. (2007). Physical violence inside prisons: Rates of victimization. *Criminal Justice and Behavior, 34,* 588–599. http://dx.doi.org/10.1177/0093854806296830