

The MMP-2-RF and College Students: Do We Remain Stuck in a Normative No-man's Land?

P. Kevin Bolinskey, Ph.D., Janice P. Guidi, M.S., Kevin R. Myers, M.S., Dianna Cooper-Bolinskey, M.S.W., Kelly M. Schuder, M.S., Alison V. James, Psy.D., Daniel V. Hudak, Psy.D., Yelena Gonzalez, M.S., Erin McTiernan, M.S., and Elizabeth A. Smith, M.S.

Abstract

Sixty-plus years of research have demonstrated that college students tend to score significantly higher than older adults on the MMPI and MMPI-2. Relatively little research has examined whether this pattern of results holds with the newer MMPI-2-RF. The present study compared MMPI-2-RF scores of college students, both by gender and combined, to a reference group composed of equal numbers of men and women from the MMPI-2 normative sample. Results demonstrated that this pattern of higher scores among college students remains a concern when using the MMPI-2-RF. Of particular concern is the finding that women's scores appear to be more affected by the use of non-gendered norms on the MMPI-2-RF than do the scores of men. Caution is urged when interpreting MMPI-2-RF results obtained from college students, particularly women.

Introduction

It has long been known that college students may respond to assessment items differently than do older adults. Hathaway and McKinley, the authors of the original MMPI, were aware of this fact, as well. After finding that several preliminary MMPI items which discriminated healthy adults from psychiatric samples also discriminated healthy adults from healthy college students, they employed a healthy college and pre-college sample as a control group in item selection to help mitigate this problem (Brown, 1948).

Despite these efforts of Hathaway and McKinley, even early studies using the MMPI with college students demonstrated that they produced significantly different MMPI T scores than the normative group across a variety of scales (cf. Brown, 1948; Gilliland & Colgin, 1951; Sopchak, 1952; Tyler & Michaelis, 1953). Gilliland and Colgin (1951), for example, found that the average MMPI T scores for college students consistently ranged from 55-64, with the only exception being T scores below 50 for male college students on scale 5. They suggested that the normative scores for the MMPI were not appropriate for college students and might lead to high false-positive rates for psychopathology and/or abnormal personality.

The first specific suggestion that separate norms were needed for use with college students appears to have been made by Sopchak in 1952 who found that male college students averaged T scores above 50 on all clinical scales. Although women participants' scores in this

study were closer to the standardization group than were those of college men, their scores averaged more than a full standard deviation higher than the normative group on scales 4 and 0.

Of course, it is well known that the original MMPI norms tended to overpathologize adults by an average of around 5 T points (cf. Pancoast & Archer, 1989), which was one factor, among others, that led to the creation of the MMPI-2. Although one might be tempted to view the above-noted differences as artifacts of the deficiencies of the original norms, discrepancies between contemporary college student and adult samples suggest that these differences may reflect developmental stage variations. Although not specifically related to college students, this realization of developmental processes resulting in MMPI differences is reflected in the decisions of Marks and Briggs (1972) and Gottesman, Hanson, Kroeker, and Briggs (1987) to create multiple sets of norms, based on age, for the MMPI in some of the earlier attempts to create adolescent norms. Such norms have failed to gain popularity in clinical application, resulting in continued use of adult norms which may fail to accurately describe young adults, including traditional-aged college students.

Pancoast and Archer (1992) characterized college student MMPI profiles in terms of developmental processes, noting that college students' MMPI scores fell in between those of adolescents and adults on scales *F*, 4, 7, 8, and 9, although the difference between college students and adults on scale *F* was observed only in women¹. Male college students scored an average of more than 3 T points higher than adult men on scales 4 and 5, more than 5 T points higher on scale 7, and more than six T points higher on scales 8 and 9. As compared to their adult counterparts, female college students scored an average of more than 3 T points higher on scales *F* and 7, more than 4 T points higher on scales 4 and 8, and more than six T points higher on scale 9.

The MMPI-2 and college students

Similar differences have been found between college-aged individuals and adults on the MMPI-2. Butcher, Graham, Dahlstrom, and Bowman (1990), for example, compared mean MMPI-2 basic validity and clinical scale raw scores of 1,312 college students to the MMPI-2 normative sample. They found that both male and female college students scored significantly higher than their respective normative groups on scales *F*, *K*, 2, 4, 6, 7, 8, 9, and 0; in addition, female participants also scored significantly higher than the female MMPI-2 normative sample on scales *L*, *I*, and 5. When T scores were compared, however, college students evidenced significantly and meaningfully higher T scores only on scales 7, 8, and 9 compared to the normative groups.

Osberg, Haseley, and Kamas (2008) likewise reported mean T scores at least 5 points above 50 for college freshmen on clinical scales 4, 7, 8, and 9. Among the validity scales, the college population evidenced T scores at least 5 points above 50 on *F*, *Fb*, and *Fp*.

Strassberg (1997) found that college students from the United States scored significantly higher on scales *F*, 4, 8, and 9 compared to the MMPI-2 normative sample. Further, in a cross-temporal meta-analysis Twenge et al. (2010) reported a trend of higher MMPI and MMPI-2 T scores among college students over time and noted that current college students scored, on average, about one standard deviation higher on the clinical scales of the MMPI-2. To further

¹ Pancoast and Archer (1992) included scale 6 among their list of scales on which college students scored in between adolescents and adults. However, the magnitude of the difference between college students and adults was only .4 T points in men and 1.5 T points in women.

complicate matters, Osberg and Poland (2002) noted that while the MMPI-2 tended to overpathologize 18 year-old college students, the MMPI-A tended to underpathologize them.

RC Scales/MMPI-2-RF

Tellegen and Ben-Porath (2011) report mean MMPI-2-RF T scores for 1,227 college men and 1,989 college women in the MMPI-2-RF technical manual; college men scored at least four T points higher than 50 on 21 of 51 scales, whereas college women scored at least four T points higher than 50 on 26 of 51 scales. Overall, however, there has been a dearth of information concerning average RC and/or MMPI-2-RF scale scores among college students, as several of the available studies (e.g., Forbey & Ben-Porath, 2007; 2008) do not report mean scores. Osberg et al. (2008), provided mean score information for a large sample of undergraduate students and found that college students' mean scores were at least four T-points higher than 50 on four of RC scales (*RC1*, *RC6*, *RC8*, and *RC9*) and more than 3.9 T-points higher on two others (*RCd* and *RC3*). It should be noted that Osberg et al. incorporated gender-based T scores from the MMPI-2. In a study incorporating non-gender-based T scores from the MMPI-2-RF, Hunter et al. (2014) observed a similar pattern of scores in their nonclinical sample, as each of these scales' mean scores were more than 5 T points above 50. In addition, they also found that their normal sample scored substantially above 50 on *RC7*. Among the remainder of the MMPI-2-RF scales, Hunter et al. reported mean T scores more than 5 points above 50 for a majority of the scales.

Non-gendered T scores on the MMPI-2-RF

As noted above, the MMPI-2-RF includes only non-gendered norms. The use of non-gendered norms complies with the provision of the Civil Rights Act of 1991 prohibiting the consideration of sex in employment practices. Some (e.g., Butcher & Williams, 2012; Nichols, 2011) have criticized the decision to use non-gendered norms, based on decades of gender-based psychological research. A comparison of gender-specific versus non-gendered norms for the MMPI-2 scales conducted by Ben-Porath and Forbey (2003) revealed a dearth of significant differences among adults, with most T score differences falling within a range of three points. However, little research has been conducted to examine what differential impact, if any, exists when non-gendered norms are used with college students.

Present Study

The present study has two goals:

- 1) To compare the MMPI-2-RF scale scores of traditionally aged college students to a comparison sample of adults that are representative of the MMPI-2-RF normative sample.
- 2) To examine whether the use of non-gendered T scores has a differential impact on the relative elevations of T scores for college men and women.

Methods

Participants. Our criterion sample included 1,329 (411 men, 918 women) college students between the ages of 18 and 22 years ($M = 19.0$, $SD = 0.98$) who received course credit for their participation in a larger, ongoing study of psychosis proneness. Men ($M = 19.2$, $SD = 1.1$) averaged approximately three months older than women ($M = 18.9$, $SD = .93$); this age difference was not significant.

For inclusion in the larger study's final sample, participants' responses had to meet the following MMPI-2-RF validity criteria: *VRIN-r* T score < 80, *TRIN-r* T score < 80, *F-r* T score < 111, *Fp-r* T score < 100, *L-r* T score \leq 81, and omitted items \leq 10. Individuals who had previously been identified as psychosis prone were not included in the current study; there were no exclusions for any other physical or mental health issues.

The reference group was made up of 2,276 individuals from the MMPI-2 normative sample (Butcher et al., 1989). In order to create a proxy of the MMPI-2-RF normative sample, data from all 1,138 men and data from a randomly selected subset of 1,138 women were used. Participants in this sample were between the ages of 18 and 84. The average age of male participants in the proxy normative sample was 41.7 years old ($SD=15.3$) and the mean age of female participants was 40.6 years old ($SD=15.2$).

Measure. The Minnesota Multiphasic Personality Inventory-2-Restructured Form (MMPI-2-RF; Ben Porath & Tellegen, 2011) is a 338-item self-report measure designed to assess an array of clinical behaviors and conditions. Items are endorsed in either a "True" or "False" direction. The MMPI-2-RF is composed of nine validity scales, three higher order scales, nine restructured clinical scales, five somatic and cognitive scales, nine internalizing scales, four externalizing scales, five interpersonal scales, two interest scales, and five PSY-5 scales (Tellegen & Ben-Porath, 2011).

Procedures. Participants completed the full MMPI-2 form. It has been demonstrated that MMPI-2-RF scale scores obtained from an MMPI-2 administration are comparable to those obtained with the MMPI-2-RF booklet (Tellegen & Ben-Porath, 2011; van der Heijden, Egger, & Derksen, 2010); thus, MMPI-2-RF scores were derived from the full set of MMPI-2 item responses.

Two series of separate one-way analyses of variance (ANOVAs) were conducted for each of the MMPI-2-RF scales using group membership as the independent variable. In the first series of analyses, male and female college students were combined into a single group and compared to the reference group without regard to gender. In the second group of analyses, male and female students were compared to the reference group separately. Note that the reference sample for these comparisons included both men and women, as the normative samples for the MMPI-2-RF includes both genders. We are interested in how well the norms of the MMPI-2-RF work with male and college students independently, but are not necessarily interested in how college men compare to adult men in our reference sample. A Bonferroni adjustment of $p < .001$ was incorporated into significance estimations to account for the 51 separate analyses in each series. Effect sizes (Cohen's d) were calculated for each comparison to estimate the magnitude of differences. Effects greater than .40, which represent, on average, a difference of four or more T points, were considered to be clinically significant.

Results

Non-gendered comparisons. Results from the series of one-way ANOVAs are shown in Table 1. Even after application of the Bonferroni correction of $p < .001$, there were significant differences in mean scores between the non-gendered college sample and reference group on 40 of the 51 MMPI-2-RF scales. Unless otherwise noted, the term *significant* refers to statistical significance.

As can be seen in Table 1, the college sample scored significantly higher than the reference sample on seven of the nine validity scales. Five of these differences were judged to be clinically significant with medium effect sizes. On both *Fp-r* ($d = 0.69$) and *F-s* ($d = 0.67$)

Table 1
Means and standard deviations for the MMPI-2-RF scales by sample, with associated *F* values and effect sizes

	Reference Group		All College		College Men		College Women		<i>F</i> ₁	<i>d</i> ₁	<i>F</i> ₂	<i>d</i> ₂	<i>F</i> ₃	<i>d</i> ₃
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>						
<i>VRIN-r</i>	51.5	10.1	53.6	10.2	52.8	10.5	53.9	10.0	36.61*	0.21	5.93	0.13	39.40*	0.25
<i>TRIN-r</i>	50.4	7.4	50.5	6.5	50.6	6.8	50.5	6.4	0.17	0.01	0.26	0.03	0.13	0.01
<i>F-r</i>	49.8	10.1	55.8	13.7	54.4	13.0	56.4	13.9	227.97*	0.52	65.50*	0.43	226.79*	0.59
<i>Fp-r</i>	49.8	10.1	57.3	12.2	55.6	12.4	58.1	12.1	397.99*	0.69	104.67*	0.55	396.67*	0.78
<i>Fs</i>	49.9	9.7	57.4	13.3	55.4	12.6	58.3	13.6	374.59*	0.67	99.89*	0.54	380.52*	0.76
<i>FBS-r</i>	49.9	9.9	52.6	11.8	47.9	10.7	54.7	11.6	53.82*	0.25	14.45*	0.20	139.54*	0.46
<i>RBS</i>	49.9	10.0	55.3	12.5	53.6	11.2	56.1	13.0	202.52*	0.49	44.13*	0.36	209.54*	0.57
<i>L-r</i>	50.1	10.0	50.2	8.8	51.8	9.1	49.5	8.6	0.15	0.01	10.42	0.17	2.37	0.06
<i>K-r</i>	50.0	9.9	45.2	9.4	47.8	9.3	44.0	9.2	205.17*	0.49	16.93*	0.22	248.15*	0.62
<i>EID</i>	49.9	9.8	53.4	11.2	49.4	10.7	55.3	10.9	100.14*	0.35	0.96	0.05	185.39*	0.53
<i>THD</i>	50.2	9.9	54.6	11.0	54.3	11.3	54.7	10.9	151.68*	0.43	57.96*	0.41	127.54*	0.44
<i>BXD</i>	50.1	9.9	51.2	8.9	53.9	9.0	50.0	8.6	11.97*	0.12	53.79*	0.39	0.04	0.01
<i>RCD</i>	49.9	9.7	55.3	10.4	52.5	10.2	56.6	10.2	254.14*	0.55	25.06*	0.27	308.51*	0.69
<i>RC1</i>	50.1	10.0	56.5	11.2	52.1	10.6	58.5	11.0	314.69*	0.61	13.85*	0.20	434.49*	0.81
<i>RC2</i>	49.7	10.0	50.3	10.9	48.3	10.9	51.3	10.8	3.42	0.06	6.82	0.14	15.82*	0.16
<i>RC3</i>	50.1	10.0	56.1	9.6	55.2	9.5	56.5	9.6	308.10*	0.61	91.50*	0.51	270.44*	0.64
<i>RC4</i>	50.0	9.9	51.3	9.0	52.3	9.3	50.8	8.8	14.27*	0.13	18.91*	0.23	4.31	0.08
<i>RC6</i>	50.3	9.9	56.9	11.4	55.9	11.3	57.4	11.4	333.86*	0.63	106.27*	0.55	304.58*	0.68
<i>RC7</i>	50.0	9.7	55.9	11.8	51.6	11.0	57.9	11.7	270.03*	0.57	9.46	0.16	387.17*	0.77
<i>RC8</i>	50.1	9.9	56.1	11.9	56.1	12.2	56.1	11.7	265.42*	0.56	119.14*	0.58	214.86*	0.57
<i>RC9</i>	50.1	10.1	54.9	10.7	56.0	11.0	54.3	10.4	182.57*	0.47	117.04*	0.58	116.18*	0.42
<i>MLS</i>	49.8	9.8	53.3	9.7	50.2	8.6	54.7	9.8	107.42*	0.36	0.57	0.04	162.71*	0.50
<i>GIC</i>	50.4	9.8	53.6	12.4	51.3	10.6	54.7	13.1	73.74*	0.30	2.66	0.09	101.77*	0.39
<i>HPC</i>	50.2	9.9	55.0	11.7	50.6	9.8	56.9	12.0	166.98*	0.45	0.41	0.03	264.41*	0.64
<i>NUC</i>	50.1	10.0	57.5	11.4	55.6	11.2	58.3	11.4	413.34*	0.70	101.83*	0.54	408.83*	0.79
<i>COG</i>	49.8	9.9	57.6	12.3	55.3	11.7	58.6	12.4	434.85*	0.72	101.77*	0.54	447.46*	0.83

<i>SUI</i>	49.7	10.1	49.8	11.3	49.6	11.3	49.8	11.3	0.00	0.00	0.05	0.01	0.04	0.01
<i>HLP</i>	49.8	10.1	50.4	10.9	48.9	10.8	51.1	10.8	3.18	0.06	2.76	0.09	11.05	0.13
<i>SFD</i>	50.0	9.9	55.1	12.5	51.5	11.3	56.7	12.7	178.28*	0.46	6.88	0.14	251.24*	0.62
<i>NFC</i>	50.0	9.9	55.5	10.9	51.8	9.9	57.2	11.0	241.65*	0.54	10.91	0.18	326.03*	0.71
<i>STW</i>	49.8	9.7	54.4	11.1	50.8	10.3	56.1	11.0	168.89*	0.45	3.53	0.10	246.52*	0.61
<i>AXY</i>	49.8	9.7	58.3	13.8	53.5	12.4	60.4	13.8	467.28*	0.75	45.65*	0.36	608.06*	0.96
<i>ANP</i>	50.1	9.8	53.8	11.4	50.5	10.5	55.3	11.4	106.81*	0.36	0.65	0.04	165.94*	0.50
<i>BRF</i>	50.2	9.8	56.3	12.8	51.0	10.5	58.6	13.0	254.88*	0.55	2.01	0.08	397.41*	0.78
<i>MSF</i>	50.1	9.8	49.5	8.2	45.1	6.6	51.4	8.0	4.43	0.07	100.02*	0.54	12.35*	0.14
<i>JCP</i>	50.0	10.1	47.9	8.7	49.6	9.7	47.2	8.1	40.37*	0.22	0.73	0.05	57.88*	0.30
<i>SUB</i>	50.1	10.0	51.2	10.7	52.6	10.7	50.6	10.7	10.50	0.11	21.27*	0.25	1.90	0.05
<i>AGG</i>	49.8	10.0	51.3	10.7	52.5	10.9	50.8	10.6	17.69*	0.15	24.60*	0.27	5.83	0.09
<i>ACT</i>	50.1	9.9	56.5	12.4	53.7	12.5	57.8	12.1	291.35*	0.59	41.49*	0.35	343.54*	0.72
<i>FML</i>	49.9	9.9	53.1	10.7	50.8	9.7	54.1	11.0	80.08*	0.31	2.87	0.09	108.29*	0.41
<i>IPP</i>	49.8	9.9	48.3	9.3	46.2	8.8	49.1	9.4	20.35*	0.16	45.50*	0.36	2.64	0.06
<i>SAV</i>	49.9	10.0	47.2	10.4	48.4	9.9	46.7	10.5	58.56*	0.26	7.61	0.15	65.78*	0.32
<i>SHY</i>	49.8	9.9	51.6	10.7	49.8	9.8	52.4	10.9	26.09*	0.18	0.01	0.01	43.20*	0.26
<i>DSF</i>	49.9	10.1	52.2	10.8	53.1	11.4	51.8	10.5	41.39*	0.22	33.09*	0.31	22.65*	0.19
<i>AES</i>	49.8	10.1	43.3	8.8	41.9	8.4	43.9	9.0	386.49*	0.68	225.60*	0.81	239.35*	0.60
<i>MEC</i>	50.1	10.0	47.9	8.8	54.5	9.7	44.9	6.5	46.99*	0.24	66.66*	0.44	214.62*	0.57
<i>AGGR-r</i>	49.0	9.1	50.4	10.0	53.3	10.4	49.1	9.6	18.88*	0.15	74.19*	0.46	0.13	0.01
<i>PSYC-r</i>	49.9	10.1	54.6	11.1	54.2	11.1	54.9	11.0	170.29*	0.45	59.35*	0.41	148.04*	0.48
<i>DISC-r</i>	49.9	10.1	51.1	8.8	55.8	9.1	49.0	7.8	13.51*	0.13	124.20*	0.60	5.65	0.09
<i>NEGE-r</i>	49.9	9.9	55.8	11.6	51.3	10.6	57.8	11.4	266.46*	0.56	7.14	0.14	387.18*	0.77
<i>INTR-r</i>	49.9	9.9	46.4	10.1	47.2	10.1	46.0	10.1	106.71*	0.36	26.87*	0.28	101.19*	0.39

Note: MC = matched comparison sample. SZT = schizotypy sample. *N* = 105 for each group. *VRIN-r* = Variable Response Inconsistency. *TRIN-r* = True Response Inconsistency. *F-r* = Infrequent Responses. *Fp-r* = Infrequent Psychopathology Responses. *Fs* = Infrequent Somatic Responses. *FBS-r* = Symptom Validity. *RBS* = Response Bias. *L-r* = Uncommon Virtues. *K-r* = Adjustment Validity. *EID* = Emotional/Internalizing Dysfunction. *THD* = Thought Dysfunction. *BXD* = Behavioral/Externalizing Dysfunction. *RCd* = Demoralization. *RC1* = Somatic Complaints. *RC2* = Low Positive Emotionality. *RC3* = Cynicism. *RC4* = Antisocial Behavior. *RC6* = Ideas of Persecution. *RC7* = Dysfunctional Negative Emotions. *RC8* = Aberrant Experiences. *RC9* = Hypomanic Activation. *GIC* = Gastrointestinal Complaints. *HPC* = Head Pain Complaints. *NUC* = Neurological Complaints. *COG* = Cognitive Complaints. *SUI* = Suicidal/Death Ideation. *HLP* = Helplessness/Hopelessness. *SFD* = Self-Doubt. *NFC* = Inefficacy. *STW* = Stress/Worry. *AXY* = Anxiety. *ANP* = Anger Proneness. *BRF* = Behavior-Restricting Fears. *MSF* = Multiple Specific Fears. *JCP* = Juvenile Conduct Problems. *SUB* = Substance Abuse. *AGG* = Aggression. *ACT* = Activation. *FML* = Family Problems. *IPP* = Interpersonal Passivity. *SAV* = Social Avoidance. *SHY* = Shyness. *DSF* = Disaffiliativeness. *AES* = Aesthetic Interests. *MEC* = Mechanical Interests. *AGGR-r* = Aggressiveness PSY-5. *PSYC-r* = Psychoticism PSY-5. *NEGE-r* = Negative Emotionality/Neuroticism PSY-5. *INTR-r* = Introversion PSY-5. F_1 and d_1 refer to comparisons between the reference sample and the combined (men and women) college sample; F_2 and d_2 refer to comparisons between the reference sample and college men; F_3 and d_3 refer to comparisons between the reference sample and college women. d = Cohen's d ; * = $p < .001$. Effect sizes > 0.40 are in boldface

college students averaged scores more than 7 T points higher than our reference group, whereas the differences were in the range of 4.8 to 6.0 T points for *F-r* ($d = 0.52$), *RBS* ($d = 0.49$), and *K-r* ($d = 0.49$). *VRIN-r*, however, evidenced a small effect of $d = 0.21$; this difference of 2.1 T points was judged to not be clinically relevant.

Among the Higher Order scales, the college sample scored significantly higher than the reference sample on all scales. Only on *THD*, however, was the difference judged to be clinically significant, as the mean score of the college sample exceeded that of the reference group by more than four T points with a medium effect size ($d = 0.43$).

Among the RC scales, the college sample scored significantly higher than the reference sample on eight of the nine scales. Medium effect sizes (see Table 1) were observed for each significant comparison, and each statistically significant difference was deemed to be clinically relevant, as they ranged from 4.8 to 6.6 T points.

The college sample scored higher than the reference sample on each of the five Somatic/Cognitive scales. *NUC* ($d = 0.70$) and *COG* ($d = 0.72$) evidenced large between-groups effect sizes, whereas *HPC* evidenced a medium ($d = 0.45$) between-groups effect size. Each of these statistical differences fell in the clinically significant range. The differences for *MLS* and *GIC* ($d = 0.36$ and 0.30 , respectively), while statistically significant, did not meet our criterion for clinical relevance of a difference of 4 or more T score points.

On the Internalizing scales, the college sample scored significantly higher than the reference sample on six of the nine scales. A large effect was observed for the between-groups comparisons on *AXY* ($d = 0.75$), with a mean difference of 8.5 T points. Medium effect sizes were observed for *SFD* ($d = 0.46$), *NFC* ($d = 0.54$), *STW* ($d = 0.45$), and *BRF* ($d = 0.55$). These differences ranged from 4.6 to 6.1 T points and were considered to be clinically significant. The magnitude of the difference on *ANP* demonstrated a small effect size ($d = 0.36$) of 2.7 T points; this difference is not considered clinically relevant.

There were significant differences in mean scores between the college and reference samples on three of the Externalizing scales. College students scored 5.4 T points ($d = 0.59$) higher on *ACT* than did the reference sample; this is a clinically meaningful difference. The differences on *JCP* and *AGG*, however, reflected small ($d = 0.22$ and 0.15 , respectively) effects which ranged from 1.5 to 2.1 T points and were not considered to be clinically relevant.

The college sample scored higher than the reference sample on each of the five Interpersonal scales. The effects for each of the comparisons, however, were small and not of clinical relevance.

College students scored lower than the reference sample on both of the Interest scales. Only *AES* ($d = 0.68$), on which college students averaged scoring 6.5 T points below our reference group, evidenced a clinically relevant difference.

Finally, significant differences were observed between the college and reference samples on three of the PSY-5 scales. Both *PSYC-r* ($d = 0.45$) and *NEGE-r* ($d = 0.56$) evidenced medium effect sizes and clinically relevant T score differences. *AGGR-r*, *DISC-r*, and *INTR-r*, evidenced small effect sizes that fell below the clinically relevant range.

Comparisons by gender. Results for all ANOVAs are shown in Table 1. College men scored significantly higher than the overall reference sample on six of the nine validity scales; see Figure 1 for a graphical representation of each group’s scores. Clinically significant differences were observed for the between-groups comparison on *F-r*, *Fp-r* and *Fs*, as each demonstrated medium effect sizes ($d = 0.43, 0.55, \text{ and } 0.54$ respectively). College women scored significantly different from the overall reference sample on seven of the nine validity scales, with clinically relevant differences from the reference sample on five of those scales. Large effects were observed for *Fp-r* ($d = 0.78$) and *Fs* ($d = 0.76$); these reflect differences of 8.3 and 6.6 T points, respectively. Medium effect sizes, reflecting T score differences of 4.8 to 6.6, were seen on *F-r* ($d = 0.59$), *FBS-r* ($d = 0.46$), *RBS* ($d = 0.57$), and *K-r* ($d = 0.62$). The difference between college women and the reference group on *VRIN-r* was of small magnitude ($d = 0.25$) and of little clinical relevance.

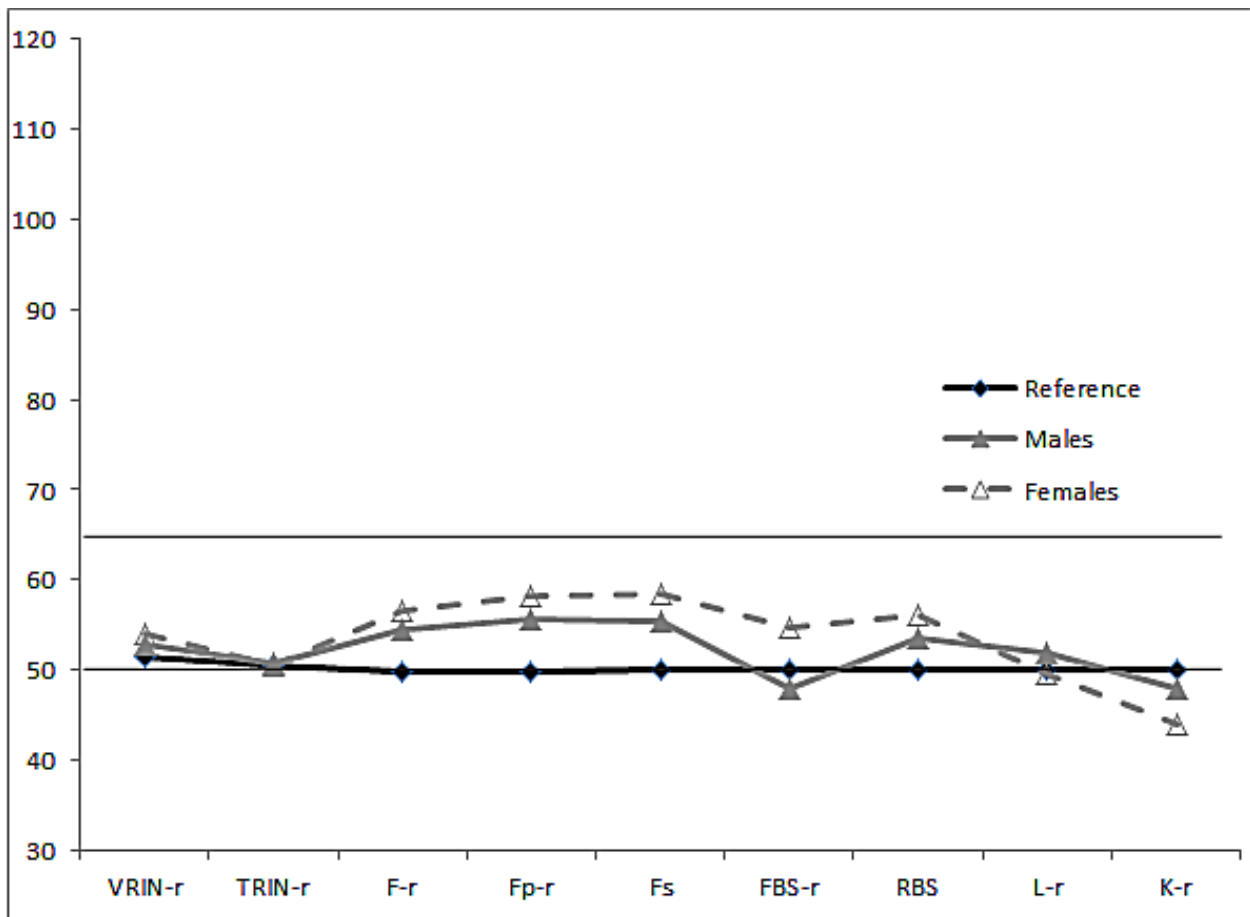


Figure 1: MMPI-2-RF Validity Scale scores by group. *VRIN-r* = Variable Response Inconsistency. *TRIN-r* = True Response Inconsistency. *F-r* = Infrequent Responses. *Fp-r* = Infrequent Psychopathology Responses. *Fs* = Infrequent Somatic Responses. *FBS-r* = Symptom Validity. *RBS* = Response Bias. *L-r* = Uncommon Virtues. *K-r* = Adjustment Validity.

Among the higher order scales (see Figure 2), college men scored significantly higher than the reference sample on *THD* and *BXD*. The difference on *THD* was of medium magnitude ($d = 0.41$) and of clinical relevance; the observed difference on *BXD* approached a medium

effect. College women scored significantly higher than the reference group on *EID* and *THD* ($d = 0.53$ and 0.44 , respectively), demonstrating medium effects and clinically relevant differences.

College men scored significantly higher than the overall reference sample on six of the nine RC scales. Clinically significant differences, ranging from 5.1 to 6.0 T points, were found for college men on four of the nine RC scales, with medium effects observed for *RC3* ($d = 0.51$), *RC6* ($d = 0.55$), *RC8* ($d = 0.58$), and *RC9* ($d = 0.58$). College women scored significantly higher than the overall reference sample on eight of the nine RC scales, however. Clinically relevant differences were found for seven of the nine scales. Large effects were noted for *RC1* and *RC7* ($d = 0.81$ and 0.77 , respectively), whereas medium effects were noted for *RCd* ($d = 0.69$), *RC3* ($d = 0.64$), *RC6* ($d = 0.68$), *RC8* ($d = 0.57$) and *RC9* ($d = 0.42$). Each of these differences was deemed to be clinically relevant, as they ranged from 4.2 to 8.4 T points. *RC2* demonstrated a small ($d = 0.16$) and clinically irrelevant difference (see Figure 2).

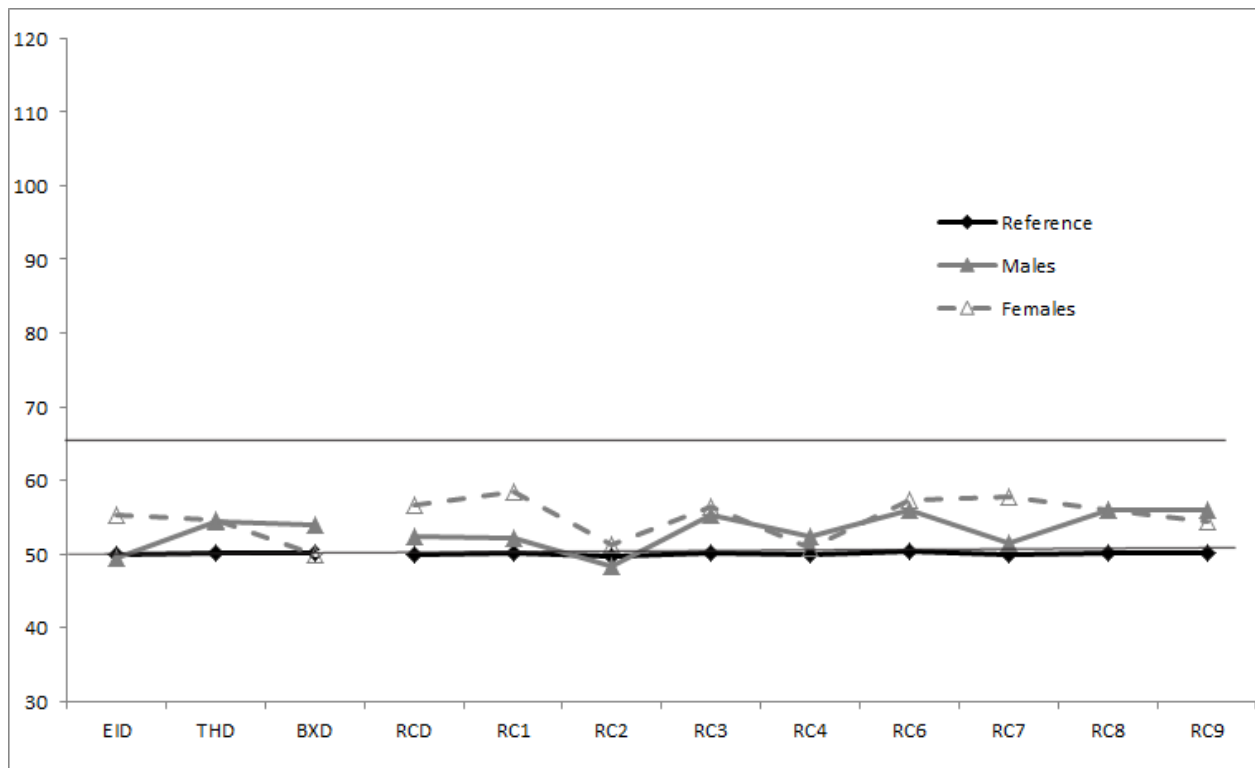


Figure 2: MMPI-2-RF Higher-Order and Restructured Clinical Scale scores by group. *EID* = Emotional/Internalizing Dysfunction. *THD* = Thought Dysfunction. *BXD* = Behavioral/Externalizing Dysfunction. *RCd* = Demoralization. *RC1* = Somatic Complaints. *RC2* = Low Positive Emotionality. *RC3* = Cynicism. *RC4* = Antisocial Behavior. *RC6* = Ideas of Persecution. *RC7* = Dysfunctional Negative Emotions. *RC8* = Aberrant Experiences. *RC9* = Hypomanic Activation.

Among the somatic and cognitive scales (see Figure 3), college men scored higher than the reference sample, with effects of medium magnitude, on both *NUC* and *COG* ($d = 0.54$ for each). College women, on the other hand scored significantly higher scores than the reference group on all five of these scales. Two scales, *NUC* ($d = 0.79$), and *COG* ($d = 0.83$), demonstrated large effect sizes, reflecting T score differences of 8.2 and 8.6 points, respectively. Two others,

MLS ($d = 0.50$) and *HPC* ($d = 0.64$) demonstrated medium effects. Each of these differences is clinically relevant. The effect size for *GIC* ($d = 0.39$) approached medium magnitude.

Among the internalizing scales, college men scored significantly higher than the reference sample on two of the nine scales. College men showed a clinically significant difference from the reference sample on *MSF*, obtaining a medium effect size ($d = 0.54$); a small effect ($d = 0.36$) was observed for *AXY*. College women obtained significantly higher scores on seven of the nine internalizing scales. Large effects were noted for *NFC*, *AXY*, and *BRF* ($d = 0.71, 0.96,$ and $0.78,$ respectively); the T score differences between college women and our reference group ranged from 7.2 to 10.6 points for these three scales. Medium effects were observed for *SFD* ($d = 0.62$), *STW* ($d = 0.61$), and *ANP* ($d = 0.50$); each of these differences reflects a clinically relevant disparity. The small effect for *MSF* ($d = 0.14$) was not clinically noteworthy. See Figure 3 for a graphical representation.

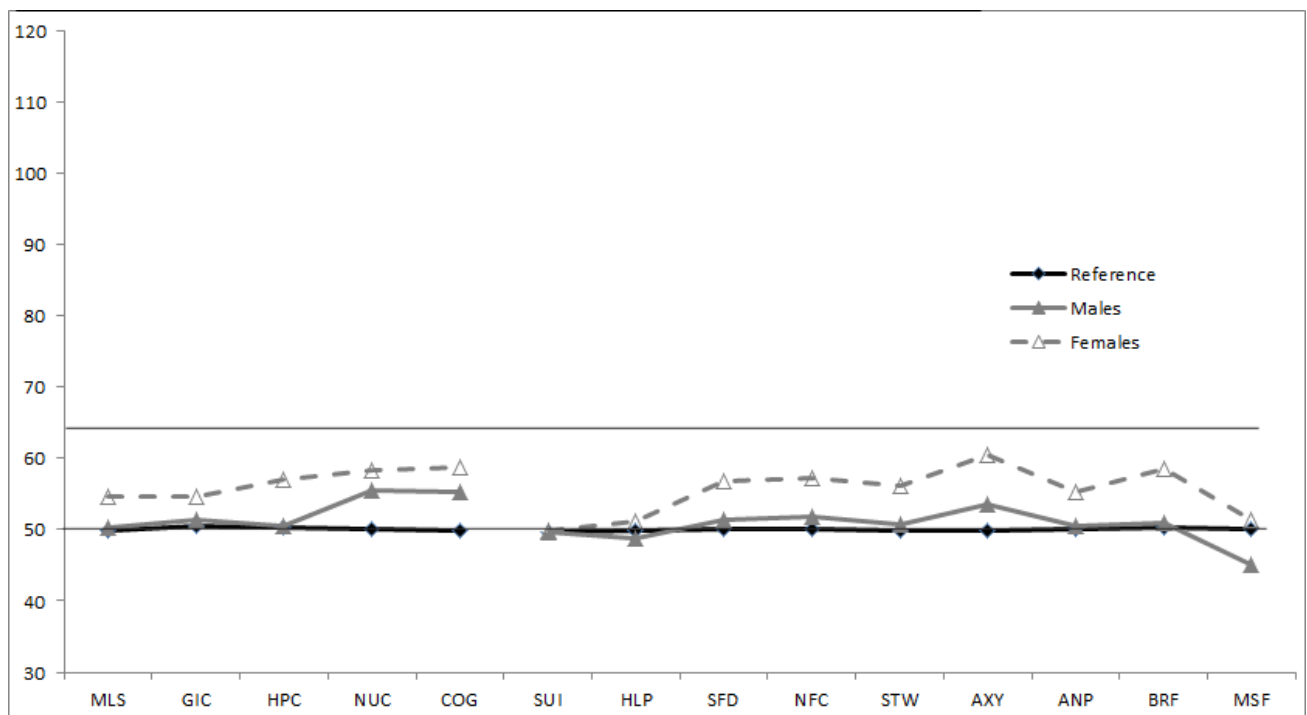


Figure 3: MMPI-2-RF Somatic-Cognitive and Internalizing Scale scores by group. *GIC* = Gastrointestinal Complaints. *HPC* = Head Pain Complaints. *NUC* = Neurological Complaints. *COG* = Cognitive Complaints. *SUI* = Suicidal/Death Ideation. *HLP* = Helplessness/Hopelessness. *SFD* = Self-Doubt. *NFC* = Inefficacy. *STW* = Stress/Worry. *AXY* = Anxiety. *ANP* = Anger Proneness. *BRF* = Behavior-Restricting Fears. *MSF* = Multiple Specific Fears.

College men scored significantly higher than the reference sample on three of the four externalizing scales. The effect sizes for *SUB*, *AGG*, and *ACT* ($d = 0.25, 0.27,$ and $0.35,$ respectively) were each of small effect and did not meet our criterion for clinical relevance. College women scored significantly higher than the reference group on two of these scales. The 7.7 T point difference on *ACT* reflected a large effect ($d = 0.72$). Additionally, women scored significantly lower than the reference sample on *JCP*; this difference was of small effect ($d = 0.30$) and of limited clinical relevance (see Figure 4).

College men obtained statistically, but not clinically, significantly different scores than the reference sample on two of the five interpersonal scales, as they scored lower on *IPP* ($d =$

0.36) and higher on *DSF* ($d = 0.31$). College women scored significantly higher than the reference group on three of the five interpersonal scales and lower on one. The higher score of college women on *FML* ($d = 0.41$) was of clinical relevance, whereas the higher scores on *SHY* and *DSF* and the lower score on *SAV* were not of clinical relevance (see Figure 4).

College men (see Figure 4) obtained a clinically and statistically significant lower score than the reference group on the *AES* ($d = 0.81$) interest scale and a clinically and statistically significant higher score than the reference group on *MEC* ($d = 0.44$). College women obtained clinically significant lower scores than the reference sample on both interest scales, with medium effect sizes observed for both *AES* ($d = 0.60$) and *MEC* ($d = 0.57$).

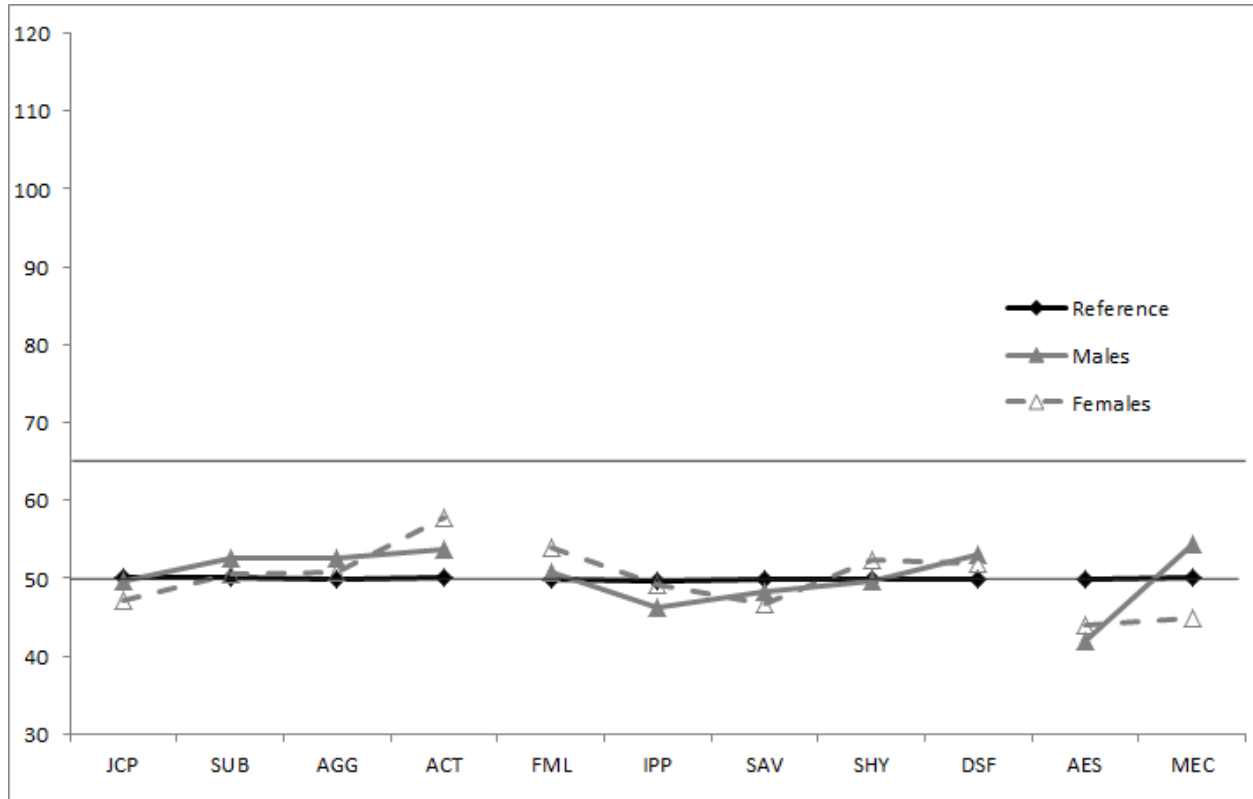


Figure 4: MMPI-2-RF Externalizing, Interpersonal, and Interest Scale scores by group. *JCP* = Juvenile Conduct Problems. *SUB* = Substance Abuse. *AGG* = Aggression. *ACT* = Activation. *FML* = Family Problems. *IPP* = Interpersonal Passivity. *SAV* = Social Avoidance. *SHY* = Shyness. *DSF* = Disaffiliativeness. *AES* = Aesthetic Interests. *MEC* = Mechanical Interests.

Among the PSY-5 scales (see Figure 5), college men scored significantly higher than the reference sample on four scales, with *AGGR-r* ($d = 0.46$), *PSYC-r* ($d = 0.41$), and *DISC-r* ($d = 0.60$) demonstrating medium effect sizes and clinically relevant differences. The effect size for *INTR-r* was small ($d = 0.28$) and of low clinical relevance. College women scored significantly higher than the overall reference sample on three PSY-5 scales. Clinically relevant effects were observed on two scales, as *NEGE-r* ($d = 0.77$) demonstrated a large effect size and a T score difference of 7.9 points. A medium effect size was observed for *PSYC-r* ($d = 0.48$). The effect size for *INTR-r* ($d = 0.39$) approached both a medium effect and clinical relevance.

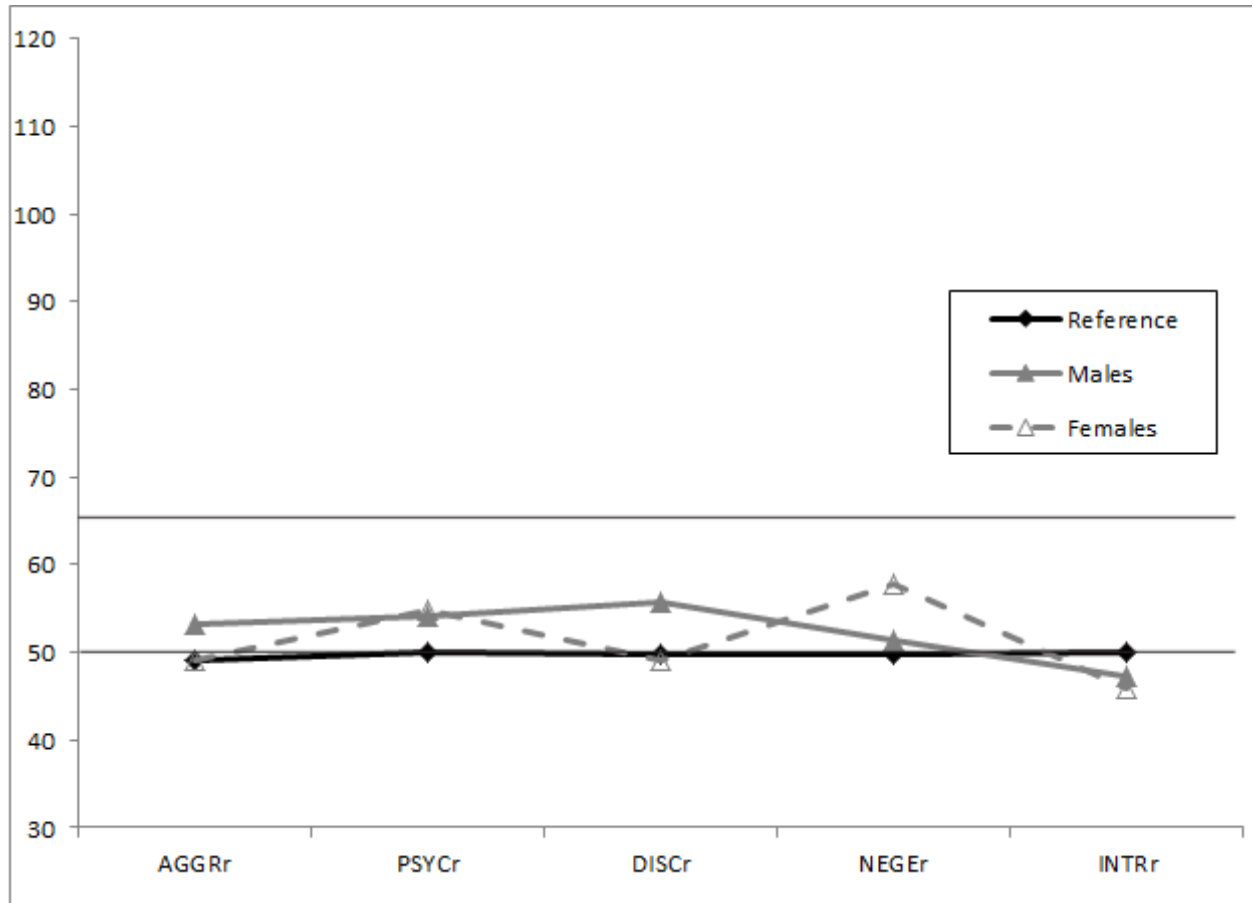


Figure 5: MMPI-2-RF PSY-5 Scale scores by group. *AGGR-r* = Aggressiveness PSY-5. *PSYCr* = Psychoticism PSY-5. *NEGE-r* = Negative Emotionality/Neuroticism PSY-5.

Discussion

The present study examined MMPI-2-RF scale score differences between college students and a reference group that was derived from the MMPI-2 normative sample in much the same way as was the MMPI-2-RF normative sample. In keeping with the past 60-plus years of research concerning the performance of college students on the earlier forms of the MMPI, our results indicated that college students had significantly and meaningfully different scores from our reference group on the majority of the MMPI-2-RF scales.

Our combined (men and women) college sample evidenced statistically significant differences on 44 of the 51 MMPI-2-RF scales when compared to our reference sample. Although we employed a relatively conservative Bonferroni correction of $p < .001$ as our significance criterion, a statistically significant difference does not always equal a clinically relevant difference – a fact which is all too often missed in our zeal for significant results. For example, although the combined college sample scored significantly higher than the reference sample on *BXD*, the actual difference in mean score was just over one T point, which is in no way clinically meaningful. Thus, we choose to focus our discussion on score differences that reflect what we consider to be clinically meaningful differences; we define this as those comparisons which resulted in a medium effect size of $d \geq 0.40$, which reflect T score difference of 4 points, or more.

By our effect size criterion, the combined college sample evidenced clinically relevant differences on 25 of the 51 MMPI-2-RF scales. The direction of the differences was in the area of greater dysfunction among students in the large majority of comparisons, as 23 of the differences reflected higher scores for the students. Lower scores were observed for students on *K-r* and *AES*, which reflect poorer adjustment and disinterest in arts or aesthetic pursuits. The most salient differences were observed on *Fp-r*, *Fs*, *NUC*, and *COG*, as each of these differences exceeded 7 T points. Differences of 6 T points were observed for *RC1*, *RC3*, *RC6*, *RC8*, *BRF*, and *ACT*.

College men evidenced statistically significant differences on 30 MMPI-2-RF scales when compared to our reference sample; however, only 16 of these differences met our criteria for clinical significance. The only scale evidencing a large effect for men was *AES*, as college men showed substantively fewer aesthetic interests and more stereotypic male interests than did the reference sample. The remainder of the clinically significant differences ranged from 4.1 (*THD*) to 6.0 (*RC8*) T points. College men scored higher on all significant comparisons, save *MSF* and *AES*.

Most concerning was that college women demonstrated statistically significant higher scores on 40 of the 51 scales of the MMPI-2-RF, and 31 of these were clinically significant differences. On 12 of these scales (*Fp-r*, *Fs*, *RC1*, *RC6*, *RC7*, *NUC*, *COG*, *NFC*, *AXY*, *BRF*, *ACT*, and *NEGE-r*) the differences between college women and the reference sample exceeded 7 T points, which means that, in many ways, the average college female is closer to producing an elevated MMPI-2-RF profile than one that would be viewed as within normal limits. Further, college women averaged score differences of at least 6 T points on an additional 9 scales (*F-r*, *RBS*, *K-r*, *RCd*, *RC3*, *RC8*, *HPC*, *SFD*, and *STW*).

The mean scores observed in our sample and the patterns of differences from an average score of 50 are similar to those reported by Tellegen and Ben-Porath (2011) in the *Technical Manual*. Thus, the present results do not appear to be anomalous to our sample. Further, they are, in many ways, consistent with results of comparisons between college students and adults on the various MMPI forms over the past 60 years. The obvious problems, then, associated with applying adult norms to the assessment of college students are not endemic to the MMPI-2-RF, but are likely a function of college students existing in a unique developmental time period wherein they are no longer like younger adolescents, but have not fully developed into adults. This echoes Osberg and Poland's (2002) earlier warning that while younger college students tend to produce higher than average elevations when adult norms are used, they produce lower than average elevations when adolescent norms are used.

Our finding that, overall, women college students evidenced greater disparity – largely in the direction of psychopathology – from our reference group than did men is especially troubling and suggests that non-gendered norms may not be appropriate for use with college students, as, in the current study, at least, college women tended to produce higher elevations than did men. Further research is needed to examine this issue. Butcher, Hass, Greene, and Nelson (2015) noted that the use of non-gendered norms in adults will result in lower men's scores and higher women's scores compared to the use of gendered norms. We observed a similar phenomenon within our sample as women's scores were suggestive of a higher degree of dysfunction than were those of men.

The degree of difference that we termed clinically significant is similar to that degree of difference that was observed when plotting contemporary samples on the original MMPI norms. A generation ago, this, among other reasons, served as an impetus for a revision and

restandardization of the MMPI that resulted in the MMPI-2. Further, differences of this magnitude have been observed among college students and other samples of similar age (cf. Butcher, Jeffrey, et al., 1990) for more than 60 years. Still, no official norms for college-aged individuals have been officially supported.

Over 60 years ago, Sopchak (1952) called for specific norms for college students. Following the earlier work of Marks and Briggs (1972), Gottesman et al. (1987) recognized that developmental differences in adolescence can lead to MMPI scale score differences. Thus, they created separate gendered norms for 15 and 18 year olds. Further, recognizing that a 20 year-old may be developmentally closer to an 18 year-old than to a 40 year-old, they recommended that individuals between the ages of 18 and 20 be scored using both adult and 18 year-old norms. Only a single set of gender-based adolescent norms was created for the MMPI-A, however. Likewise, a single set of non-gendered norms was created for the MMPI-A-RF using a subsample of the MMPI-A normative group. We urge that future revisions of the MMPI create a separate set of norms for late adolescents/early adults that more clearly reflect their developmental stage.

For now, we present the results of our study and caution those who may use the MMPI-2-RF with college-aged individuals to be careful with regard to over-interpreting scores that may appear to reflect pathology. The data that were obtained in this study may be especially useful in college counseling centers where clinicians should be cautious when interpreting results of the MMPI-2-RF within such samples. This concern is especially relevant with regard to use of the instrument with college-aged women, as endorsing only a few items beyond the mean for this age group may result in scales that cross over the clinical threshold.

About the Authors

P. Kevin Bolinskey, Ph.D. is with the Department of Psychology at Indiana State University in Terre Haute, Indiana. Correspondence concerning this article should be addressed to: P. Kevin Bolinskey; Department of Psychology; B-214 Root Hall; Indiana State University; Terre Haute, IN 47809; **Phone:** 812.237.2463; **Email:** *Kevin.Bolinskey@indstate.edu*

Janice P. Guidi, M.S. is with the Department of Psychology at Indiana State University in Terre Haute, Indiana.

Kevin R. Myers, M.S. is with the Department of Psychology at Indiana State University in Terre Haute, Indiana.

Dianna Cooper-Bolinskey, M.S.W. is with the Department of Social Work at Indiana University in Terre Haute, Indiana.

Kelly M. Schuder, M.S. is with the Department of Psychology at Indiana State University in Terre Haute, Indiana.

Allison V. James, Psy.D. is with the Department of Psychology at Indiana State University in Terre Haute, Indiana.

Daniel V. Hudak, Psy.D. is with the Department of Psychology at Indiana State University in Terre Haute, Indiana.

Yelena Gonzalez, M.S. is with the Department of Psychology at Indiana State University in Terre Haute, Indiana.

Erin McTiernan, M.S. is with the Department of Psychology at Indiana State University in Terre Haute, Indiana.

Elizabeth A. Smith, M.S. is with the Department of Psychology at Indiana State University in Terre Haute, Indiana.

References

- Archer, R. P. (2005). *MMPI-A: Assessing adolescent psychopathology*. (3rd ed.). Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers.
- Ben-Porath, Y. S., & Tellegen, A. (2011). *Minnesota Multiphasic Personality Inventory-2 Restructured Form: Manual for administration, scoring, and interpretation*. Minneapolis, MN: University of Minnesota Press.
- Brown, H. S. (1948). Similarities and differences in college populations on the Multiphasic. *Journal Of Applied Psychology*, 32, 541-549. doi:10.1037/h0056219
- Butcher, J. N., Dahlstrom, W. G., Graham, J. R., Tellegen, A., & Kaemmer, B. (1989). *Minnesota Multiphasic Personality Inventory-2 (MMPI-2): Manual for administration and scoring*. Minneapolis: University of Minnesota Press.
- Butcher, J. N., Graham, J. R., Dahlstrom, W., & Bowman, E. (1990). The MMPI-2 with college students. *Journal of Personality Assessment*, 54, 1-15. doi:10.1207/s15327752jpa5401&2_1
- Butcher, J. N., Hass, G. A., Greene, R. L., & Nelson, L. D. (2015). The MMPI-2-Restructured form. In J. N Butcher, G. A. Hass, R. L. Greene, and L. D. Nelson (Eds.), *Using the MMPI-2 in forensic assessment* (pp. 249-263). Washington, DC: American Psychological Association. doi: 10.1037/14571-015
- Butcher, J. N., Jeffrey, T., Cayton, T. G., Colligan, S., DeVore, J. R., & Minegawa, R. (1990). A study of active duty military personnel with the MMPI-2. *Military Psychology*, 2, 47-61. doi:10.1207/s15327876mp0201_4
- Butcher, J. N., & Williams, C. L. (2012). Problems with using the MMPI-2-RF in forensic evaluations: A clarification to Ellis. *Journal of Child Custody: Research, Issues, And Practices*, 9(4), 217-222. doi:10.1080/15379418.2012.748347
- Butcher, J. N., Williams, C. L., Graham, J. R., Archer, R. P., Tellegen, A., Ben-Porath, Y. S., & Kaemmer, B. (1992). *Minnesota Multiphasic Personality Inventory-Adolescent Version (MMPI-A): Manual for administration, scoring and interpretation*. Minneapolis, MN: University of Minnesota Press.
- 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(1), 14-24. doi: 10.1037/1040-3590.19.1.14
- Forbey, J. D., & Ben-Porath, Y. S. (2008). Empirical correlates of the MMPI-2 Restructured Clinical (RC) scales in a nonclinical setting. *Journal of Personality Assessment*, 90(2), 136-141. doi: 0.1080/00223890701845161

- Gilliland, A. R., & Colgin, R. (1951). Norms, reliability, and forms of the MMPI. *Journal of Consulting Psychology, 15*, 435-438. doi:10.1037/h0057211
- Gottesman, I. I., Hanson, D. R., Kroeker, T. A., & Briggs, P. R. (1987). New MMPI normative data and power-transformed T score tables for the Hathaway-Monachesi Minnesota Cohort of 14,019 15-year-olds and 3,674 18-year-olds. In R. P. Archer, *Using the MMPI with adolescents* (pp. 241-297). Hillsdale, NJ: Erlbaum.
- Hunter, H. K., Bolinskey, P. K., Novi, J. H., Hudak, D. V., James, A. V., Myers, K. R., & Schuder, K. M. (2014). Using the MMPI-2-RF to discriminate psychometrically-identified schizotypic college students from a matched comparison sample. *Journal of Personality Assessment*. doi: 10.1080/00223891.2014.922093
- Marks, P. A., & Briggs, P. (1972). Adolescent norm tables for the MMPI. In W. G. Dahlstrom, G.S. Welsh, & L. E. Dahlstrom (Eds.), *An MMPI handbook (Vol. 1): Clinical interpretation* (pp.388-399). Minneapolis: University of Minnesota Press.
- Nichols, D. S. (2011). *Essentials of MMPI-2 Assessment* (2nd ed.). New York: John Wiley & Son.
- Osberg, T. M., Haseley, E. N., & Kamas, M. M. (2008). The MMPI-2 Clinical Scales and Restructured Clinical (RC) Scales: Comparative psychometric properties and relative diagnostic efficiency in young adults. *Journal of Personality Assessment, 90*, 81-92. doi:10.1080/00223890701693801
- Osberg, T. M., & Poland, D. L. (2002). Comparative accuracy of the MMPI-2 and the MMPI-A in the diagnosis of psychopathology in 18-year-olds. *Psychological Assessment, 14*, 164-169. doi:10.1037/1040-3590.14.2.164
- Pancoast, D. L., & Archer, R. P. (1989). Original adult MMPI norms in normal samples: A review with implications for future developments. *Journal of Personality Assessment, 53*(2), 376-395. doi:10.1207/s15327752jpa5302_14
- Pancoast, D. L., & Archer, R. P. (1992). MMPI response patterns of college students: Comparisons to adolescents and adults. *Journal of Clinical Psychology, 48*(1), 47-53. doi:10.1002/1097-4679(199201)48:1<47::AID-JCLP2270480106>3.0.CO;2-B
- Sopchak, A. L. (1952). College student norms for the Minnesota Multiphasic Personality Inventory. *Journal of Consulting Psychology, 16*, 445-448. doi:10.1037/h0057545
- Strassberg, D. S. (1997). A cross-national validity study of four MMPI-2 content scales. *Journal of Personality Assessment, 69*, 596-606. doi:10.1207/s15327752jpa6903_12
- Tellegen, A., Ben-Porath, Y. S. (2011). *Minnesota Multiphasic Personality Inventory-2 Restructured Form: Technical Manual*. Minneapolis: University of Minnesota Press.

- Twenge, J. M., Gentile, B., DeWall, C. N., Ma, D., Laceyfield, K., & Schurtz, D. R. (2010). Birth cohort increases in psychopathology among young Americans, 1938–2007: A cross-temporal meta-analysis of the MMPI. *Clinical Psychology Review, 30*(2), 145-154. doi:10.1016/j.cpr.2009.10.005
- Tyler, F. T., & Michaelis, J. U. (1953). A comparison of manual and college norms for the MMPI. *Journal of Applied Psychology, 37*, 273-275. doi:10.1037/h0060366