Personality and Mental Health
3: 12–25 (2009)
Published online in Wiley InterScience
(www.interscience.wiley.com) DOI: 10.1002/pmh.61

Comparative gender biases in models of personality disorder

DOUGLAS B. SAMUEL AND THOMAS A. WIDIGER, Department of Psychology, University of Kentucky, USA

ABSTRACT

The potential of gender bias within the DSM personality disorders has long been a concern of scholars and clinicians. Over the past three decades, research findings utilizing the case vignette methodology have repeatedly indicated a gender bias within the histrionic diagnosis. The current study replicates these findings using a novel case vignette, but extends them to investigate the potential for gender biases within an alternative dimensional model of personality—the Five-Factor Model (FFM). One hundred and forty-one practicing clinicians rated either a male or a female version of a case vignette in terms of either the FFM or the personality disorders from the Diagnostic and Statistical Manual of Mental Disorders (DSM-IV-TR). The results supported the concern of gender bias, with the female case less likely to be diagnosed as antisocial and the male case less likely to be diagnosed as histrionic. However, when the FFM conceptualizations of these two disorders were compared, no significant differences were noted. The results indicate that the FFM may be less prone to gender bias than the current DSM nomenclature. Copyright © 2009 John Wiley & Sons, Ltd.

Introduction

The possibility of gender bias within the personality disorders (PDs) has been one of the most controversial aspects of the American Psychiatric Association's (APA) Diagnostic and Statistical Manual (DSM-IV-TR; APA, 2000) over the past few decades (Garb, 2005; Haslam, 2006; Ross, Frances, & Widiger, 1995; Widiger, 2007). Kaplan's (1983) paper was perhaps the first and has certainly been the most widely cited critique since the appearance of DSM-III (APA, 1980). Kaplan suggested in her *American Psychologist* paper that 'a healthy woman automatically earns the diagnosis of Histrionic Personality Disorder' (p. 789), due in large part to the codification of a gender biased criteria set constructed by a DSM-III task force comprised largely of males.

Kaplan's (1983) suggestion that a normal woman would be diagnosed with histrionic PD (HPD) was perhaps overstated, but quite a number of subsequent studies have supported her concern of gender bias in the assessment and diagnosis of PDs. In fact, these studies have indicated consistently that

histrionic is the PD that is most prone to potential gender biases.

One of the first studies to illustrate the potential for gender bias in the diagnosis of HPD actually predated Kaplan's critique and the publication of DSM-III. Warner (1978) presented a hypothetical clinical profile to 175 mental health professionals. The gender of the case vignette was varied to determine if this would impact diagnostic rates. Warner reported that when the case was presented as a female, 76% of the clinicians assigned the diagnosis of hysterical while only 22% chose the diagnosis of antisocial PD (APD). In contrast, when the case was described as a male, only 49% of the participants assigned the diagnosis of hysterical and 41% opted for antisocial.

Case vignette methodology

Since the publication of Warner (1978), there have been numerous other studies that have also investigated the possibility of gender bias in the diagnosis of PDs using the case vignette methodology. There have been other methods used, including the examination of perceived dysfunction of male-typed and female-typed diagnostic criteria (Howell & Watson, 2002), differential impairment (Boggs et al., 2005), clinicians' ratings of their own clients (Blashfield & Herkov, 1996; Morey & Ochoa, 1989), and the provision of diagnoses by novice undergraduates (Flanagan & Blashfield, 2003). Nevertheless, a predominant method for exploring the possibility of gender bias has been the case vignette methodology, which has also been used for the study of potential race and ethnicity biases in PD diagnosis (Mikton & Grounds, 2007; Minnis, McMillan, Gillies, & Smith, 2001).

The most consistent finding has been the tendency for clinicians to more frequently diagnose HPD in a female case than in a male case. There has also been support for a complementary bias favoring the diagnosis of antisocial in males relative to female cases, albeit not quite as consistent or strong as obtained for the histrionic diagnosis (Garb, 2005; Morey, Warner, & Boggs, 2002; Widiger, 1998).

For example, Hamilton, Rothbart, and Dawes (1986) developed five case vignettes that varied in their level of histrionic and antisocial pathology, ranging from purely histrionic to purely antisocial. These five cases were then presented, along with several 'foil' cases, as either male or female, to a sample of psychologists. Analysis of variance (ANOVA) tests revealed that the main effect for gender was significant for the diagnosis of HPD, but not for antisocial, indicating that females were likely to be rated as more histrionic than males exhibiting the same symptoms.

Ford and Widiger (1989) developed three vignettes that described a case of HPD, APD, or one with equal criteria for each disorder. The client was presented as male, female, or neuter (i.e. the gender of the case was not specified). For the histrionic case, clinicians significantly more often failed to diagnose HPD in male patients (44%) than in female patients (76%). For the antisocial case, clinicians significantly more often failed to diagnose APD for female patients (15%) than in male patients (46%). In fact, antisocial female patients were more likely to be diagnosed with HPD (46%) than with APD (15%).

Fernbach, Wimstead, and Derlega (1989) similarly found that when male and female versions of an antisocial case were rated, the clinicians were significantly less likely to correctly diagnose the female as antisocial, than the male. Crosby and Sprock (2004) also employed an antisocial case and again found that while the most common diagnosis for both genders was indeed APD, the female version was significantly more likely to be diagnosed with HPD than the male version.

Adler, Drake, and Teague (1990) developed a case vignette that met diagnostic criteria for the APD, HPD, borderline PD (BPD), and narcissistic PDs. Overall, the case was most often diagnosed as borderline, regardless of gender. However, when this case was described as a female, the mental health professionals were more likely to provide the diagnosis as histrionic, whereas the male

DOI: 10.1002/pmh

version was more likely to be diagnosed as narcissistic.

Becker and Lamb (1994) used case vignettes representing either BPD or post-traumatic stress disorder (PTSD). No significant gender effects were noted for the PTSD case, but results indicated that for the borderline case, the mental health professionals were less likely to diagnose the male as histrionic and the female as antisocial, replicating the findings of Ford and Widiger (1989).

In the most recent demonstration of this finding, Flanagan and Blashfield (2005) used the original case vignette of Warner (1978) as well as other vignettes used in subsequent studies. Statistical tests revealed that across all cases, males were more likely to be diagnosed as antisocial and females as histrionic. More specifically, the HPD diagnosis was less likely to be used for the male than for the female version of a purely histrionic case as well as a mixed case. The APD diagnosis was less likely to be used for the female relative to the male version of a purely antisocial case, the Warner (1978) case, and a purely histrionic case.

Five-factor model of PD

For reasons well beyond concerns regarding gender bias, researchers have called for the abandonment of the current categorical system of PD diagnosis in favor of an alternative dimensional model of PD (Clark, 2007; Livesley, 2003; Widiger & Trull, 2007). Amongst the more heavily researched of these alternative models is the Five-Factor Model (FFM; McCrae & Costa, 1999), consisting of the five broad dimensions of extraversion (vs. introversion), agreeableness (vs. antagonism), neuroticism (vs. emotional stability), conscientiousness (vs. undependability) and openness (vs. closedness to experience). Costa and McCrae (1992) have proposed that each of these domain is, in turn, underlain by six facets.

Over the past two decades, the FFM has provided a useful dimensional framework for understanding the DSM PDs (Clark, 2007; O'Conner, 2005; Samuel & Widiger, in press; Widiger & Costa, 2002). It is important though for research to go beyond simply indicating whether the PDs can be understood in terms of the FFM and to indicate whether there are any benefits or advantages in converting to this dimensional model of personality structure (Clark, 2007; Verheul, 2005; Widiger & Trull, 2007). The purpose of the current study was to determine whether an FFM description of a client, using the case vignette methodology, was as susceptible to gender bias as the DSM.

The FFM and DSM-IV-TR provide comparable hierarchical models of personality description. At the highest levels are the three clusters of the DSM-IV-TR (i.e. odd-eccentric, dramatic-emotional and anxious-fearful) and the five domains of the FFM (i.e. neuroticism, extraversion, openness, agreeableness and conscientiousness). Beneath this broad level of description are the 30 facets of the FFM (e.g. anxiousness and mistrust) and the 10 PDs (e.g. avoidant and paranoid). Comparisons are perhaps most appropriate at equivalent levels of the hierarchy (e.g. facets of the FFM vs. PDs of the DSM-IV-TR). Findings and hypotheses of bias have been, almost exclusively, at the level of the 10 PDs, thus it is most appropriate to examine the FFM at the level of the 30 facets. Nonetheless, because it is possible that higher levels of a hierarchy may be more prone to broad conceptualizations and thus potential bias, it would be important that comparisons also are made at these higher levels (i.e. FFM domains and DSM-IV-TR clusters).

Widiger, Trull, Clarkin, Sanderson, and Costa (2002) translated each PD diagnostic criterion into a specific, relevant facet of the FFM to develop profile descriptions of each DSM-IV-TR PD. More specifically, they indicated that HPD would be translated as high standings on the extraversion facets of warmth, gregariousness, excitement seeking and, positive emotions; the neuroticism facets of depression and self-consciousness; the openness facets of fantasy and feelings and the

DOI: 10.1002/pmh

agreeableness facet of trust. Widiger and colleagues indicated APD would be characterized by low standings on the agreeableness facets of straightforwardness, altruism, compliance and tender mindedness; the conscientiousness facets of dutifulness, self-discipline and deliberation; and high standings on the extraversion facet of excitement seeking and the neuroticism facet of angry hostility. These hypothesized FFM profiles for HPD and APD have since been confirmed by the consensus descriptions provided by researchers (Lynam & Widiger, 2001) and practicing clinicians (Samuel & Widiger, 2004).

In the current study, the HPD and APD FFM facets were investigated to determine whether the FFM conceptualizations are as prone to gender bias as the DSM diagnoses. For example, to the degree that a female case vignette is rated as more histrionic than its male counterpart, one might also expect the female version to be rated more highly on the corresponding facets of the FFM. Conversely, if these FFM facets are not significantly different between male and female versions of the same case, then the FFM could be considered less prone to this form of gender bias. The current study sought to examine this hypothesis using a brief case vignette of a real (female) individual, as well as an alternative vignette where the gender of the individual was portrayed as male.

Method

Fourteen hundred participants from the APA's Division 42 (Private Practitioners) were solicited via postal mail. Each participant completed a brief demographic questionnaire and read a 1.5 page case vignette (describing either a male or a female) and then provided ratings of that case for either the FFM or the DSM PDs. Thus, the four conditions consisted of a female case/DSM-IV, female case/FFM, male case/DSM-IV, and male case/FFM. Participants returned the materials in a postage-paid envelope.

Case history

The case history used in the current study was 'Madeline', drawn from the recently published text *Paradigms of Personality Assessment*, by Dr. Jerry Wiggins (Wiggins, 2003). In this text, Wiggins asked leading experts from five different paradigms of personality assessment (i.e. psychodynamic, multivariate, interpersonal, personological and empirical) to assess and describe the same person. The person he chose was 'Madeline G' who provided an intriguing case study with surprising depth and colour of character. A recent review of Wiggins' text suggests, 'Madeline G may go down in history as one of the best case studies ever published' (Strack, 2005, p. 106). The description of Madeline within the 1.5 page, single-spaced vignette was drawn largely from the life history interview conducted by Dr. Dan McAdams and the peer description provided by Dr. Krista Trobst and Dr. Jerry Wiggins (Wiggins, 2003). This case history was originally developed for and used in a previous study (Samuel & Widiger, 2006).

This vignette was considered to be appropriate for the current study as it concerned a real person who was deemed by 87% of the clinicians studied by Samuel and Widiger (2006) to meet criteria for HPD (49% diagnosed her with APD). The vignette used in the current study was identical to the original with only one exception. The vignette used in Samuel and Widiger indicated that Madeline would '"flash" men at parties (e.g. exposing her breasts)'. It was felt that a direct translation of this behaviour for a male would artefactually alter its meaning in a manner that would bias the findings in a way favourable to the hypothesis of the current study (i.e. a male who flashes women at parties by exposing himself would likely be suggestive of a disorder other than HPD). Although the elimination of this particular behaviour might decrease the likelihood that Madeline would be perceived as histrionic, including the behaviour was felt to be more problematic. For the male version of this vignette, the name of Madeline became 'Matthew' and all the personal pronouns were switched to

DOI: 10.1002/pmh

their male equivalent. There was otherwise no change.

Instruments

FFM rating form. The FFM rating form (FFMRF) is a one-page measure of the 30 facets of the FFM. The individual is described on each facet using a 1–5 Likert scale where 1 = extremely low, 2 = low, 3 = neutral, 4 = high and 5 = extremely high. Each of the 30 facets is labelled with a trait term (e.g. the first facet of neuroticism is labelled 'anxiousness'). In addition to this label, both the high and low pole of each facet contains 2–3 trait descriptors to assist the user in making the ratings. For example, the facet of 'gregariousness' is described by the words 'sociable' and 'outgoing' at the high pole and 'withdrawn' and 'isolated' at the low pole. The FFMRF has been utilized in previous studies to aggregate clinicians' FFM ratings (i.e. Samuel & Widiger, 2004; 2006). Additionally, it has been used as a self-report instrument and has shown good convergent validity with more extensive measures of the FFM (Mullins-Sweatt, Jamerson, Samuel, Olson, & Widiger, 2006).

DSM rating form (DSMRF). The DSMRF is a one-page measure that assists the clinician in providing dimensional ratings for each of the 10 DSM-IV-TR PDs. All 10 DSM PDs are listed along with a brief one-sentence description (e.g. 'pattern of distrust and suspiciousness such that others' motives are interpreted as malevolent' for paranoid PD). The clinician rates the extent to which the individual is characterized by each of the disorders on a 1–5 Likert scale where 1 = absent, 2 = subthreshold, 3 = threshold, 4 = above threshold, and 5 = prototypic. After rating each disorder, the clinician is asked to provide a final DSM-IV diagnosis, for which they may select: (A) One or more of the above diagnoses, (B) PD not otherwise specified or (C) no PD diagnosis. The DSMRF has been employed in previous studies as a way to collect clinicians' diagnostic impressions of the DSM-IV-TR PDs (Samuel & Widiger, 2004; 2006).

Demographic questionnaire. Each clinician was also asked to complete a questionnaire that gathered basic demographic data as well as information about training, experience, direct clinical contact hours and theoretical orientation. They were also asked to rate their level of familiarity with the rating system they had used (i.e. DSM or FFM) as 'not at all familiar', 'vaguely familiar', 'average level of familiarity', 'moderately familiar' or 'very familiar'.

Results

Demographics

Of the 1 400 total surveys mailed, 492 envelopes were returned unopened, leaving 908 that presumably reached their intended targets. Of this number, 141 psychologists returned the survey, yielding a total response rate of 15.5%, with comparable rates within each of the four groups (ranging from 14.4% to 17.4%). This response rate is similar to that of previous samples of practicing clinicians (e.g. 19.4% from Samuel & Widiger, 2006; 17% in Crosby & Sprock, 2004; 16% in Flanagan & Blashfield, 2005) and superior to that obtained from a recent survey of a broader population of mental health professionals (e.g. 3.3% Spitzer, First, Shedler, Westen, & Skodol, 2008). Three participants were later eliminated because of incomplete data, leaving a total of 138 usable responses. Of this total, 40 psychologists provided DSM ratings of Matthew, 32 provided DSM ratings of Madeline, 34 provided FFM ratings of Matthew and 32 provided FFM ratings of Madeline.

A series of one-way ANOVAs were carried out on the demographic variables to detect potential differences amongst the four groups. As none of these tests indicated significant differences, the following demographic information is provided for the entire sample. The sample was predominantly male (60.8%) and Caucasian (96.2%). All respondents were doctoral level psychologists, with 93.8% listing their highest degree as a PhD, 4.6% as an EdD and 1.5% as a PsyD. Seventy-eight percent

DOI: 10.1002/pmh

listed their subfield as clinical psychology, while 18.0% were in counseling psychology, and 3.1% indicated they were educational psychologists. The clinical experience of the participants ranged from a low of 2 years to a high of 66 years, with a mean of 30 years since earning their degree. The majority of the participants were full-time clinicians, with a mean of 71.9% of their working hours spent directly providing services to clients. Sixty-one percent of the respondents identified themselves as having a cognitive theoretical orientation, 44.6% psychodynamic, 39.2% behavioural, 36.2% interpersonal, 19.2% humanistic, and 19.2% endorsed 'other' (respondents could choose more than one option).

An independent samples *t*-test indicated that the clinicians were significantly more familiar with the DSM-IV system than the FFM, $t(128) = 10.81$, $p < 0.001$. In addition, the psychologists who rated Madeline were not more familiar with the FFM or DSM than those who rated Matthew.

Reliability

The reliability of the composite FFM and DSM profiles were calculated separately for each of the four subsamples. With raters serving as variables and facets/PDs as cases, the values for Cronbach's alpha indicated excellent reliability with values above 0.98 for each group. To provide a more stringent measure of agreement, we calculated intraclass correlation coefficients (ICCs) using an absolute agreement definition. These ICCs appear slightly lower for the DSM, with a value of 0.58 for Madeline and 0.59 for Matthew, compared to 0.72 and 0.61, respectively, for the FFM. However, neither of these comparisons was statistically significant ($z = 1.16$ for Madeline and 0.13 for Matthew; both $p > 0.05$). The ICC results are also consistent with the values (i.e. 0.58 for the DSM and 0.67 for the FFM) previously reported for the case of Madeline (Samuel & Widiger, 2006).

PD diagnoses

Table 1 presents the mean consensus ratings for all 10 DSM-IV PDs provided by the clinicians for both Madeline and Matthew. It is apparent that the clinicians considered Madeline to be suffering primarily from narcissistic and histrionic PDs, consistent with the findings of Samuel and Widiger (2006). Most important for the purpose of this study, however, was the impact of gender. When the case was a male, the mean rating for HPD dropped to 2.53, which is below the threshold for the HPD diagnosis and statistically significantly different from the female case, $t(70) = 2.75$, $p = 0.006$. Additionally, the clinicians rated the male version more highly on APD than the female version of the case, a difference that was also significant, $t(70) = 2.10$, $p = 0.039$. Finally, the male version of the case was also rated as significantly more schizoid than the female version, but this difference was not predicted based on previous research. No significant effects were observed for any of the other PDs.

Tests were also conducted to examine potential gender differences for the DSM-IV-TR clusters. These scores were calculated by summing the ratings for each of the PDs that comprise the

Table 1: Comparison of DSM-IV PD ratings

DSM-IV PDs	Madeline ($n = 32$)		Matthew ($n = 40$)		*t*
	Mean	SD	Mean	SD	
Paranoid	1.19	0.40	1.36	0.49	–1.64
Schizoid	1.23	0.56	**1.74**	0.91	–2.93*
Schizotypal	1.31	0.59	1.58	0.93	–1.45
Antisocial	2.16	0.90	**2.68**	1.16	–2.10*
Borderline	2.81	1.33	2.44	1.02	1.32
Histrionic	**3.28**	0.99	2.53	1.28	2.75*
Narcissistic	4.03	1.05	4.38	0.77	–1.59
Avoidant	1.16	0.37	1.30	0.72	–1.09
Dependent	1.16	0.45	1.13	0.52	0.27
Obsessive–compulsive	1.38	0.61	1.63	0.93	–1.38

Note: *t* values with an asterisk indicate those which were significant at $p < 0.05$, df = 70. Mean values in bold indicate the value that was significantly higher

DSM-IV-TR, Diagnostic and Statistical Manual of Mental Disorders—4th Edition—Text Revision; PDs, personality disorders.

DOI: 10.1002/pmh

respective cluster. The results of these analyses indicated that cluster A scores were significantly higher for the male version, than for the female version, $t(67) = 2.42$, $p = 0.02$. Analyses for clusters B ($t(67) = -0.40$, $p = 0.69$) and C ($t(70) = 1.07$, $p = 0.29$) did not reveal statistically significant differences across the case gender.

FFM ratings

Table 2 presents the means and standard deviations of the FFM ratings for both Madeline and Matthew. The FFM description of HPD includes high standings on the extraversion facets of warmth, gregariousness, excitement seeking and

Table 2: Comparison of FFM ratings

FFM domains and facets	Madeline ($n = 32$)		Matthew ($n = 34$)		t
	Mean	SD	Mean	SD	
Neuroticism					
(N1) Anxiousness	**3.09**	1.15	2.29	1.22	2.74*
(N2) Angry hostility	3.97	0.47	3.88	0.81	0.53
(N3) Depressiveness	2.38	0.83	2.62	1.05	−1.04
(N4) Self-consciousness	1.44	0.56	1.33	0.78	0.62
(N5) Impulsivity	3.97	1.00	3.79	1.18	0.65
(N6) Vulnerability	2.16	0.99	2.12	1.10	0.15
Extraversion					
(E1) Warmth	2.35	1.05	2.18	0.97	0.71
(E2) Gregariousness	4.50	0.67	4.68	0.54	−1.18
(E3) Assertiveness	4.81	0.40	4.82	0.39	−0.11
(E4) Activity	4.78	0.42	**5.00**	0.00	−2.95*
(E5) Excitement-seeking	4.81	0.40	4.65	0.49	1.52
(E6) Positive emotions	4.00	0.57	4.00	0.92	0.00
Openness					
(O1) Fantasy	3.03	1.00	2.65	1.23	1.40
(O2) Aesthetics	**3.26**	0.97	2.65	0.95	2.57*
(O3) Feelings	2.28	0.92	1.85	0.89	1.92
(O4) Actions	**4.41**	0.56	3.94	0.81	2.69*
(O5) Ideas	3.48	1.12	3.09	1.03	1.49
(O6) Values	4.19	0.83	3.71	1.22	1.86
Agreeableness					
(A1) Trust	2.06	0.62	1.91	0.75	0.89
(A2) Straightforwardness	1.72	0.81	1.61	1.12	0.46
(A3) Altruism	2.13	0.92	1.76	0.78	1.72
(A4) Compliance	1.78	0.55	1.85	0.78	−0.43
(A5) Modesty	1.48	0.81	1.26	0.51	1.32
(A6) Tender-mindedness	1.88	0.79	1.59	0.61	1.65
Conscientiousness					
(C1) Competence	4.41	0.62	4.18	0.76	1.35
(C2) Order	3.88	0.87	3.82	0.90	0.24
(C3) Dutifulness	3.23	1.15	2.91	1.16	1.09
(C4) Achievement	4.69	0.47	4.74	0.62	−0.35
(C5) Self-discipline	3.81	1.01	3.59	1.16	0.81
(C6) Deliberation	2.52	0.89	2.47	1.13	0.18

Note: t values with an asterisk indicate those which were significant at $p < 0.05$, df = 64. Mean values in bold indicate the value that was significantly higher.
SD, standard deviation; FFM, Five-Factor Model.

DOI: 10.1002/pmh

positive emotions, as well as high standings on the neuroticism facets of depression and self-consciousness; the openness facets of fantasy and feelings; and the agreeableness facet of trust (Widiger et al., 2002). There were no significant differences between Madeline and Matthew on any of these facets.

Because Matthew was rated as significantly higher than Madeline on APD, he might also be expected to be significantly lower than Madeline on the relevant FFM agreeableness facets of straightforwardness, altruism, compliance and tender mindedness; the conscientiousness facets of dutifulness, self-discipline and deliberation; and higher on the extraversion facet of excitement seeking and the neuroticism facet of angry hostility (Widiger et al., 2002). However, no significant differences were obtained for any of these predicted facets.

Matthew was rated as significantly lower on the anxiousness facet of neuroticism, the facets of openness to aesthetics and actions, and the extraversion facet of activity. However, as with the schizoid finding obtained for the DSM ratings, none of these differences are related to the HPD or APD diagnoses, which were the foci of the current investigation. Similarly, analyses of the FFM domains revealed that openness to experience ratings were significantly higher for Matthew than for Madeline, $t(62) = 3.16$, $p < 0.01$. Ratings for the remaining FFM domains were not significantly different for the two cases.

Effects of participant gender

In order to determine whether the gender of the participant had a significant effect on the case ratings, a set of multivariate ANOVAs (MANOVAs) were carried out and the interaction terms were analyzed. For the DSM-IV-TR, a 10 (PD) × 2 (case) × 2 (gender) MANOVA revealed a significant main effect for case, supporting the presence of bias in the PD ratings; $F(10, 49) = 2.69$, $p = 0.01$. The main effect for participant gender was non-significant, $F(10, 49) = 1.73$, $p = 0.10$, indicating that the gender of the therapist providing the ratings did not have an overall effect on the ratings. Finally, the case by gender interaction term was also non-significant, $F(10, 49) = 1.12$, $p = 0.37$. Furthermore, probes of the interaction for each 10 PD variables were also non-significant, indicating that the gender of the clinician providing the ratings did not interact with the finding of differences across the two cases.

A 30 (facet) × 2 (case) × 2 (gender) MANOVA of the FFM ratings revealed that unlike the DSM-IV-TR ratings, the main effect for case was non-significant, $F(30, 23) = 1.69$, $p = 0.10$. However, as with the DSM-IV-TR ratings, both the main effect for gender ($F(30, 23) = 0.82$, $p = 0.70$) and the case by gender interaction ($F(30, 23) = 1.12$, $p = 40$) were non-significant. Nonetheless, a probe of this interaction for the 30 FFM facets did reveal a significant value for the extraversion facet of activity, $F(1, 52) = 5.42$, $p = 0.02$. This finding suggests that the gender of the clinician providing the ratings did interact with the case being rated for the facet of activity. However, even this significant finding might be considered artefactual due to the lack of variability in the ratings provided. Every clinician (male and female) who rated Matthew, as well as all the female clinicians who rated Madeline provided a rating of '5' (i.e. the highest rating) for this particular facet, indicating a ceiling effect on this variable. Thus, although the male clinicians who rated Madeline provided quite high ratings, with a mean of 4.61, the lack of variability in the other cells caused the difference to be statistically significant.

Discussion

The findings of the current study indicate that clinicians were able to provide reliable ratings for the both DSM-IV-TR and FFM that agreed well with descriptions provided of Madeline in a previous study (Samuel & Widiger, 2006). However, the primary results of interest were whether these ratings for Madeline would be significantly

DOI: 10.1002/pmh

different when the gender of the case was experimentally manipulated. Consistent with past research, when the case was presented as male it was seen as significantly less histrionic, while the female case was seen as significantly less antisocial. These findings are consistent with prior gender bias studies of these PDs (Adler et al., 1990; Becker & Lamb, 1994; Fernbach et al., 1989; Flanagan & Blashfield, 2005; Ford & Widiger, 1989; Hamilton et al., 1986).

The novel contribution of the current study was the inclusion of FFM ratings and the ability to examine possible gender effects for this alternative model of PD. From the perspective of the FFM, each DSM-IV-TR PD can be conceptualized as an extreme or maladaptive variant of general personality structure. HPD, in particular, is characterized primarily in terms of facets of extraversion (warmth, gregariousness, excitement seeking and positive emotions), along with the neuroticism facets of depression and self-consciousness, the openness facets of fantasy and feelings, and the agreeableness facet of trust. The results of the current study indicated that when the case vignette was described in terms of the FFM conceptualization of this PD, no effect of sex on the ratings was obtained, suggesting that the FFM might be less prone to gender bias than the current DSM-IV nomenclature.

The findings of the current study are also consistent with an FFM meta-analytic review of sex differences amongst the PDs. Lynam and Widiger (2007) identified the sex differences amongst the DSM-IV PDs that would be expected if they were indeed maladaptive variants of the facets of the FFM. The authors then compared these expected sex differences with the observed sex differences reported in 32 published studies. The results of the meta-analysis indicated that the agreement between expected and observed sex differences was quite good for 8 of the 10 PDs. One of the two significant exceptions occurred for HPD.

From the perspective of the FFM, women and men should have similar rates of HPD (Lynam & Widiger, 2007). HPD, when described in terms of general personality functioning, should not yield a differential sex prevalence rate. Prior research has suggested that the diagnostic criteria for HPD bear a strong relation to caricatures of stereotypically feminine traits (e.g. Rienzi & Scrams, 1991; Sprock, Blashfield, & Smith, 1990), which some have suggested represent a biased characterization of women and femininity (Caplan, 1995; Chodoff, 1982; Kaplan, 1983; Sprock et al., 1990). The current results are consistent with this concern. The DSM-IV criterion set may too heavily weigh features that relate to stereotypically feminine behaviour, such as sexually seductive behaviour, overconcern with physical appearance, impressionistic speech and self-dramatization (Caplan, 1995; Kaplan, 1983), rather than describing the disorder in terms of more fundamental personality traits that may be less prone to gender biased applications.

In fact, concern over the extent to which the conceptualization and diagnosis of HPD represent cultural stereotypes rather than fundamental personality traits even predate DSM-III's publication in 1980. Chodoff (1974) suggested that 'the hysterical personality represents a caricature of femininity' (p. 1076) and 'young male residents may classify as a hysterical personality any reasonably attractive woman with whom they come into therapeutic contact' (p. 1076). The early origin of the diagnosis was explicitly gender specific as it referred to the presence of a malpositioned uterus (Chodoff, 1982).

The results of the current study were consistent with prior studies that suggest that the construct (Kaplan, 1983) and the diagnostic criteria (Sprock et al., 1990) continue to evoke a gender-specific stereotypic image. The FFM of PD is less prone to such a bias in part because each disorder is assessed with respect to component parts, rather than a global categorical construct (and one originally tied to a wandering womb). One strength of the FFM (and other dimensional models) is the allowance for a more precise description of any particular individual, rather than attempting to force him or her into categories that might be prone to

DOI: 10.1002/pmh

stereotypes. This finding provides additional evidence to support the belief that future versions of the diagnostic manual should adopt a dimensional approach to PD diagnosis. Gender-biased assessments and diagnoses can have a very pernicious and harmful effect on clinical assessment, conceptualization and treatment (Kaplan, 1983; Ross et al., 1995; Worell, in press). The possibility of reducing biases, whether related to gender, race, sexual orientation, or other variables, should be an important consideration for revisions to the diagnostic manual.

It is possible that the FFM holds no particular advantage in this area, relative to other dimensional models. However, it would be informative for future research to compare the DSM-IV-TR and the FFM along with other alternative dimensional models, such as those presented within the Schedule for Non-adaptive and Adaptive Personality (SNAP-2; Clark, Simms, Wu, & Casillas, in press), the Dimensional Assessment of Personality Pathology (DAPP; Livesley & Jackson, in press), the Temperament and Character Inventory-Revised (TCI-R; Cloninger, 1999), and the Shedler-Westen Assessment Procedure (SWAP-200; Shedler, 2002). It is quite possible that a description in terms of the personality traits included within these models, which lack any explicit reference to gender specific stereotypes (e.g. exhibitionism from the SNAP-2, extravagance from the TCI-R, and stimulus seeking from the DAPP), would fail to reveal any effects of gender in their application. The SWAP-200, however, does have a scale that refers to 'histrionic sexualization' (but it also goes by the alternative title of 'oedipal conflict').

Levels of the model hierarchy

In order to examine whether certain levels of a hierarchy may be more prone to potential biases, we also compared the two cases with respect to the clusters of the DSM-IV-TR and the domains of the FFM. The results of these analyses were somewhat surprising as there is little to no research that would suggest a gender bias within the way clinicians provide ratings of FFM openness to experience or DSM-IV-TR cluster A. Furthermore, the absence of a significant finding for cluster B could simply reflect the possibility that the subcomponents of this construct (e.g. HPD and APD) are biased in opposite directions and thus effectively cancelled each other out in the calculation of this cluster score. On the one hand, the lack of a gender bias within cluster B could be seen as a strength of the clusters, compared to the PDs themselves. However, to borrow a colloquial phrase 'two wrongs do not make a right'. Similarly, the presence of two, opposite gender biases that cancel one another out do not speak well for the validity of the DSM-IV-TR clusters. In one respect, this should not perhaps be surprising, as the three-cluster organization 'does not arise from any particular evidence' (Frances, 1980, p. 1052). It was largely an impressionistic organization serving more as a mnemonic device than representing any particular empirical or theoretical perspective (Millon, 1981). Nevertheless, quite a few studies are being conducted with respect to the DSM-IV-TR clusters (e.g. Ehrensaft, Cohen, & Johnson, 2006; Lenzenweger & Willet, 2007).

In sum, it appears that diagnostic ratings made at the level of the DSM-IV-TR PDs are particularly prone to biases. The task of assigning a single rating to heterogeneous categories might pull for global impressions that rely more heavily on a few casually central features of the disorder (Kim & Ahn, 2002), rather than the careful consideration of each diagnostic criterion. This would be consistent with previous research that has demonstrated that clinicians do not always closely adhere to the diagnostic criteria when assigning PD diagnoses to their clients (Morey & Ochoa, 1989; Blashfield & Herkov, 1996). Ford and Widiger (1989) had demonstrated that gender bias is considerably less when clinicians conduct assessments at the level of individual diagnostic criteria. Future studies that ask clinicians to assess both the DSM-IV-TR and FFM models at the lowest levels of their hierarchy might be additionally informative. For example, asking clinicians to complete the entire

NEO Personality Inventory—Revised (NEO PI-R) and rate the presence of each individual diagnostic criterion to determine if there would be gender differences at these, the lowest levels of each respective hierarchy.

Limitations

A potential limitation of the current study is the 15.5% response rate. However, this rate is typical for postal mail surveys and is consistent with previous studies within the literature (e.g. 17% in Crosby & Sprock, 2004; and 16% in Flanagan & Blashfield, 2005). Perhaps a more concerning limitation, with respect to external validity, is the reliance on the case vignette methodology. While the case vignette is the most commonly employed methodology for examining the potential of bias related to gender as well as race (Mikton & Grounds, 2007), it is not without flaws. The brief description that is inherent to a case vignette is something of a double-edged sword. It certainly standardizes the presentation and eases the collection of data, but also limits the amount of information that can be communicated about the individual being described. It is quite possible that clinicians would provide different ratings if they were able to interact with a real individual rather than simply read a brief case vignette.

Future studies may, in fact, explore the possibility of developing a computer program that allows a clinician to interact with a virtual client that answers live questions with programmed responses (Flanagan & Blashfield, 2005). Considering the advances in computer programming and animation, this could be functionally possible, albeit perhaps prohibitively expensive.

An additional limitation is the reliance on a single case vignette, which precludes comparison across cases. Without this, it is possible that the finding of non-significant differences between the ratings of the male and female case might actually reflect a limitation, rather than a strength of the model. For example, to the extent that individuals with differing personality pathology obtain similar FFM profiles, as suggested by Morey et al. (2002), then the results for the FFM from the current study could actually reflect an undesirable lack of discriminant validity, rather than a true lack of gender bias. However, because the case of Madeline has been used in previous studies alongside other vignettes with different personality pathology, it is possible to clarify the FFM's ability to discriminate amongst individuals. Samuel and Widiger (2006) asked clinicians to rate the case of Madeline, as well as two additional cases, in terms of the FFM facets and the DSM-IV-TR PDs. ANOVA tests revealed that 29 of the 30 FFM variables were, in fact, significantly different across the three cases (self-discipline from the domain of conscientiousness was the lone non-significant finding). Madeline, for instance, was much higher than Earnst in facets of extraversion (gregariousness, assertiveness and excitement-seeking), and higher than Ted in openness to values, impulsivity and positive emotions, but lower in deliberation. These findings support the notion that the FFM may have adequate discriminant validity to distinguish amongst individuals with PDs. Nonetheless, future research that replicates and extends the current study using additional vignettes would further address this concern.

Conclusions and clinical implications

The results of the current study, a survey of experienced members of the private practitioners division of the APA, have clear implications for the assessment and diagnosis of PDs within clinical practice. The ultimate goal of any diagnostic system is to provide valid information about clients (First et al., 2004). However, research over the past three decades has drawn into question the validity and equity of the DSM PDs by repeatedly suggesting that these diagnoses (particularly HPD) are prone to gender-biased applications. The current study extends these findings, but more importantly, it suggests that FFM descriptions of PD may be less prone to such biases. The potential for the FFM, or another alternative dimensional model, to

DOI: 10.1002/pmh

decrease gender bias provides further support for a transition to a dimensional classification that is integrated with models of general personality structure developed within psychology. For this reason, future research should continue the exploration of gender bias within additional alternative dimensional models of PD.

References

Adler, D., Drake, R., & Teague, G. (1990). Clinicians' practices in personality assessment: Does gender influence the use of DSM-III Axis II? *Comprehensive Psychiatry, 31*, 125–133.

American Psychiatric Association (APA). (1980). *Diagnostic and statistical manual of mental disorders* (3rd ed.). Washington, DC: APA.

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

Becker, D., & Lamb, S. (1994) Sex bias in the diagnosis of borderline personality disorder and posttraumatic stress disorder. *Professional Psychology: Research and Practice, 25*, 55–61.

Blashfield, R. K., & Herkov, M. J. (1996). Investigating clinician adherence to diagnosis by criteria: A replication of Morey and Ochoa (1989). *Journal of Personality Disorders, 10*, 219–228.

Boggs, C. D., Morey, L. C., Skodol, A.E., Shea, M. T., Sanislow, C. A., Grilo, C., McGlashan, T., Zanarini, M. C., & Gunderson, J. G. (2005). Differential impairment as an indicator of sex bias in DSM-IV criteria for four personality disorders. *Psychological Assessment, 17*, 492–496.

Caplan, P. J. (1995). *They say you're crazy: How the world's most powerful psychiatrists decide who's normal*. Reading, MA: Addison-Wesley.

Chodoff, P. (1974). The diagnosis of hysteria: An overview. *American Journal of Psychiatry, 131*, 1073–1078.

Chodoff, P. (1982). Hysteria and women. *American Journal of Psychiatry, 139*, 545–551.

Clark, L. A. (2007). Assessment and diagnosis of personality disorder. Perennial issues and an emerging reconceptualization. *Annual Review of Psychology*, *58*, 227–257.

Clark, L. A., Simms, L. J., Wu, K. D., & Casillas, A. (in press). *Manual for the Schedule for Nonadaptive and Adaptive Personality (SNAP-2)*. Minneapolis, MN: University of Minnesota Press.

Cloninger, C. R. (1999). *Temperament and Character Inventory—Revised (TCI-R)*. Unpublished Test. St Louis, Missouri: Washington University, Center for Psychobiology of Personality.

Costa, P. T., & McCrae, R. R. (1992). *Revised NEO Personality Inventory (NEO PI-R) and NEO Five-Factor Inventory (NEO-FFI) professional manual*. Odessa, FL: Psychological Assessment Resources.

Crosby, J. P., & Spock, J. (2004). Effect of patient sex, clinician sex, and sex role on the diagnosis of antisocial personality disorder: Models of underpathologizing and overpathologizing biases. *Journal of Clinical Psychology, 60*, 583–604.

Ehrensaft, M. K., Cohen, P., & Johnson, J. G. (2006). Development of personality disorder symptoms and the risk for partner violence. *Journal of Abnormal Personality, 115*, 474–483.

Fernbach, B. E., Wimstead, B. A., & Derlega, V. J. (1989). Sex differences in diagnosis and treatment recommendations for antisocial and somatization disorders. *Journal of Social and Clinical Psychology, 8*, 238–255.

First, M. B., Pincus, H. A., Levine, J. B., Williams, J. B. W., Ustun, B., & Peele, R. (2004). Clinical utility as a criterion for revising psychiatric diagnoses. *American Journal of Psychiatry, 161*, 946–954.

Flanagan, E. H., & Blashfield, R. K. (2003). Gender bias in the diagnosis of personality disorders: The roles of base rates and social stereotypes. *Journal of Personality Disorders*, *17*, 431–446.

Flanagan, E. H., & Blashfield, R. K. (2005). Gender acts as a context for interpreting diagnostic criteria. *Journal of Clinical Psychology, 61*, 1485–1498.

Ford, M., & Widiger, T. A. (1989). Sex bias in the diagnosis of histrionic and antisocial personality disorders. *Journal of Consulting and Clinical Psychology, 57*, 301–305.

Frances, A. J. (1980). The DSM-III personality disorders section: A commentary. *American Journal of Psychiatry, 137*, 1050–1054.

Garb, H. (2005). Clinical judgment and decision making. *Annual Review of Clinical Psychology, 1*, 67–89.

Hamilton, S., Rothbart, M., & Dawes, R. (1986). Sex bias, diagnosis, and DSM-III. *Sex Roles, 15*, 269–274.

Haslam, N. (2006). Bias in psychopathology research. *Current Opinion in Psychiatry, 19*, 625–630.

Howell, A. J., & Watson, D. C. (2002). Perceived dysfunction of male-typed and female-typed DSM-IV personality disorder criteria. *Journal of Personality Disorders, 16*, 536–548.

Kaplan, M. (1983). A woman's view of DSM-III. *American Psychologist, 38*, 786–792.

Kim, N. S., & Ahn, W. (2002). Clinical psychologists' theory-based representations of mental disorders predict their diagnostic reasoning and memory. *Journal of Experimental Psychology, 131*, 451–476.

Lenzenweger, M. F., & Willett, J. B. (2007). Predicting individual change in personality disorder features by simultaneous individual change in personality dimensions linked to neurobehavioral systems: The longitudinal study of personality disorders. *Journal of Abnormal Psychology, 116*, 684–700.

Livesley, W. J. (2003). Diagnostic dilemmas in classifying personality disorder. In K. A. Phillips, M. B. First, & H. A. Pincus (Eds.), *Advancing DSM. Dilemmas in psychiatric diagnosis* (pp. 153–190). Washington, DC: American Psychiatric Association.

Livesley, W. J., & Jackson, D. (in press). *Manual for the dimensional assessment of personality pathology-basic questionnaire*. Port Huron, MI: Sigma Press.

Lynam, D. R., & Widiger, T. A. (2001). Using the Five-Factor Model to represent the DSM-IV personality disorders: An expert consensus approach. *Journal of Abnormal Psychology, 110*, 401–412.

Lynam, D. R., & Widiger, T. A. (2007). Using a general model of personality to understand sex differences in the personality disorders. *Journal of Personality Disorders, 21*, 583–602.

McCrae, R. R., & Costa, P. T., Jr. (1999). A Five-Factor theory of personality. In L. A. Pervin, & O. P. John (Eds.) *Handbook of personality: Theory and research* (2nd ed., pp. 139–153). New York: Guilford Press.

Mikton, C., & Grounds, A. (2007). Cross-cultural clinical judgment bias in personality disorder diagnosis by forensic psychiatrists in the UK: A case-vignette study. *Journal of Personality Disorders, 21*, 400–417.

Millon, T. (1981). *Disorders of personality: DSM-Ill, Axis 11*. New York: Wiley.

Minnis, H., McMillan, A., Gillies, M., & Smith, S. (2001). Racial stereotyping: Survey of psychiatrists in the United Kingdom. *British Medical Journal, 323*, 905–906.

Morey, L., & Ochoa, E. (1989). An investigation of adherence to diagnostic criteria: Clinical diagnosis of the DSM-III personality disorders. *Journal of Personality Disorders, 3*, 180–192.

Morey, L., Warner, M., & Boggs, C. (2002). Gender bias in the personality disorders criteria: An investigation of five bias indicators. *Journal of Psychopathology and Behavioral Assessment, 24*, 55–65.

Morey, L. C., Gunderson, J. G., Quigley, B. D., Shea, M. T., Skodol, A. E., McGlashan, T. H., Stout, R. L., & Zanarini, M. C. (2002). The representations of borderline, avoidant, obsessive-compulsive, and schizotypal personality disorders by the Five-Factor Model. *Journal of Personality Disorders, 16*, 215–234.

Mullins-Sweatt, S. N., Jamerson, J. E., Samuel, D. B., Olson, D. R., & Widiger, T. A. (2006). Psychometric properties of an abbreviated measure of the Five-Factor Model. *Assessment, 13*, 119–137.

O'Connor, B. P. (2005). A search for consensus on the dimensional structure of personality disorders. *Journal of Clinical Psychology, 61*, 323–645.

Rienzi, B., & Scrams, D. (1991) Gender stereotypes for paranoid, antisocial, compulsive, dependent, and histrionic personality disorders. *Psychological Reports, 69*, 976–978.

Ross, R., Frances, A. J., & Widiger, T. A. (1995). Gender issues in DSM-IV. In J. M. Oldham and M. B. Riba, *Review of psychiatry* (Vol 14, pp. 205–226). Washington, DC: American Psychiatric Press.

Samuel, D. B., & Widiger, T. A. (2004). Clinicians' descriptions of prototypic personality disorders. *Journal of Personality Disorders, 18*, 286–308.

Samuel, D. B., & Widiger, T. A. (2006). Clinicians' judgments of clinical utility: A comparison of the DSM-IV and Five-Factor Models. *Journal of Abnormal Psychology, 115*, 298–308.

Samuel, D. B., & Widiger, T. A. (2008). A meta-analytic review of the relationships between the Five-Factor Model and DSM-IV-TR personality disorders: A facet level analysis. *Clinical Psychology Review, 28*, 1326–1342. doi:10.1016/j.cpr.2008.07.002.

Shedler, J. (2002). A new language for psychoanalytic diagnosis. *Journal of the American Psychoanalytic Association, 50*, 429–456.

Spitzer, R. L., First, M. B., Shedler, J., Westen, D., & Skodol, A. E. (2008). Clinical utility of five dimensional systems of personality diagnosis: A 'consumer preference' study. *The Journal of Nervous and Mental Disease, 196*, 356–374.

Sprock, J., Blashfield, R. K., & Smith, B. (1990). Gender weighting of the DSM-III-R personality disorder criteria. *American Journal of Psychiatry, 147*, 586–590.

Strack, S. (2005). An integrationist perspective on personality assessment. *Journal of Personality Assessment*, *84*, 105–107.

Verheul, R. (2005). Clinical utility for dimensional models of personality pathology. *Journal of Personality Disorders, 19*, 283–302.

Warner, R. (1978). The diagnosis of antisocial and hysterical personality disorders. *Journal of Nervous and Mental Disease, 166*, 839–845.

Widiger, T. A. (1998). Sex biases in the diagnosis of personality disorders. *Journal of Personality Disorders, 12*, 95–118.

Widiger, T. A. (2007). DSM's approach to gender: History and controversies. In W. E. Narrow, M. B. First, P. J. Sirovatka, & D. A. Regier (Eds.), *Age and gender considerations in psychiatric diagnosis*. A research agenda for DSM-V (pp. 19–29). Washington, DC: American Psychiatric Association.

Widiger, T. A., & Costa, P. T., Jr. (2002). Five-Factor Model personality disorder research. In P. T. Costa, Jr., & T. A.

Widiger (Eds.) *Personality disorders and the Five-Factor Model of personality* (2nd ed., pp. 59–87). Washington, DC: American Psychological Association.

Widiger, T. A., & Trull, T. J. (2007). Plate tectonics in the classification of personality disorder: Shifting to a dimensional model. *American Psychologist*, *62*, 71–83.

Widiger, T. A., Trull, T. J., Clarkin, J. F., Sanderson, C. J., & Costa, P. T., Jr. (2002). A description of the *DSM-IV* personality disorders with the Five-Factor Model of personality. In P. T. Costa Jr., & T. A. Widiger (Eds.), *Personality disorders and the Five-Factor Model of personality* (2nd ed., pp. 89–99). Washington, DC: American Psychological Association.

Wiggins, J. S. (2003). *Paradigms of personality assessment*. New York: Guilford.

Worell, J. (in press). Issues in clinical assessment with women. In J. Butcher (Ed.), *Oxford handbook of personality assessment*. New York: Oxford University Press.

Address correspondence to: Douglas B. Samuel, Department of Psychology, University of Kentucky, Lexington, KY, 40506-0044. Email: dsamuel@uky.edu