Master's Theses and Doctoral Dissertations

Master's Theses, and Doctoral Dissertations, and Graduate Capstone Projects

2014

Measurement Invariance Between Genders on Two Measures of Borderline Personality Disorder

Amy Paggeot

Measurement Invariance Between Genders on Two Measures of

Borderline Personality Disorder

by

Amy Paggeot

Master's Thesis

Submitted to the Department of Psychology

Eastern Michigan University

in partial fulfillment of the requirements

for the degree of

MASTER OF SCIENCE

in Clinical Psychology

Thesis Committee:

Steven Huprich, PhD, Chair

Tamara Loverich, PhD

Stephen Jefferson, PhD

May 8, 2014

Ypsilanti, Michigan

Abstract

Gender bias in the diagnosis and assessment of Borderline Personality Disorder (BPD) has been the subject of much controversy in the psychological research literature. Evidence regarding differential prevalence rates and diagnostic rates in clinical settings between genders has been mixed, and measurement bias is one potential explanation for these mixed findings. Measurement invariance (MI) is considered a prerequisite for comparing group means on any measure or latent construct. Two structured clinical interviews of BPD were subjected to sequential CFAs to evaluate MI between genders. The SCID-II BPD Scale (First, Gibbon, Spitzer, Williams, & Benjamin, 1997) was found to be measurement invariant between genders, while the PDI-IV BPD Scale (Widiger, Mangine, Corbitt, Ellis, & Thomas, 1995) did not meet criteria for strict measurement invariance. Gender differences were then examined using the SCID-II BPD Scale, and no significant gender differences were found in diagnosis or overall criteria endorsed. However, in the rates of endorsement of individual criteria, differences were found in the unstable relationships item. Implications and study limitations are discussed.

Table of Contents

List of Tables

Literature Review

Introduction

Borderline personality disorder (BPD) is a serious psychological disorder, characterized by a pervasive pattern of affective instability, impulsivity, self-harm behaviors, identity disturbance, and chaotic interpersonal relationships (American Psychiatric Association, 2000; Distel, Hottenga, Trull & Boomsma, 2008; Domes, Schulze, & Herpertz, 2009; Selby & Joiner, 2009; Tragesser, Solhan, Schwartz-Mette, & Trull, 2007). It is the most commonly diagnosed personality disorder in clinical settings, and individuals with BPD have significant functional impairments and a high rate of death by suicide (Selby & Joiner, 2009). One of the most controversial aspects of BPD is the common finding of a differential rate of BPD between the genders; specifically, BPD is more prevalent among women than men (American Psychiatric Association, 2000). A number of researchers have questioned whether this finding reflects true differences in the etiology and pathology of the disorder, or whether this finding reflects a bias in the diagnostic construct, the assessment methods, or the application of these assessment methods in the diagnosis of BPD (Widiger, 1998). The literature review that follows will explore the findings regarding differential rates of BPD among men and women, some of the proposed explanations for these findings including sex bias, and will review analyses of measurement invariance as a way to explore whether the construct of BPD is being measured in the same way between men and women. Subsequently, a study is proposed to analyze measurement invariance with two structured clinical interviews on a mixed clinical and nonclinical population, to find whether there is measurement invariance between the genders on these two measures of BPD. Those measures which meet criteria for measurement invariance will then be used to examine

differences between men and women in BPD diagnosis, number of criteria endorsed, and

differences with regards to endorsement of individual criteria.

Gender Bias in Personality Disorder Diagnosis

Gender differences in the rates of BPD have been cited in the most recent edition of the

Diagnostic and Statistical Manual of Mental Disorders (DSM-IV TR; American Psychiatric

Association, 2000). These diagnostic rates – prevalences—and their gender differences have

been noted for several other personality disorders as well; for example, both Dependent and

Histrionic Personality Disorders are said to also be more prevalent among females than males

(Widiger, 1998). Widiger (1998) has outlined six possible explanations for different ways in

which gender differences in prevalence rates might reflect a sex bias in personality disorder

diagnosis (see Table 1).

Table 1

Types of Sex Bias (Widiger, 1998)

Type of Sex Bias	Description
Diagnostic Construct	The diagnostic construct itself is biased, and reflects a sexist characterization of one gender.
Diagnostic Thresholds	Different thresholds are applied to male-typed versus female-typed personality disorders.
Application of Diagnostic Criteria	Diagnostic criteria are applied differently to women and men by clinicians.
Biased Population Sampling	The population from which a sample is drawn has more of one gender than another, therefore more individuals of one gender are assigned the diagnosis.
Assessment Instruments	The assessment instruments being used carry a systemic sex bias of their own, and will apply a diagnosis to more individuals of one gender than of another.
Diagnostic Criteria	One or more of the diagnostic criteria might carry a sex bias, and might be applicable more to one gender than another.

First, the diagnostic construct itself may be biased. That is, the diagnostic criteria for BPD as currently conceptualized could reflect an implicit bias that tends to characterize stereotypically feminine characteristics as pathological. Second, the diagnostic thresholds may vary across disorders more prevalent in males versus those more prevalent in females; that is, different thresholds may be applied for those personality disorders often seen as more prevalent in men versus those seen as more prevalent in women. For example, it may generally be easier to receive a "female-typed" personality disorder diagnosis, such as Histrionic Personality Disorder or BPD, than a "male-typed" personality disorder, such as Antisocial Personality Disorder. Thus, personality disorder diagnoses in general may be more frequently applied to women than to men because those that are found more frequently in women have lower thresholds. Third, the application of the diagnostic criteria by clinicians may be biased; this reflects the way in which criteria are applied by clinicians to men versus women. For example, a clinician might more readily apply the diagnostic criterion of "intense, inappropriate episodes of anger" of BPD to a woman than to a man, though the criterion objectively may not differ between genders. Fourth, there may be a natural bias that occurs when sampling a given population. Specifically, if there are more women than men in a given population, then more women than men would be given any diagnosis, including BPD, simply by chance. Fifth, the assessment instruments used may carry a systemic sex bias of their own. For example, an item of an inventory meant to assess BPD might be more readily endorsed by women than men, perhaps because of its wording or some type of response bias, despite reflecting no significant functional impairment that one would expect to see with this type of psychopathology. Sixth, the diagnostic criteria themselves may be biased. This form of bias is particularly problematic to identify and address. For instance, if the criteria described behavior associated with a feminine

gender role, such behavior might be considered normative in women and pathological in men, leading women displaying the behavior to be undiagnosed and men displaying the behavior to be pathologized (Widiger, 1998).

Others have noted that it is possible that differential rates in the diagnosis or prevalence of BPD between genders simply reflects the true state of affairs, rather than any kind of sex bias at all (e.g. Lynam & Widiger, 2007). For example, Linehan's (1993) biopsychosocial model of BPD etiology suggests that the escalated rate of BPD among women is a consequence of several factors, such as escalated rates of sexual abuse among women, invalidating cultural ideals for women, and sexism aimed at girls who fail to be appropriately "feminine" or whose talents are in areas considered to be "masculine" such as mathematics and mechanics (pp. 52-56). Thus, women are more likely to encounter these forms of invalidation, and because invalidation is a critical factor leading to the development of BPD within Linehan's (1993) model, women are more likely to develop BPD in adulthood.

To examine the issue of whether or not a sex bias exists in BPD diagnosis, and if so, what kind of sex bias it is, one must start by establishing first whether a difference in prevalence rates between the sexes exists. The DSM-IV-TR (APA, 2000) reports a difference in prevalence rates of 3 females receiving a diagnosis for every male diagnosed. A meta-analysis of 75 studies found that about 75% of those diagnosed with BPD in clinical settings are female (Widiger & Trull, 1993), consistent with APA reports.

However, non-clinical samples produce mixed results. Several large, representative epidemiological samples have found no difference in prevalence between men and women. For example, in a sample of 34,653 individuals, Grant, Chou, Goldstein, Huang, Stinson, Jackson, and Burgess (2008) found no statistically significant difference in prevalence rates of BPD

between men and women who were assessed with the Alcohol Use Disorder And Associated Disabilities Interview Schedule for DSM-IV (AUDADIS-IV; Grant, Dawson, & Hasin, 2001), though they did find a trend weighted in favor of women (5.6% to 6.2%, respectively). Trull, Jahng, Tomko, Wood, and Sher (2010) re-analyzed the same data set used by Grant et al. (2008) and required each criterion to be associated with distress or impairment before it could be used to count towards a PD diagnosis; using this technique, the authors found a significantly greater prevalence rate of BPD in women than men (3.0% versus 2.4%, respectively). Lenzenweger, Lane, Loranger, and Kessler (2007) also evaluated a large, representative sample within the United States ($n = 9,282$) who were assessed with the International Personality Disorder Examination (IPDE; Loranger, 1999), but found no significant gender differences in BPD's diagnostic frequency. Torgersen et al. (2001) also evaluated the diagnostic frequency of BPD in a Norwegian sample of 2,053 individuals who were interviewed with the Structured Interview for DSM-III-R Personality Disorders (SIDP-R; Pfohl, Blum, Zimmerman, & Stangl, 1989). They found BPD was more prevalent in women than men (0.9% vs. 0.4%, respectively, showing that women were twice as likely to be diagnosed as men); however, this difference did not reach statistical significance.

Further complicating this debate are the findings of Coid et al. (2006). Specifically, they evaluated a sample of 638 individuals diagnosed with BPD using the Structured Clinical Interview for Personality Disorders (SCID-II; First et al., 1997). Their analyses revealed a BPD diagnostic gender bias weighted in favor of men rather than women (5.6 % vs. 4.2%), a reversal of the trend found in previously cited studies. However, this difference was not statistically significant.

Upon reviewing this literature, it appears that most of the empirical evidence suggests that some kind of sex bias may exist which affects the assignment of a BPD diagnosis. Notably, clinical samples appear to display a bias toward diagnosing women with BPD more frequently than men, while epidemiological samples more frequently find nonsignificant trends, which sometimes are consistent with this bias, though sometimes fluctuate in the opposite direction. The reason for these trends, and explanations for differences between samples, remains to be answered.

Some authors have sought to explore the issue of gender bias in BPD diagnosis by examining clinicians' application of diagnostic criteria through the use of case vignettes. For example, Becker and Lamb (1994) asked clinicians from a variety of backgrounds (i.e., psychiatrists, psychologists, and clinical social workers) to rate one of three slightly varying vignettes, labeled as either male or female, according to how applicable a variety of diagnoses were, including BPD. The vignettes were left deliberately vague in that the authors made an effort to keep the vignettes from clearly meeting criteria for either of the disorders of interest, those being BPD and posttraumatic stress disorder (PTSD). The authors found a significant main effect for the sex of the subject in the vignette, such that female vignettes were rated significantly higher for the applicability of a BPD diagnosis than were male vignettes.

Woodward, Taft, Gordon, and Mais (2009) conducted a similar study examining clinicians' diagnosis of case vignettes. Using slightly modified vignettes originally created by Becker and Lamb (1994), they counterbalanced the order of the presentation of the abuse history in order to examine the effect of presentation order, and took out a reference to excessive jealousy, as this did not fit either BPD or PTSD. They also replaced a suicidal threat with a reference to chronic insomnia. The authors found no significant differences in diagnosis

between case genders and, therefore, found no support for clinician bias in regards to gender and

BPD. Flanagan and Blashfield (2003) also utilized vignettes to study clinician bias in BPD

diagnosis. The authors selected novice undergraduate students who had not been exposed to

diagnostic information yet, and manipulated instructional information about cluster B personality

disorder diagnosis base rates between genders in three studies. They reasoned that this

manipulation would offer greater insight into the origins of clinician bias as discussed by

Widiger (1998), specifically with regard to gender bias in Cluster B personality disorder

diagnoses. Novices were taught about the Cluster B PDs by using paragraphs of information

taken from the DSM-IV (American Psychiatric Association, 1994). The authors checked the

novices' learning of the information through a computerized task where they were asked to

associate relevant traits based on the diagnostic criteria with the correct PD.

In the first study, the novices were not given any base rate information during the

learning phase; then, they were presented with vignettes meant to be prototypes of Narcissistic,

Antisocial, and Histrionic Personality Disorders, as well as two vignettes with mixed narcissistic

and histrionic features. The vignettes were either gender-neutral or described as male or female.

They were asked to rate these vignettes on a 1 to 7 scale of how well each vignette represented

each of the four Cluster B personality disorders and their associated traits. There were no

significant gender differences found in male or female case vignette trait ratings. In the second

study, base rate information was provided in the expected or stereotypic direction, and three case

vignettes were added to the original five. The new vignettes were a prototypical Antisocial PD

case, a prototypical BPD case, and an additional Histrionic PD case. Gender differences were

found for Borderline trait ratings given to the prototypical Antisocial PD case and one of the

mixed Antisocial and Histrionic feature vignettes such that women were rated higher for

Borderline traits, though this was not found in the prototypical BPD case vignette. In the third study, novices were taught base rate information contrary to the expected or stereotypic direction. In this case, the novices displayed greater difficulty in learning the information, particularly for BPD and Antisocial Personality Disorder traits, and only displayed gender bias for Histrionic Personality Disorder and Narcissistic Personality Disorder ratings. These associations were smaller than when base rate information was consistent with expected or stereotypic patterns in the second study. The authors concluded that novices rated cases in accordance with known base rates, but that the consistency and magnitude of this association was affected by correspondence with stereotypic associations.

Further complications regarding gender bias present themselves when BPD gender differences are examined at the individual criterion or symptom level. In the Collaborative Longitudinal Personality Disorders Study (CLPS), Johnson et al. (2003) did not find a significant difference in the mean number of BPD criteria endorsed by men and women using the Diagnostic Interview for DSM-IV Personality Disorders (DIPD-IV; Zanarini, Frankenburg, Sickel, & Young, 1996),; however, they did find a difference in the rate of endorsement in one criterion, (identity disturbance), such that more female participants were rated by interviewers as endorsing this criteria than male participants. In contrast to this finding, De Moor et al. (2009) and McCormick et al. (2007) found the dissociation criterion to be significantly endorsed by more female participants than male participants on both the PAI-BOR (Morey, 1991) and DIPD-IV (Zanarini et al, 1996), respectively, unlike any of the other eight criteria for BPD in both studies.

Boggs et al. (2009) also examined gender bias at the level of the diagnostic criteria by examining the relationship of each of the criteria and functional impairment in both men and

women, to determine if the criteria were differentially predictive of impairment between genders. They found that eight of the nine BPD criteria were differentially predictive of impairment between gender, such that relying on any of these criteria for diagnosis for women would lead to an underestimate of functioning in women relative to men. In other words, women who were considered to have endorsed or to be "positive" for any of these criteria would, on average, be functioning better than a man who was considered to have endorsed the same criteria. The sole exception to this finding was the BPD criterion related to impulsivity.

The previous findings, while mixed, seem to suggest that one or all of the forms of sex bias outlined by Widiger (1998) could be present in the assessment of BPD; however, the most fundamental bias would be that the diagnostic construct itself is biased. Testing this hypothesis empirically is difficult and confounded with the method of assessment and the other forms of bias outlined by Widiger (1998). Specifically, it is almost impossible to detach the assessment tool from the criterion itself, hence making it challenging to determine if the instrument itself is biased, or whether it is the construct or the criterion. One can compare two separate assessment tools, but when findings are mixed, as has typically been the case with BPD assessment, it is possible that one or both assessment instruments are biased, which makes it difficult to determine which kind(s) of bias exist, and for which instrument.

Measurement

One way in which researchers can begin to statistically determine whether a given assessment instrument is accurately reflecting differences in a latent construct is to examine measurement invariance (MI). Measurement invariance on a given measure is considered a prerequisite to making comparisons between groups. In fact, it has been suggested that conclusions drawn from other comparative analyses may be biased or invalid if the measures do

not have the same meaning across groups; that is, conclusions regarding mean differences

between groups on a given measure are invalid if the underlying latent construct that the measure

purports to assess is different across groups (Vandenberg & Lance, 2000). Thus, in order to

make meaningful interpretations of group differences, one must first establish MI, which is not

evaluated by mean differences, but rather typically by examining differences in the pattern of

variation across groups.   This allows for a more conceptual-level comparison that should not be

sensitive to mean score differences between groups, which is unable to distinguish between true

differences between the groups or some other form of bias (Chen, 2007; Slof-Op 't Landt, van

Furth, Rebollo-Mesa, Bartels, van Beijsterveldt, Slagboom, et al., 2009; Vandenberg & Lance,

2000). Statistically, MI is commonly assessed through a confirmatory factor analysis (CFA)

framework, using tests of factorial invariance in a sequence determined by the results of each test

(Chen, 2007; Vandenberg & Lance, 2000). Factorial invariance refers to the procedure of

comparing the similarity of factor loadings across groups in order to ascertain whether or not the

items of an assessment tool are loading in the same direction and magnitude, thus allowing

researchers to determine whether they are measuring the same latent construct in both groups

(Chen, 2007; Slof-Op 't Landt, et al.,   2009). Typically, a measure is considered to have

factorial invariance when comparisons of increasingly restricted CFA models indicate adequate

fit across groups.

MI and Measures of BPD

De Moor et al. (2009) examined MI with respect to both age and sex in BPD with a

single measure, the Personality Assessment Inventory – Borderline Scale (PAI-BOR; Morey,

1991), a self-report measure of BPD symptomatology. They did this by dividing a large sample

(n = 8,527) collected from the Netherlands Twin Registry into four groups: young male and

female groups (ages 18-35) and older male and female groups (ages 36-90). The authors found that the PAI-BOR was measurement invariant across both sex and age, although they did find that women generally scored higher on the affective instability, impulsivity, and negative relationships subscales of the PAI-BOR.

Aggen, Neale, Roysamb, Reichborn-Kjennerud, and Kendler (2009) also examined MI with respect to age and sex, using the SIDP-IV (Pfohl, Blum, & Zimmerman, 1995), collected from a large epidemiological sample of Norwegian twins ($n = 2,794$). In this study, the authors included age as a continuous variable in the analyses rather than dividing the sample into groups. The authors found that the SIDP-IV BPD criteria were not measurement invariant. In particular, they found that the impulsivity and affective lability criteria functioned differently between sexes; specifically, affective instability was reported more often in women, given the same BPD factor level or trait level; similarly, impulsivity was reported to be higher in men than in women, and this difference increased disproportionately with age. The authors of this study also found that women reported higher overall mean levels of BPD symptomatology.

This finding is particularly notable given the importance of both affective instability and impulsivity to the construct of BPD. Both of these traits have been theorized to be core underlying features of the disorder, either together or separately, by numerous researchers (e.g. Koenigsberg et al., 2001; Links et al., 1999; Tragesser & Robinson, 2009; Tragesser, Solhan, Schwartz-Mette, & Trull, 2007). For example, affective instability has been shown to be independently predictive of BPD features assessed two years later (Tragesser et al., 2007) and suicide attempts in BPD patients (Yen et al., 2004). Impulsivity has also been shown to be independently predictive of BPD features 7 years later (Links et al., 1999). Both traits together have been shown to be predictive of suicidal behavior and total BPD features (Koenigsberg et

al., 2001; Tragesser & Robinson, 2009). Crowell, Beauchaine, and Linehan (2009) theorize that

poor impulse control and emotional sensitivity are early biological vulnerabilities for BPD,

which interact with various environmental influences in the development of BPD in adolescents

and adults. If there were a gender difference or a gender bias in either or both of these traits, as

was found by Aggen et al. (2009), this would have important theoretical implications for the

meaning of gender differences in BPD diagnosis. For example, it could be suggested that a

gender difference in emotional sensitivity or impulsivity might be responsible for overall gender

differences in BPD diagnosis. This would lend support and justification for true differences in

differential rates of BPD diagnosis, rather than gender differences as a result of bias.

Given the importance of MI in determining whether there is gender bias in a latent

construct, and the fundamental importance of MI to establishing whether other forms of gender

bias might be present, clarifying whether or not BPD is measurement invariant is an important

element in the ongoing controversy surrounding the issue of gender bias in BPD. Moreover, MI

should be established across multiple measures of assessment, in order to diminish the possibility

that MI (or the lack thereof) is solely a function of the assessment tool. With this in mind, the

present study seeks to use a pre-existing data set to examine MI across gender in two structured

clinical interviews of patients being seen in outpatient mental health care treatment clinics and

undergraduate students in a university setting. Differences will also be explored using two self-

report measures of affective instability and impulsivity. These traits are particularly relevant,

given that Aggen et al. (2009) found these two traits to lack MI between genders when using

other measures, and thus differences in these traits might provide a theoretical explanation for

any gender differences found in the analyses. MI will first be established through tests of

factorial invariance within a CFA framework for each measure, then mean score differences will be examined using *t*-tests for each measure that has demonstrated satisfactory MI.

Methods

Participants

Participants were recruited from two populations across five locations. Nonclinical, undergraduate students were recruited from psychology courses at a Midwestern university between 2006-2008. Psychiatric outpatients were recruited from four locations between 2007-2009: a university-based psychology clinic; a hospital-based outpatient behavioral health treatment facility; and two offices of a community mental health center. Institutional review boards at all facilities reviewed and approved the study. Therapists at all outpatient locations were informed about the study and provided an information sheet to give to potential participants. To be included in the study, participants had to range between 18-75 years-old and could not be actively psychotic, or have a primary diagnosis of schizophrenia, major depression with psychosis, bipolar disorder with psychotic features, or psychotic disorder NOS. Their mental status had to be intact, as judged by their clinician, and they could not have an organic or medical condition that accounted for their diagnosis. They also could not be actively using or abusing substances, nor in such acute psychological distress that their therapists believed answering questions about their mood, thoughts, relationships, and personality would evoke a strong negative emotional reaction. In other words, based upon the ethical principle of nonmaleficence, participants were ruled out by their therapist if it was believed that asking such questions would destabilize their therapeutic progress or general functioning. Though 192 patients were referred and participated in the study, a review of their primary diagnoses after

participation identified 10 individuals who had an exclusionary clinician-assigned diagnosis.

Hence, their data were excluded.

Procedures

After being informed about the study, potential participants contacted the Principal

Investigator, Dr. Steven Huprich, by phone or email expressing their interest. They were

referred to one of three doctoral student research assistants, who rotated their responsibility in

scheduling participants. The assistants did an initial screening on the phone with the participant

to determine if s/he met eligibility criteria and scheduled to meet them at a time that was

mutually convenient with one of six graduate student research assistants. At one of the

community mental health center locations, a staff member coordinated the patient scheduling, as

it was more feasible to manage the scheduling in the center this way. Study participation took

approximately two hours and involved administration of diagnostic questionnaires, as well as a

completion of a set of self-report questionnaires. For their participation, clinical participants

received $75 cash. Undergraduate participants were granted extra credit by their instructors in

an undergraduate psychology course.

Measures

Structured Clinical Interview for DSM-IV-TR Personality Disorders - Borderline

Scale (SCID-II; First, Gibbon, Spitzer, Williams, & Benjamin, 1997). The SCID-II Borderline

Scale is a nine-item, structured clinical interview where each of the nine items is based on one of

the nine criteria for BPD as defined in the DSM-IV-TR (APA, 2000). Each item is scored from

1 to 3, with 1 indicating the absence of the given criteria, 2 indicating the presence of the criteria

at a subthreshold level, and 3 indicating the presence of the criteria at the given threshold. The

scale had satisfactory internal consistency for the sample ($\alpha = 0.85$) and has demonstrated good

convergent validity (Ryder, Costa, & Bagby, 2007). In the present study, the scale has also

demonstrated satisfactory inter-rater item reliability, with ICC values for the current sample

ranging from .83 - .98 and averaging .92 for the items.

        Personality Disorder Interview for DSM-IV-TR (PDI-IV; Widiger, Mangine, Corbitt,

Ellis, & Thomas, 1995). The PDI-IV Borderline scale is a nine-item, structured clinical

interview where each of the nine items is based on one of the nine criteria for BPD as defined in

the DSM-IV-TR (APA, 2000). Each item is scored on a three-point scale from 0 to 2, where 0

indicates the absence of a criterion, 1 indicates the presence of a criterion, and 2 indicates the

respondent exceeds the criterion. In the present study, the scale had satisfactory internal

consistency for the sample (α = .84) and demonstrated satisfactory inter-rater item reliability,

with ICC values for the current sample ranging from .64 to 1.00 and averaging .80 for the items.

        Affective Lability Scale (ALS; Harvey, Greenberg, & Serper, 1989). The ALS is a 54-

item self-report measure designed to assess affective lability. The ALS consists of six subscales

examining shifts in mood from euthymia to depression, elation, anger, and anxiety, as well as

shifts between elation and depression, and between anxiety and depression. Items are rated on a

four-point scale ranging from 0 (Very Undescriptive) to 3 (Very Descriptive). The ALS has

been found to have satisfactory internal consistency (α ranging from .72 to .89), satisfactory test-

retest reliability over the period of a month (r ranging from .48 to .86), and good divergent

validity with measures of affective intensity and depression. All six subscales have been found

to correlate significantly with each other, suggesting a unitary factor structure (Harvey et al.,

1989). The scale was found to have very good internal consistency in the present sample as a

whole (α = .96) as well as in each subsample (clinical, α = .96; non-clinical, α = .96).

Barratt Impulsiveness Scale-11 (BIS-11; Stanford, Mathias, Dougherty, Lake,

Anderson, & Patton, 2009). The Barratt Impulsiveness Scale 11th Revision (BIS-11) is a 30-item

self-report measure of impulsiveness. It consists of questions designed to be answered on a four-

point Likert-type scale (Rarely/Never, Occasionally, Often, Almost Always/Always) and has

three, second-order subscales: Attentional Impulsivity, Motor Impulsivity, and Non-Planning

Impulsivity. It was found to have satisfactory internal consistency in the present sample (entire

sample, $\alpha = .82$; clinical, .82; non-clinical, .81) as well as for each of the subscales (Attentional

Impulsivity $\alpha = .66$ [clinical, $\alpha = .66$; non-clinical, $\alpha = .67$]; Motor Impulsivity $\alpha = .60$ [clinical,

$\alpha = .60$; non-clinical, $\alpha = .62$]; Non-Planning Impulsivity $\alpha = .69$ [clinical, $\alpha = .68$; non-

clinical, $\alpha = .68$]). Because a previous version of this measure, originally published by the

authors, was employed for data collection, some items were excluded in order to align more

closely with the final BIS-11 structure[1]. The BIS-11 has also demonstrated significant between-

groups differences in the expected directions across undergraduate and clinical populations, and

good convergent validity with other self-report measures of trait impulsivity (Stanford et al.,

2009).

Beck Depression Inventory-II (BDI; Beck, Steer, & Brown, 1996). The BDI-II is a 21-

item, four-point Likert questionnaire that assesses thoughts, feelings, behaviors, and physical

symptoms associated with major depressive disorder in individuals age 13 or older. Each item

has four responses, ranging from 0 to 3, in which 0 indicates the absence of a depressive

symptom and 3 indicates the presence of the symptom in its most severe form. Individuals are

---

[1] The authors used a correction published by Dr. Marijn Lijffijt online at
http://impulsivity.org/BIS-11/bis-11%20file%20storage/BIS-
11A_to_BIS11%20prorating%20procedure.doc. The International Society for Research on
Impulsivity has stated that this procedure is "a reasonable way to deal with BIS-11A data".

asked to rate each of these items for their applicability over the past two weeks. The 25-year

reliability and validity of the BDI-II is well documented (Beck et al., 1996). In the present study,

Cronbach's alpha was computed to be .95.

Bell Object Relations and Reality Testing Inventory (BORRTI; Bell, 1995). The

BORRTI is a 45-item, true-false measure that assesses four dimensions of an individual's object

relations: Alienation, Insecure Attachment, Egocentricity, and Social Incompetence. Cronbach's

alpha and split-half reliabilities are reported for all four scales and range from .78 to .90.

Twenty-six-week test-retest reliabilities across the scales range between .58 and .85. A number

of validation studies are reported by Bell (1995) and Huprich and Greenberg (2003). Because of

the complex scoring algorithm, Cronbach's alpha could not be computed in the current study.

Analyses

All four measures were examined for MI using the procedure outlined by Slof-Op 't

Landt et al. (2009). The procedure involves fitting increasingly restrictive CFA models to the

data, culminating in a highly constrained model, in order to establish MI. In the first step, a

saturated model is fitted to the data to obtain estimates of item thresholds and the polychoric

correlation among items. In the second step, items are evaluated for unidimensionality between

men and women; the items will be regressed onto a single common factor without any equality

constraints between genders, though thresholds in men will be constrained to equal those in

women to obtain a common metric. In the third model, factor loadings are constrained to be

equal over gender. In the fourth model, mean liabilities (intercepts) in the men are constrained at

zero, and the common factor mean estimated. Finally, in the fifth model, a final restraint is

added: the variance in the separate items not explained by the common factor will be constrained

to be equal between men and women. It is this model, the fifth model, where full measurement

invariance is represented. If this final model displays a reasonable fit, then the measure is determined to have established full measurement invariance. This is determined by several fit statistics: the hierarchical chi-square, the comparative fit index (CFI) and the residual mean square error of approximation (RMSEA). The chi-square is a difference test and produces a p value, and as such, a p value of $< .05$ will be used as a cutoff. CFI values of > 0.95 are generally considered an acceptable fit (Slof-Op 't Landt et al., 2009) and it has been suggested that an RMSEA of < 0.06 should be considered acceptable in factorial invariance analyses as well (Schermelleh-Engel, Moosbrugger, & Muller, 2003). This would suggest that the construct being measured is the same in both genders, and therefore can be meaningfully compared using t-tests. If both measures of BPD and the two measures of traits underlying BPD are measurement invariant, this would further suggest gender differences may be due to true differences in the etiology and pathology of the disorder. If measurement invariance is not found for any of the measures, it would suggest that the diagnostic construct itself is not being measured the same between men and women, and this form of sex bias is present. Finally, if measurement invariance is found for only one or some of the measures, one of the other forms of sex bias outlined by Widiger (1998) may be present.   As a final set of analyses, t-tests between each measure that has established MI will be conducted in order to examine any meaningful sex differences. No *a priori* hypotheses will be made, given the inconsistency in similar studies and overall lack of research to guide such hypotheses.

Results

Sample Characteristics

The sample consisted of a total 340 participants, 169 of whom were psychiatric outpatients and 171 of whom were undergraduate students. The combined sample had a mean age of 31.61 (SD = 15.07) years, and consisted of 234 female participants and 97 male participants. Nine individuals did not report their gender, all of whom were in the clinical sample. Within the undergraduate sample, participants ranged in age from 18 to 59, with a mean of 21.91 (SD = 6.23) years. There were 124 female and 47 male participants. Participants identified themselves as Caucasian (69%; n = 235), African American (18%; n = 60), Asian (3%; n = 10), Hispanic (3%; n = 11), Middle Eastern (1%; n = 4), and other (5%; n = 16). Four participants did not report their ethnic heritage. The majority of the undergraduate sample designated their relationship status as single (80%), although 10% were cohabiting, 7% were married, and 4% were divorced or separated.

Within the outpatient subsample, 37.9% were recruited from two community mental health facilities (n = 64), 21.3% were recruited from a university psychology clinic (n = 36), and 40.8% from an outpatient behavioral health facility associated with a hospital (n = 69). Participants ranged in age from 18 to 76, with a mean of 42.08 (SD = 14.83) years. There were 110 female and 50 male participants. Participants identified themselves as Caucasian (77.7%; n = 129), African American (12.0%; n = 20), Middle Eastern (0.6%; n = 1), Hispanic (3.6%; n = 6), and other (6.0%; n = 10). Three individuals did not report their ethnic heritage. The majority of the sample designated their relationship status as single (40.2%; n = 68), although 24.3% (n = 41) were married, 24.3% (n = 41) were divorced or separated, 8.3% (n = 14) were cohabiting with a partner, and 3.0% (n = 5) were widowed.

Measurement Invariance Analyses

      SCID-II. The SCID-II BPD scale was subjected to an exploratory factor analysis, to determine factor structure. Mplus 5.2 (Muthén & Muthén, 2008) was utilized for all factor analyses. Initial results appeared to support a one-factor model, with satisfactory fit indices (CFI = 0.97, RMSEA = 0.05), and eigenvalues falling well below 1 after the first factor's variance is taken into account (known as the Kaiser-Guttman rule; Henson & Roberts, 2006; Kaiser, 1960). The SCID-II BPD scale was then fitted to a unidimensional scale for males and females separately; results indicated slightly better model fit in females than in males (CFI = .97, RMSEA = 0.054, and CFI = .93, RMSEA = 0.074, respectively) though both models still indicated adequate fit.

      Multigroup confirmatory factor analysis was then conducted to compare model structure and fit between males and females. Due to the limitations of weighted least squares (WLS) when using smaller sample sizes, and due to the relative robustness of maximum likelihood (ML) estimation against violations of normality assumptions (Schermelleh-Engel, Moosbrugger, & Müller, 2003), ML estimation was used for all CFA models. When a model comparing overall factor structure between groups was fitted to the data, often referred to as configural invariance, model fit deteriorated slightly from initial model fit but still indicated satisfactory model fit. Model fit was determined by examining multiple fit statistics, as recommended by Schermelleh-Engel et al. (2003) and Chen (2007). Subsequent models were also compared to the initial configural invariance model using change in chi-square. Models comparing factor loading equivalency and intercepts, error variance equality, and homogeneity of variance all continued to display acceptable model fit (see Table 1).

PDI-IV. The PDI-IV BPD scale was also subjected to an exploratory factor analysis to determine factor structure. Results also appeared to support a single-factor model, with satisfactory fit indices (CFI = 0.95, RMSEA = 0.068), and eigenvalues falling below 1 after the first factor's variance is taken into account. Results indicated acceptable fit for unidimensionality of the scale, CFI = 0.95, RMSEA = 0.07. The PDI-IV BPD scale was then fitted to a unidimensional scale separately among males and females; results indicated slightly better model fit in males than in females (CFI = 0.99, RMSEA = 0.022, and CFI = 0.95, RMSEA = 0.076, respectively) though again both models continued to indicate adequate model fit in both groups.

Models comparing model structure and fit between males and females were then fitted to the data. When a model comparing overall factor structure between groups was fitted to the data, model fit deteriorated slightly from initial model fit but still indicated satisfactory model fit. A model comparing factor loading and intercept equivalency continued to indicate similarly satisfactory model fit. Model fit deteriorated significantly when a model comparing error variance equality was fitted to the data, and model fit deteriorated again when a model comparing homogeneity of variance was fitted to the data (see Table 2). Thus, the PDI-IV Borderline Scale does not appear to meet criteria for strict measurement invariance between genders.

In order to examine the items to determine whether any single item was contributing significantly to model fit deterioration in the model examining associated error terms, each item was freed while the others were constrained in keeping with the model. The changes in chi-square and model fit indices (CFI, TLI, RMSEA) were examined to see if allowing any individual item to vary between groups would result in a greater increase in model fit or decrease

in chi-square. Results indicated that the item related to assessing impulsivity was contributing the most to the decrease in model fit between groups, and when this item was allowed to vary between groups, the chi-square value was no longer significantly different from the configural model (see Table 3).

ALS. The ALS was subjected to an exploratory factor analysis initially, to determine model structure. Results indicated that the ALS might be better conceptualized as multifactorial than unifactorial, as with the SCID-II and PDI-IV. According to the Kaiser-Guttman rule, where one examines eigenvalues to find all eigenvalues greater than 1 (Henson & Roberts, 2006; Kaiser, 1960) the ALS appears to have 10 factors. According to the scree plot test (Henson & Roberts, 2006), where one examines the slope of eigenvalues as the number of factors increases to find the "elbow," or where the curve flattens out, the ALS appears to have either 4 or 6 factors. According to parallel analysis, often considered to be the most accurate method of determining the number of factors within an EFA (Henson & Roberts, 2006), the ALS appears to have 4 factors. However, model fit at 4 factors was still unsatisfactory (CFI = 0.87, TLI = 0.85, RMSEA = 0.054). Given the difficulty in specifying a well-fitting model for the ALS, and given that it is considered optimal to have 15-20 subjects per variable in an analysis, and that many factor analyses fail to utilize adequate sample sizes (Henson & Roberts, 2006) it was decided that MI analyses with the ALS on the current sample would be inadvisable and might provide misleading results.

BIS-11. The BIS-11 was subjected to an exploratory factor analysis to determine factor structure. Results indicated that the BIS-11 might be better conceptualized as multifactorial, rather than unifactorial as with the SCID-II and PDI-IV. According to the Kaiser-Guttman rule, the BIS-11 appears to have 7 factors. According to the scree plot test, the BIS-11 appears to

have 5 factors. According to parallel analysis, the BIS-11 appears to have 5 factors. Like the

ALS, the BIS-11 did not have satisfactory model fit at 5 factors (CFI = 0.93, TLI = 0.88,

RMSEA = 0.051). Again, given that it is considered optimal to have 15-20 subjects per variable

in an analysis, and that many factor analyses fail to utilize adequate sample sizes (Henson &

Roberts, 2006) it was decided that MI analyses with the BIS-11 on the current sample would be

inadvisable and might provide misleading results due to small sample sizes, particularly among

males.

Gender Differences

       SCID-II Diagnosis and Total Criteria Endorsed. A chi-square analysis between men

and women with regards to diagnosis on the SCID-II indicated no significant differences, $\chi^2(1, N$

$= 331) = 0.94$, $p = .331$ (see Table 3). A t-test was conducted to examine gender differences with

regard to number of total criteria endorsed on the SCID-II, which also showed no significant

differences between men and women, $t(331) = -0.13$, $p = .90$ (see Table 4).

       SCID-II and PDI-IV Individual Criteria. The individual criteria endorsement for both

the SCID-II and PDI-IV were compared between genders using t-tests. It was decided to

compare individual criteria on the PDI-IV, despite lack of strict measurement invariance, in

order to discuss any differences found in the context of the sources of invariance, as discussed

above. Differences on level of endorsement between genders were found on the items for the

criteria related to unstable interpersonal relationships for the SCID-II and differences were found

on level of endorsement for the item related to the criteria of impulsivity, though differences

were only marginally significant in both cases ($p = .05$ and $p = .06$, respectively). Women

endorsed the unstable interpersonal relationships item of the SCID-II at a significantly higher

rate than men, while the reverse was true of the impulsivity item of the PDI-IV. No other items showed significant gender differences on either measure (see Table 5).

Gender Differences on Similar Measures

In order to examine whether or not gender differences found regarding individual criteria on the SCID-II BPD scale were specific to BPD or generalized to other measures of similar constructs within the same sample. *T*-tests were also conducted between genders on the following measures.

BDI-II. Overall BDI-II score and rating on the BDI-II item related to suicidality were compared between men and women using *t*-tests. No significant gender differences were found, $t(316) = 0.64$, $p = .53$. When comparing men and women specifically on the rating of the item related to suicidal ideation, there were also no significant gender differences found, $t(327) = -0.47$, $p = .64$.

BORRTI-O. Scores on each of the four BORRTI-O subscales were also compared between men and women using *t*-tests. Again, no significant gender differences were found for any of the subscales (Alienation, $t[317] = -0.65$, $p = .51$, Insecure Attachment, $t[321] = 1.10$, $p = .27$, Egocentricity, $t[324] = 0.26$, $p = .79$, and Social Incompetence, $t[160.09] = -1.28$, $p = .20$).

<div align="center">

Discussion

</div>

Measurement Invariance

This study set out to examine measurement invariance on two measures of Borderline Personality Disorder in a mixed clinical and nonclinical sample of participants, then to examine gender differences on those measures that establish measurement invariance. Only one of the two clinical interviews of BPD, the SCID-II Borderline Scale, was able to demonstrate strict measurement invariance between men and women. The other clinical interview, the PDI-IV,

failed to meet criteria for full measurement invariance in models where error variance was

constrained to be equal between genders. Specifically, it appears that the item assessing the

criterion relating to impulsivity is contributing the most variance to the model.

Given that the SCID-II was able to demonstrate MI with an item assessing the same

impulsivity criterion, answered by the same sample of individuals and administered and scored

by the same interviewers, it may be that there is something about the way in which the item is

worded or the scoring criteria that is contributing to this difference. The SCID-II item assessing

BPD impulsivity is phrased, "Have you often done things impulsively?" This initial question, if

answered affirmatively by the interviewee, is followed up with "What kinds of things?" At this

point, the interviewer is provided a number of prompts to provide to the interviewee to help

clarify scoring. These include, "How about … buying things you really couldn't afford? Having

sex with people you hardly knew, or 'unsafe sex'? Drinking too much or taking drugs? Driving

recklessly? Uncontrollable eating?" A further prompt states, "If yes to any of the above: Tell

me about that. How often does it happen? What kinds of problems has it caused?" The

interviewee's responses are then assessed on a scale from 1 to 3, and the scoring criteria state:

"Impulsivity in at least two areas that are potentially self-damaging (e.g. spending, sex,

substance abuse, reckless driving, binge eating). [Do not include suicidal or self-mutilating

behavior]." A further note states that a 3 is only to be scored if "several examples indicating a

pattern of impulsive behavior (not necessarily limited to examples above)."

However, the PDI-IV opens with an item stating, "Ever spend so much money that you

had trouble paying it off?" This is followed up with four other questions asked in series: "Ever

go on a drinking or eating binge? Have you ever taken any major chances or risks with drugs?

Ever do anything impulsive that was risky or dangerous? Have you ever become sexually

involved with someone in a risky or dangerous way?" These items are then assessed on a scale

from 0 to 2, with 0 indicating an absence of this criterion, 1 being scored if the interviewer

judges there is "Impulsivity in at least two areas that are potentially self-damaging" and (like the

SCID-II) restricting this to exclude self-harm or suicidal behavior. A score of 2 is given if the

interviewer judges that the interviewee has "impulsivity in at least 3 areas, at least one of which

has been physically self-damaging." Here, a higher rating requires greater impulsivity than on

the SCID-II, where a higher rating simply means meeting the full criteria listed in the DSM-IV-

TR. Interestingly, while the SCID-II item demonstrated MI and the PDI-IV item did not, the

SCID-II item also did not show significant gender differences while the PDI-IV item showed a

significant gender difference such that men were rated significantly higher on this item than

women. This is consistent with the MI analysis conducted on the SIDP-IV by Aggen et al.

(2009), where they found that the SIDP-IV was not measurement invariant, and similarly found

that the impulsivity criterion was problematic, with men reporting higher overall levels of

impulsivity.

One potential explanation for this finding is that the type of impulsivity assessed as part

of the BPD construct is not the same as the broader construct of impulsivity, on which men have

historically demonstrated significant across-the-board gender differences, particularly at younger

ages (i.e., Trent & Davies, 2012). The SCID-II item asks specifically about impulsivity that has

the potential to be self-damaging across specific behavioral areas, such as risky sex, substance

abuse, binge eating, and reckless driving. While the SCID-II requires that this impulsivity not be

assessed on the basis of suicidal or self-harm behaviors (i.e. cutting, burning, or scratching

oneself), it seems as though the behaviors being assessed have the potential to overlap with self-

harming behaviors in terms of their function, if not their form. The DSM-IV-TR (American

Psychiatric Association, 2000) defines Criterion 4 of BPD thusly: "Impulsivity in at least two areas that are potentially self-damaging (e.g., spending, sex, substance abuse, reckless driving, binge eating)." This criterion also explicitly excludes suicidal and self-harm behaviors as assessed by Criterion 5.

Several authors have noted confusion regarding the construct of impulsivity (Trent & Davies, 2012; Whiteside & Lynam, 2001). For example, Trent and Davies (2012) reported that impulsivity may break down into at least two distinct traits: impulsive action, or behavioral disinhibition, and impulsive choice, or decision-making without appropriate consideration beforehand. These traits are commonly measured via neuropsychological task performance, such as Go/No-go and delay discounting paradigms, respectively. The results of this study, in which the BIS-11 failed to fit a unifactorial factor structure and appears to have multiple factors, supports this conceptualization of impulsivity as multifactorial.

Further supporting the imprecision of previous models of impulsivity are Whiteside and Lynam (2001), who reported utilizing a number of measures of impulsivity (including items from the NEO-PI-R and BIS-11, among others) to obtain four constructs: Lack of Premeditation, Urgency, Sensation Seeking, and Lack of Perseverance. The authors term this the "UPPS" model of impulsivity. Interestingly, the construct Whiteside and Lynam (2001) refer to as "urgency" appears the most conceptually relevant to the assessment of BPD. The authors report that this subscale within the UPPS model "appeared to reflect a tendency to commit rash or regrettable actions as a result of intense negative affect" (p. 677). Whiteside, Lynam, Miller, and Reynolds (2005) further tested this model in relation to BPD, and found Urgency to be the most significant predictor of BPD features, though Sensation Seeking and Lack of Perseverence were also significant predictors. Similarly, DeShong and Kurtz (2013) found that these four

constructs within the domain of impulsivity differentiated between BPD and Antisocial Personality Disorder, such that BPD was uniquely predicted by Urgency and Lack of Perseverence. Unfortunately, neither of these studies examined gender differences among the different subscales of UPPS impulsivity.

Miller, Flory, Lynam, and Leukefeld (2003), however, examined gender differences in the relationship between these four aspects of impulsivity and BPD symptoms. They approximated the four Impulsivity constructs using the NEO-PI-R (Costa & McCrae, 1992) items used by Whiteside and Lynam (2001) in developing the UPPS model, and approximated the symptoms of BPD based on items from the Brief Symptom Inventory (Derogatis, 1993), and found that Urgency was significantly correlated with BPD symptoms only for women, while both Lack of Perseverence and Lack of Premeditation were significantly correlated regardless of gender.

Thus, the error variance detected by this set of analyses in this population might arise from a confusion regarding the assessment of the construct of impulsivity, particularly between assessing a more global conceptualization of impulsivity including the aforementioned separable traits, and the impulsive behaviors specifically associated with BPD, as defined by the DSM-IV-TR. The PDI-IV may assess more generalized sex differences in trait impulsivity, such as impulsive action, which are not present in BPD specifically or are not necessarily a function of BPD.

There is also evidence to support the assertion that at least some types of impulsivity naturally vary between men and women. Trent and Davies (2012) report some preliminary evidence suggestive of a biological basis for these sex differences in trait impulsivity. For example, they report that there is some evidence that biological females may outperform males

on tasks of response inhibition, such as the Stroop task, indicating males may have greater levels of trait impulsive action. However, the imprecision with which impulsivity has been measured and a lack of consistency in experimental findings raises caution regarding generalizing from these findings. Still, this baseline difference may mean that, in a population with heightened impulsivity (e.g. individuals with BPD), relative gender differences remain intact, while overall levels of impulsivity increase proportionately to a healthy population. Thus, an alternative explanation for the findings of this study is that the scaling of the second threshold on the PDI-IV, where the first threshold indicates meeting criteria and the second indicates a higher level of the trait, allows for the detection of variance associated with an overall sex difference in trait impulsivity that is not a consequence of BPD specifically.

Prevalence and Frequency Differences

When men and women were compared in terms of frequency of diagnosis on the SCID-II, as well as in terms of number of criteria endorsed on the SCID-II, no significant gender differences emerged. An absence of significant gender differences in diagnosis and overall number of BPD symptoms endorsed is consistent with several epidemiological studies, such as Grant et al. (2008), and Lenzenweger et al. (2007), which also found no significant differences in diagnosis and/or number of criteria endorsed on measures of BPD.

However, one difference emerged when examining endorsement of individual criteria on the SCID-II between genders. Specifically, it appears that women's responses to the item relating to chaotic and unstable interpersonal relationships were given a significantly higher rating, though only marginally so ($p = .05$), by the interviewer. Notably, mean differences did not always display the same trend. Men tended to be rated slightly higher for the criteria related to impulsivity, difficulty controlling anger, and affective instability.

Given these results, the SCID-II Borderline Scale appears to be measurement invariant with regard to gender. Using the SCID-II Borderline Scale on a mixed clinical and nonclinical sample does not yield significant gender differences in terms of rates of diagnosis, nor number of criteria endorsed. However, some trends and significant gender differences emerge when examining men and women's endorsement of individual criteria. It appears that, while the same construct is being measured in both groups of individuals, the exact manifestation of that construct may differ slightly according to the individual, and that the individual's gender may play some role in how these differences emerge.

It stands to reason that the individual's gender may play a role in how the underlying vulnerabilities, hypothesized to result in the disorder we currently call BPD, manifest within the context surrounding that individual. For example, it is well-known that women are stereotyped and socialized to be more relationship-oriented than men (e.g. Perrin, Heesacker, Tiegs, Swan, Lawrence, Smith et al., 2010). Since this is the case, women may be more likely to notice and complain of instability in interpersonal functioning. Similarly, affective instability could be said to be considered relatively more normative in women than men, as women are stereotyped as "hyper-emotional" (pp. 614, Perrin et al., 2010); therefore, men might be more likely to complain when they notice this trait in themselves. Men might also be more likely to express this difficulty in a more socially acceptable manner. For example, frequent media portrayals of men reacting in violent or angry ways might lead men to be more willing to express heightened emotional reactivity in terms of anger, or to simply be more willing to endorse difficulty with anger as opposed to difficulty with other emotions. This might explain a trend toward women endorsing relational difficulties at a slightly higher rate, while men were slightly more likely to endorse difficulties with anger in the current sample, despite finding measurement invariance

with the SCID-II BPD scale. In line with this hypothesis, previous studies have shown different patterns of comorbid disorders between genders with individuals who meet criteria for BPD (for example, men with BPD have been found to endorse criteria of intermittent explosive disorder at a higher rate than women with BPD; Zlotnick, Rothschild, & Zimmerman, 2002). In other words, men and women might struggle with BPD at relatively equal rates, but the manifestation and consequences of the disorder might differ according to the differences in men and women's psychosocial history and environment. Future research may want to explore this hypothesis in greater depth; for example, looking at socialization and femininity/masculinity as potential moderators of this relationship.

Summary

The results of this study found that the SCID-II BPD scale demonstrated strict measurement invariance with regards to gender differences in BPD symptoms, while the PDI-IV does not meet strict measurement invariance criteria. Specifically, the construction of the question regarding impulsivity appears to be contributing significant variance between genders on the PDI-IV. Potential explanations include poor framing of the type of impulsive behaviors associated with BPD, as well as a higher ceiling that captures variance attributable to natural gender differences in overall trait impulsivity, rather than differences associated with BPD specifically. This study also found no significant gender differences in overall frequency of BPD diagnosis using the SCID-II, nor in number of BPD criteria endorsed. However, some differences in the rates of endorsement of individual criteria between men and women were found. While BPD appears to be a relatively unitary construct with nonsignificant gender differences in the present sample, the specific constellation of symptoms endorsed appears to vary slightly between men and women.

Several limitations of the present study are noteworthy. Specifically, substance abusing participants were excluded from the sample. Given the findings that impulsivity was problematic in MI analyses for the PDI-IV, and that impulsivity is a common correlate of substance abuse, and men tend to abuse substances at a higher rate than women in both general populations as well as specifically within BPD populations (McCormick et al., 2007), the inclusion of substance abusing participants could have altered the present findings significantly. The sample size utilized for the MI analyses was also on the smaller end of the recommended sample size for this type and complexity of analyses, and a larger sample would be preferable for future studies. Finally, multi-method comparisons within a given sample should also be undertaken, to determine whether the method of the measure (i.e. structured clinical interview versus self-report measure) contributes to measurement variance in any way.

References

Aggen, S. H., Neale, M. C., Røysamb, E., Reichborn-Kjennerud, T., & Kendler, K. S. (2009). A psychometric evaluation of the DSM-IV borderline personality disorder criteria: age and sex moderation of criterion functioning. *Psychological Medicine*, *39*(12), 1967-78.

Beauchaine, T. P., Klein, D. N., Crowell, S. E., Derbidge, C., & Gatzke-Kopp, L. (2009). Multifinality in the development of personality disorders: a Biology x Sex x Environment interaction model of antisocial and borderline traits. *Development and Psychopathology*, *21*(3), 735-70.

Beck, A. T., Steer, R. A., & Brown, G. K. (1996). *Beck Depression Inventory II Manual.* San Antonio, TX: Psychological Corporation.

Becker, D., & Lamb, S. (1994). Sex bias in the diagnosis of Borderline Personality Disorder and Posttraumatic Stress Disorder. *Professional Psychology: Research and Practice, 25*(1), 55-61.

Becker, D., & Lamb, S. (1995). Addressing criticisms of sex bias research on Borderline Personality Disorder. *Professional Psychology: Research and Practice*, *26*(5), 525-527.

Bell, M. D. (1995). *Bell object relations and reality testing inventory*. San Antonio, TX.

Billieux, J., Gay, P., Rochat, L., & Linden, M. Van Der. (2010). The role of urgency and its underlying psychological mechanisms in problematic behaviours. *Behaviour Research and Therapy*, *48*(11), 1085–1096.

Boggs, C. D., Morey, L. C., Skodol, A. E., Shea, M. T., Sanislow, C. A, Grilo, C. M., McGlashan, T. H., et al. (2009). Differential impairment as an indicator of sex bias in DSM-IV criteria for four personality disorders. *Personality Disorders: Theory, Research, and Treatment*, *S*(1), 61-68.

Chen, F. F. (2007). Sensitivity of Goodness of Fit Indexes to Lack of Measurement Invariance. *Structural Equation Modeling*, *14*(3), 464–504.

Coid, J., Yang, M., Tyrer, P., Roberts, A., & Ullrich, S. (2006). Prevalence and correlates of personality disorder in Great Britain. *The British Journal of Psychiatry*, *188*, 423-431.

Crowell, S. E., Beauchaine, T. P., & Linehan, M. M. (2009). A biosocial developmental model of borderline personality: Elaborating and extending Linehan's theory. *Psychological Bulletin*, *135*(3), 495-510.

De Moor, M. H. M., Distel, M. a, Trull, T. J., & Boomsma, D. I. (2009). Assessment of borderline personality features in population samples: is the Personality Assessment Inventory-Borderline Features scale measurement invariant across sex and age? *Psychological Assessment*, *21*(1), 125-30.

DeShong, H. L., & Kurtz, J. E. (2013). Four factors of impulsivity differentiate Antisocial and Borderline Personality Disorders. *Journal of Personality Disorders*, *27*(2), 144–56.

Fernbach, B. E., Winstead, B. A., & Derlega, V. J. (1989). Sex differences in diagnosis and treatment recommendations for Antisocial Personality and Somatization Disorders. *Journal of Social and Clinical Psychology*, *8*(3), 238-255.

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*(5), 431–446.

Flanagan, E. H., & Blashfield, R. K. (2005). Gender acts as a context for interpreting diagnostic criteria. *Journal of Clinical Psychology*, *61*(12), 1485-1498.

Grant, B. F., Chou, S. P., Goldstein, R. B., Huang, B., Stinson, F. S., Saha, T. D., Smith, S. M., et al. (2008). Prevalence, correlates, disability, and comorbidity of DSM-IV Borderline

Personality Disorder: Results from the Wave 2 National Epidemiologic Survey on Alcohol and Related Conditions. *Journal of Clinical Psychiatry*, *69*, 533-545.

Grant, B. F., Dawson, D. A., & Hasin, D. S. (2001). *The Alcohol Use Disorders and Associated Disabilities Interview Schedule–DSM-IV Version*. Bethesda, Md: National Institute on Alcohol Abuse and Alcoholism. Available at: http://www.niaaa.nih.gov.

Gratz, K. L., Tull, M. T., Reynolds, E. K., Bagge, C. L., Latzman, R. D., Daughters, S. B., & Lejuez, C. W. (2009). Extending extant models of the pathogenesis of Borderline Personality Disorder to childhood borderline personality symptoms: The roles of affective dysfunction, disinhibition, and self- and emotion-regulation deficits. *Development and Psychopathology*, *21*(4), 1263-91.

Hartung, C. M., & Widiger, T. A. (1998). Gender differences in the diagnosis of mental disorders : Conclusions and controversies of the DSM-IV. *Psychological Bulletin, 123*(3), 260-278.

Harvey, P. D., Greenberg, B. R., & Serper, M. R. (1989). The Affective Lability Scales: Development, reliability, and validity. *Journal of Clinical Psychology, 45*(5), 786-793.

Henson, R. K., & Roberts, J. K. (2006). Use of exploratory factor analysis in published research: Common errors and some comment on improved practice. *Educational and Psychological Measurement*, *66*(3), 393–416.

Huprich, S. K., Schmitt, T. A., Richard, D. C. S., Chelminski, I., & Zimmerman, M. A. (2010). Comparing factor analytic models of the DSM-IV personality disorders. *Personality Disorders: Theory, Research, and Treatment*, *1*(1), 22–37.

Huprich, S. K., Zimmerman, M., & Chelminski, I. (2006). Disentangling depressive personality disorder from avoidant, borderline, and obsessive-compulsive personality disorders. *Comprehensive Psychiatry*, *47*(4), 298-306.

Jackson, H. J., & Burgess, P. M. (2000). Personality disorders in the community: A report from the Australian National Survey of Mental Health and Wellbeing. *Social Psychiatry and Psychiatric Epidemiology*, *35*, 531-538.

Jane, J. S., Oltmanns, T. F., South, S. C., & Turkheimer, E. (2007). Gender bias in diagnostic criteria for personality disorders: an item response theory analysis. *Journal of Abnormal Psychology*, *116*(1), 166-75.

Johnson, D. M., Shea, M. T., Yen, S., Battle, C. L., Zlotnick, C., Sanislow, C. A., Grilo, C. M., et al. (2003). Gender Differences in Borderline Personality Disorder: Findings from the Collaborative Longitudinal Personality Disorders Study. *Comprehensive Psychiatry*, *44*(4), 284-292.

Koenigsberg, H. W., Harvey, P. D., Mitropoulou, V., New, A. S., Goodman, M., Silverman, J., et al. (2001). Are the interpersonal and identity disturbances in the borderline personality disorder criteria linked to the traits of affective instability and impulsivity? *Journal of Personality Disorders, 15,* 358-370.

Lenzenweger, M. F., Lane, M. C., Loranger, A. W., & Kessler, R. C. (2007). DSM-IV personality disorders in the National Comorbidity Survey Replication. *Biological Psychiatry*, *62*(6), 553-64.

Linehan, Marsha M. (1993). *Cognitive behavioral treatment of Borderline Personality Disorder*. New York: Guilford Press.

Links, P. S., Heslegrave, R., & van Reekum, R. (1999). Impulsivity: Core aspect of borderline personality disorder. *Journal of Personality Disorders, 13,* 1-9.

Loranger, A. W. (1999). *International Personality Disorder Examination: DSM-IV and ICD-10 Interviews*. Odessa, Florida: Psychological Assessment Resources.

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*(6), 583-602.

McCormick, B., Blum, N., Hansel, R., Franklin, J. A., St John, D., Pfohl, B., Allen, J., et al. (2007). Relationship of sex to symptom severity, psychiatric comorbidity, and health care utilization in 163 subjects with borderline personality disorder. *Comprehensive Psychiatry*, *48*(5), 406-12.

McGrath, R. E., Mitchell, M., Kim, B. H., & Hough, L. (2010). Evidence for response bias as a source of error variance in applied assessment. *Psychological bulletin*, *136*(3), 450-70.

Miller, J., Flory, K., Lynam, D., & Leukefeld, C. (2003). A test of the four-factor model of impulsivity-related traits. *Personality and Individual Differences*, *34*(8), 1403–1418.

Millsap, R. E. (2007). Invariance in measurement and prediction revisited. *Psychometrika*, *72*(4), 461–473.

Morey, L. C. (1991). *The Personality Assessment Inventory: Professional manual*. Lutz, FL: Psychological Assessment Resources.

Patton, J. H., Stanford, M. S., Barratt, E. S. (1995). Factor structure of the Barratt Impulsiveness Scale. *Journal of Clinical Psychology, 51,* 768 – 774.

Pfohl, B., Blum, N., & Zimmerman, M. (1995). *Structured Interview for DSM-IV Personality (SIDP-IV)*. Department of Psychiatry, University of Iowa: Iowa City.

Pfohl, B., Blum, N., Zimmerman, M., & Stangl, D. (1989). *Structured Interview for DSM-III-R Personality Disorders (SIDP-R).* Iowa City, Iowa: Department of Psychiatry, University of Iowa.

Samuel, D. B., Ansell, E. B., Hopwood, C. J., Morey, L. C., Markowitz, J. C., Skodol, A. E., & Grilo, C. M. (2010). The impact of NEO PI-R gender norms on the assessment of personality disorder profiles. *Psychological Assessment*, *22*(3), 539-45.

Sansone, R. A., & Sansone, L. A. (2011). Gender patterns in Borderline Personality Disorder. *Clinical Neuroscience*, *8*(5), 16-20.

Schermelleh-Engel, K., Moosbrugger, H., & Müller, H. (2003). Evaluating the Fit of Structural Equation Models : Tests of Significance and Descriptive Goodness-of-Fit Measures. *Methods of Psychological Research Online*, *8*(2), 23–74.

Skodol, A. E., & Bender, D. S. (2003). Why are women diagnosed borderline more than men? *Psychiatric Quarterly*, *74*(4), 349-60.

Skodol, A. E., Gunderson, J. G., Pfohl, B., Widiger, T. a, Livesley, W. J., & Siever, L. J. (2002). The borderline diagnosis I: Psychopathology, comorbidity, and personality structure. *Biological Psychiatry*, *51*(12), 936-50.

Skodol, A. E., Siever, L. J., Livesley, W. J., Gunderson, J. G., Pfohl, B., & Widiger, T. A. (2002). The borderline diagnosis II: Biology, genetics, and clinical course. *Biological Psychiatry*, *51*(12), 951-963.

Slof-Op't Landt, M. C. T., Dolan, C. V., Rebollo-Mesa, I., Bartels, M., van Furth, E. F., van Beijsterveldt, C. E. M., Meulenbelt, I., et al. (2009). Sex differences in sum scores may be hard to interpret: The importance of measurement invariance. *Assessment*, *16*(4), 415-23.

Sprague, J., Javdani, S., Sadeh, N., Newman, J. P., & Verona, E. (2011). Borderline Personality Disorder as female phenotypic expression of psychopathy? *Personality Disorders: Theory, Research, and Treatment*, 1-13.

Sprock, J., Blashfield, R. K., & Smith, B. (1990). Gender weighting of DSM-III-R personality disorder criteria. *The American Journal of Psychiatry, 147*(5), 586-90.

Stanford, M. S., Mathias, C. W., Dougherty, D. M., Lake, S. L., Anderson, N. E., & Patton, J. H. (2009). Fifty years of the Barratt Impulsiveness Scale: An update and review. *Personality and Individual Differences, 47,* 385 – 395.

Stepp, S. D., Pilkonis, P. a, Hipwell, A. E., Loeber, R., & Stouthamer-Loeber, M. (2010). Stability of borderline personality disorder features in girls. *Journal of Personality Disorders*, *24*(4), 460-72.

Torgersen, S., Kringlen, E., & Cramer, V. (2001). The prevalence of personality disorders in a community sample. *Archives of General Psychiatry*, *58*(6), 590-596.

Tragesser, S. L., & Robinson, R. J. (2009). The role of affective instability and UPPS impulsivity in Borderline Personality Disorder features. *Journal of Personality Disorders, 23*(4), 370-383.

Tragesser, S. L., Solhan, M., Schwartz-Mette, R., & Trull, T. J. (2007). The role of affective instability and impulsivity in predicting future BPD features. *Journal of Personality Disorders, 21*(6), 603-614.

Trent, S., & Davies, W. (2012). The influence of sex-linked genetic mechanisms on attention and impulsivity. *Biological Psychology*, *89*(1), 1–13.

Trull, T. J., Jahng, S., Tomko, R. L., Wood, P. K., & Sher, K. J. (2010). Revised NESARC personality disorder diagnoses: gender, prevalence, and comorbidity with substance dependence disorders. *Journal of Personality Disorders, 24*(4), 412-426.

Trull, T. J., Tomko, R. L., Brown, W. C., & Scheiderer, E. M. (2010). Borderline Personality Disorder in 3d: Dimensions, symptoms, and measurement challenges. *Social and Personality Psychology Compass*, *4*(11), 1057-1069.

Vandenberg, R. J., & Lance, C. E. (2000). A Review and Synthesis of the Measurement Invariance Literature: Suggestions, Practices, and Recommendations for Organizational Research. *Organizational Research Methods*, *3*(1), 4–70. doi:10.1177/109442810031002

Whiteside, S. P., & Lynam, D. R. (2001). The Five Factor Model and impulsivity: Using a structural model of personality to understand impulsivity. *Personality and Individual Differences*, *30*(4), 669–689.

Whiteside, S. P., Lynam, D. R., Miller, J. D., & Reynolds, S. K. (2005). Validation of the UPPS impulsive behaviour scale: A four-factor model of impulsivity. *European Journal of Personality*, *19*(7), 559–574.

Widiger, T. A. (1998). Invited essay: Sex biases in the diagnosis of personality disorders. *Journal of Personality Disorders*, *12*(2), 95-118.

Widiger, T. A., Mangine, S., Corbitt, E. M., Ellis, C. G., & Thomas, G. V. (1995). *Personality Disorder Interview—IV: A semi-structured interview for the assessment of personality disorders*. Odessa, FL: Psychological Assessment Resources.

Widiger, T. A., & Trull, T. J. (1993). Borderline and Narcissistic Personality Disorders. *Comprehensive Handbook of Psychopathology* (pp. 371–394). New York, NY: Plenum Press.

Woodward, H. E., Taft, C. T., Gordon, R. A, & Meis, L. A. (2009). Clinician bias in the

diagnosis of posttraumatic stress disorder and borderline personality disorder.

Psychological Trauma: Theory, Research, Practice, and Policy, *1*(4), 282-290.

doi:10.1037/a0017944

Yen, S., Shea, M. T., Sanislow, C. A., Grilo, C. M., Skodol, A. E., Gunderson, J. G., et al.

(2004). Borderline Personality Disorder criteria associated with prospectively observed

suicidal behavior. *American Journal of Psychiatry, 161*, 1296-1298.

Zanarini, M.C., Frankenburg, F.R., Sickel, A.E., & Young, L.. (1996). *The Diagnostic Interview

for DSM-IV Personality Disorders (DIPD-IV)*. Belmont, MA: McLean Hospital.

Zlotnick, C., Rothschild, L., & Zimmerman, M. (2002). The role of gender in the clinical

presentation of patients with Borderline Personality Disorder. *Journal of Personality

Disorders*, *16*(3), 277-282.

Appendix A

Table 2

SCID-II Measurement Invariance Model Fit

Model	χ2	df	Δ Chi-Square	CFI	RMSEA	SRMR
Baseline Model – Unifactorial Factor Structure	46.77	27		0.98	0.047	0.03
Baseline Females Only	45.17	27		0.97	0.054	0.035
Baseline Males Only	33.32	27		0.97	0.049	0.051
Model 1 – Factor Structure Equivalency	97.35	62		0.96	0.059	0.049
Model 2 – Factor Loading Equivalency	99.62	70	2.27	0.97	0.051	0.048
Model 3 – Error Variance Equality	123.56	79	26.21	0.95	0.058	0.067
Model 4 – Error Variance Homogeneity	124.19	80	26.84	0.95	0.058	0.07

Table 3

PDI-IV Measurement Invariance Model Fit

Model	$\chi 2$	*df*	Δ Chi-Square	CFI	RMSEA	SRMR
Baseline Model – Unifactorial Factor Structure	69.91	27		0.95	0.93	0.038
Baseline Females Only	74.65	27		0.93	0.91	0.046
Baseline Males Only	38.26	27		0.93	0.91	0.059
Model 1 – Factor Structure Equivalency	127.79	62		0.93	0.91	0.057
Model 2 – Factor Loading Equivalency	142.12	70	14.33	0.92	0.92	0.068
Model 3 – Error Variance Equality	165.32	79	37.53	0.9	0.91	0.081
Model 4 – Error Variance Homogeneity	171.23	80	43.44	0.9	0.91	0.105

Table 4

PDI-IV Model Fit Comparison, Model 1 and Model 3

PDI-IV Model	$\chi2$	*df*	Δ Chi-Square	CFI	RMSEA	SRMR
PDI-IV, Model 1	127.79	62		0.93	0.08	0.057
Identity Problems	162.92	78	35.13	0.9	0.081	0.079
Abandonment	162.77	78	34.98	0.9	0.081	0.075
Unstable Relationships	165.3	78	37.51	0.9	0.082	0.081
Anger	164.89	78	37.1	0.9	0.082	0.08
Impulsivity	153.42	78	25.63	0.91	0.077	0.078
Emptiness	164.59	78	36.8	0.9	0.082	0.08
Affective Instability	161.19	78	33.4	0.91	0.081	0.08
Suicidal Behavior	165.21	78	37.42	0.9	0.082	0.081
Dissociation	164.18	78	36.39	0.9	0.082	0.08

Each item listed is freed within Model 3 while the other items continue to be constrained across groups. Change in chi-square is assessed by comparing to Model 1, testing configural invariance.

Table 5

SCID-II and PDI-IV Diagnosis by Gender

	Women	Men	χ^2	*p*
SCID-II	25 (7.6%)	7 (2.1%)	0.94	.33
PDI-IV	57 (17.4%)	19 (5.8%)	0.87	.35

Three females and one male did not receive a diagnosis

Table 6

SCID-II and PDI-IV T-Test by Gender and Total Criteria Endorsed

	Mean (SD) Women	Mean (SD) Men	*t*	*p*
SCID-II Total Criteria Endorsed	1.38 (2.05)	1.41 (1.89)	-0.13	.90
PDI-IV Total Criteria Endorsed	2.62 (2.65)	2.54 (2.35)	0.24	.81

Table 7

SCID-II and PDI-IV T-test by Gender and Individual Criteria Rating

Criterion	Measure	Mean (SD) Women	Mean (SD) Men	t	p
Fear of Abandonment	SCID-II	1.38 (0.69)	1.28 (0.55)	1.47*	.142
	PDI-IV	0.21 (0.51)	0.13 (0.40)	1.52*	.131
Unstable Relationships	SCID-II	1.58 (0.82)	1.40 (0.73)	1.96*	.052
	PDI-IV	0.38 (0.61)	0.31 (0.55)	1.10*	.272
Identity Problems	SCID-II	1.27 (0.55)	1.27 (0.59)	0.18	.986
	PDI-IV	0.30 (0.63)	0.25 (0.54)	0.71	.477
Impulsivity	SCID-II	1.53 (0.75)	1.68 (0.82)	-1.51*	.133
	PDI-IV	0.44 (0.71)	0.62 (0.80)	-1.92*	.057
Suicidal Behavior	SCID-II	1.47 (0.78)	1.38 (0.73)	0.96	.339
	PDI-IV	0.41 (0.69)	0.33 (0.64)	0.93	.351
Affective Instability	SCID-II	1.61 (0.85)	1.62 (0.85)	-0.11	.909
	PDI-IV	0.49 (0.66)	0.46 (0.6)	0.30	.765
Feelings of Emptiness	SCID-II	1.52 (0.82)	1.57 (0.85)	-0.50	.618
	PDI-IV	0.43 (0.7)	0.40 (0.66)	0.36	.723
Difficulty Controlling Anger	SCID-II	1.42 (0.69)	1.53 (0.77)	-1.19*	.234
	PDI-IV	0.41 (0.64)	0.35 (0.63)	0.75	.457
Dissociation	SCID-II	1.35 (0.65)	1.43 (0.72)	-1.02	.311
	PDI-IV	0.36 (0.63)	0.38 (0.64)	-0.25	.806

Corrected for inequality of variance