January 2019

Ethnic Bias Within The NESARC Personality Disorders Using Item Response Theory

Susan Tara Scoullar

ETHNIC BIAS WITHIN THE NESARC PERSONALITY DISORDERS USING ITEM RESPONSE THEORY

by

Susan Tara Scoullar

Bachelor of Arts (Hons.), University of British Columbia, 2011

A Thesis

Submitted to the Department of Psychology

Graduate Faculty

of the

University of North Dakota

in partial fulfillment of the requirements

for the degree of

Master of Arts

Grand Forks, North Dakota

December

2019

Correspondence concerning this article should be addressed to Susan Scoullar, Department of Psychology, University of North Dakota, Grand Forks, ND

Contact: TaraScoullar@gmail.com

This thesis, submitted by Susan Tara Scoullar in partial fulfillment of the requirements for the Degree of Master of Arts from the University of North Dakota, has been read by the Faculty Advisory Committee under whom the work has been done and is hereby approved.

Justin Douglas McDonald, Ph.D.

Alison Kelly, Ph.D.

Alan King, Ph.D.

This thesis is being submitted by the appointed advisory committee as having met all of the requirements of the School of Graduate Studies at the University of North Dakota and is hereby approved.

Chris Nelson
Associate Dean of the School of Graduate Studies

9/3/19
Date

PERMISSION

Title Ethnic Bias Within the NESARC Personality Disorders Using Item Response Theory

Department Psychology

Degree Master of Arts

In presenting this thesis in partial fulfillment of the requirements for a graduate degree from the University of North Dakota, I agree that the library of this University shall make it freely available for inspection. I further agree that permission for extensive copying for scholarly purposes may be granted by the professor who supervised my thesis work, or in his absence, by the Chairperson of the department or the Dean of the School of Graduate Studies. It is understood that any copying or publication or other use of this thesis or part thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and to the University of North Dakota in any scholarly use which may be made of any material in my thesis.

Susan Tara Scoullar

08/26/2019

TABLE OF CONTENTS

LIST OF TABLES

Disclaimer

Any opinions and conclusions expressed herein are those of the author and do not necessarily reflect the views of the U.S. Census Bureau. All results have been reviewed to ensure that no confidential information is disclosed. The statistical summaries reported in this document have been cleared by the Census Bureau's Disclosure Review Board release authorization numbers CBDRB-FY19-422, CBDRB-FY19-448 and CBDRB-FY19-536.

Acknowledgements

A special thank you to Jahn K. Hakes, statistician at U.S. Census Bureau, for running the requested analyses and Norman Johnson, U.S. Census Bureau for his assistance. Thank you to the U.S. Census Bureau for maintaining these records. Thank you to the NIAAA and their data collection team for gathering all the information, and Bridget Grant for her assistance with the original project.

A special thank you to Dr. Brian O'Connor, who supervised and assisted me in my original work on this topic in my undergraduate thesis, and gave me the idea to do this project. I hope my analyses make you proud.

I wish to express my appreciation to my supervisor, Doug McDonald, and committee members, Alan King and Alison Kelly, for their support and guidance.

To my family, my daughter, Mia, husband, Stephen, and my parents, Kim and Gail. Thank you for all your love, patience and encouragement. I love you all.

Conflict of Interest Statement

The author does not have any known conflicts of interest at this time.

Abstract

Racial bias within the Diagnostic and Statistical Manual 5th Edition, or DSM5, has been recognized in studies of personality disorder diagnoses. Item Response Theory Differential Functioning of Items and Tests (IRT DIF and DTF) was used to examine the National Institute on Alcohol Abuse and Alcoholism's (NIAAA) National Epidemiologic Survey on Alcohol and Related Conditions (NESARC) Wave 1 and 2 data. A review of prevalence information by gender and ethnicity was undertaken. Hypotheses included that bias would be present in the items and total scores of personality diagnostic scales. In particular, items and scales which rely on cultural comparisons, and out of body experiences, magical thinking, hallucinations or delusions would be diagnosed more frequently in Native American participants, given that the study would not have accounted for ethnic background when the algorithm for classifying probable diagnosis was used. Results indicated test level bias within five of the scales for African Americans, two for females, one for males, and one for Native Americans. The schizotypal scales was biased for Native Americans and African American participants, as hypothesized. Only three scales contained no scale level bias: avoidant, dependent and obsessive-compulsive.

Keywords: NESARC, Racial Bias, Personality Disorders, IRT

Ethnic Bias within the NESARC Personality Disorders using Item Response Theory

Cultural considerations should always be used when formulating cases and proceeding with diagnosis. Personality disorders in particular are prone to misinterpretation of cultural backgrounds, given the persistent and pervasive nature of the disorder with longstanding cultural beliefs. The Diagnostic and Statistical Manual, Fifth Edition (DSM5), defines a personality disorder as a pervasive, inflexible and enduring (stable over time) pattern of experience and behaviour that deviates markedly from sociocultural expectations that has an onset in late adolescence or early adulthood and leads to distress or impairment (APA, 2013). It includes criteria for ten defined personality disorders in three clusters which will be explored for cultural biases in the following study.

Many minority populations have differences in the way in which they view mental health, which can be affected by interactions within their society, acculturation level in more Western ways of thinking and culture, ancestral history impact on present day views and traditions, and spiritual beliefs (among many, many other factors; for a review, see Trianis & Suh, 2002 or Dana, 1988). Hispanic or Latino individuals tend to live in more collectivist or family focused environments. Native American individuals have also been raised in more collectivist communities with a focus on traditional spiritual beliefs and healing methods. Cultural background may influence factors which effect mental health, for example, socio-economic status, education, and experience of racial discrimination. Lewis-Fernandez & Kleinman (1994) described this distinct influence of a culture on a person's mental health and personality when they stated that "personality and psychopathology take form in distinct local worlds, characterized by behavioral environments consisting of consensual orientations to self, objects, space, time, motivations, and moral norms that are culturally constituted, shared to different

degrees, and invoked differently in specific situations by members of the social group" and that "behavioral environments vary greatly both across and within local worlds, leading to multiple versions of self and personality that are constructed in relation to not just other individuals but also whole communities, institutions, and even, in some settings, spirits and gods."

Diversity in psychological research

Cultural and gender differences have long been acknowledged in psychological diagnosis, treatment, and research. With regards to research and test construction, diversity concerns are not always taken into account. As mentioned with personality disorders above, the DSM5 accounts for cultural and gender differences in presentation and prevalence of disorders. These considerations are not always accounted for in test construction. Participants who have been raised in a collectivist culture may appear more abnormally dependent on others than those raised in a more individualist culture, while those who hold strong religious or spiritual beliefs outside of Christian beliefs may appear to have strange ideas to those with them. Cultural background may also influence other important areas which influence mental health, including socioeconomic status, education, experience of discrimination, or other environmental factors.

Anxiety disorders have also been shown to have cultural differences, with Caucasians being more likely to be diagnosed with social anxiety disorders, generalized anxiety disorders, and panic disorders than African Americans, Hispanics, or Asians (Asnaani, Richey, Dimaite, Hinton & Hofmann, 2010). African Americans were more likely to experience PTSD, while Asians were less likely to meet criteria for an anxiety disorder. Examination of prevalence of generalized anxiety disorder (GAD) within the NESARC indicated a ratio of 1:1.9 for men and 1:2.2 for women (Vesga-López, Schneier, Wang, Heimberg, Liu, Hasin & Blanco, 2008). Men with GAD were less likely to be African American.

Sue (1996) detailed a variety of concerns related to ethical assessment of minority groups and ethnic bias, including test motivation, practice of test items, assessor bias and knowledge of ethnic background, and equivalency of test materials and test norms. One example used is of a depression inventory score of 50 as a cutoff for severe depression in the United States as not necessarily being a good cutoff for those in another country, where a higher or lower score may be more appropriate.

The University of North Dakota recently published study on use of BDI-II, BHS, BAI, SCL-90-R, CES-D on Northern Plains Native American populations (Gray, Brionez, Petros & Gongaza, 2018). Overall, measures were acceptable for use with existing norms (CES-D had trouble differentiating anxiety and depression). The Psychopathy Checklist has also been examined for validity with Native Americans (Stockdale, Wong & Olver, 2010; McCuish, Mathesius, Lussier & Corrado, 2018). The authors determined that it was valid for use, but the antisocial factor in a four factor model accounted the most for recidivism.

Thomason (1999) reviewed important information for the use of assessments to ensure the validity of a measure for minority populations. These concerns included checking for appropriate test content (rather than items that reflect Caucasian middle-class values, ensure applicability of the items or have a separate test version), appropriateness of standardizations samples (minorities may need their own norms), examiner and language biases (such as over-pathologizing or lack of familiarity with the language or culture), equitability of social consequences (ensuring it is an individual deficit rather than a systemic problem), ensuring measurement of similar constructs, assessing for differential predictive validity, and differences in test taking ability.

Diversity in personality research

Assessment of the influence of culture on personality has been a relatively recent undertaking. Sue (1996) evidenced a variety of research on biases resulting in overpathologizing African Americans and other minority populations in assessments, while underpathologizing may also occur in some situations. One such study indicated that less acculturated Asian American students had greater elevation on the Minnesota Multiphasic Personality Inventory-2 (MMPI-2), while more acculturated Asian American students had greater elevations than Caucasians. Sue did indicate that response sets of acquiescence or agreement and social desirability may also play a role in cultural differences; the Asian Americans may not actually have more pathology, but may be using an assessment that may not be valid for different minority groups.

C'de Baca, Castillo, Mackaronis, and Qualls (2014), highlighted the importance of using measures which are culturally cognizant. They discovered that African American veteran participants who experienced post-traumatic stress disorder (PTSD) were more likely to encounter racial discrimination, and were more likely to elevate on items of Paranoid PD within the Millon Clinical Multiaxial Inventory, Third Edition (MCMI-III). This resulted in a three times greater chance of a diagnosis of a Cluster A personality disorder for this population when compared to Caucasian participants. Women with a trauma history were twice as likely to have a Cluster B or C personality disorder. The Personality Assessment Inventory (PAI) Italian version and Argentinean version were both assessed for validity in their respective populations (Pignolo et al., 2018; Stover, Solano & Liporace, 2015). The Italian version was found to have convergent validity, while the Argentinean version was found to have gender differences for most of the

scales and subscales. These two studies together indicate varying validity for the use of a test in other cultures.

Ascoli, Lee, Warfa, Mairura, Persaud, and Bhui (2011) stated that, "So far, little has been published about Race, Ethnicity, Culture and Personality Disorders. The wider topic of the interface between culture, race, ethnicity and Personality Disorders still remains relatively unexplored, as compared to other aspects of Personality Disorders. People from Black Minority Ethnic (BME) communities access PD services at a lower rate than the general population. This has been attributed to general patterns of misdiagnosis and lower access to talking therapies." They highlight the need for diagnosis and research to recognize that each society will promote and praise certain personality attributes, and have its own standards of normality and what is abnormal or atypical. Strauss (1979) indicated that within a diagnostic category, cultural differences in symptom profiles may emerge. Previous research on prevalence of Schizotypal PD within the NESARC indicated that odds were greater for meeting criteria if participants were female African American, and lower if they were male Asian participants (Pulay et al., 2009). A meta-analysis of personality disorders and ethnicity indicated some differences in African American and Caucasian diagnosis rates, with African Americans receiving less diagnoses (odds ratio 0.476), with no differences in Hispanic or Asian groups (McGilloway, Hall, Lee & Bhui, 2010).

With regards to validity of scales for personality disorder assessment, some studies have indicated varied findings. An additional study indicated that Europeans and Americans generally scoring higher in Extraversion than Asians and Africans on the NEO-PI-R (McCrae & Terracciano, et al. 2005). One study indicated that low rates of personality disorder in Asian-origin samples was likely related to lack of understanding cultural contexts (Ryder, Sun, Dere &

Fung, 2014). The PID-5 (for DSM5) demonstrated acceptable reliability and structural validity with factor analysis between Norway and USA (Thimm, Jordan & Bach, 2017). The MCMI-2 had different profile patterns for Native Americans and non-Native Americans (Glass, Bieber & Tkachuk, 1996). Dana (1983) stated that objective tests require local norms for valid assessment, and advised caution for interpretation with Native Americans for the MMPI, MCMI, and 16PF (as reviewed in Thomason, 1999). This opinion was also endorsed by Triandis and Suh (2002) in their review of culture and personality, who stated that as most studies have not included culturally specific (emic) traits and references in addition to more universal (etic) traits, and that the standardization and sample studies completed are typically similar in culture to Western samples.

The MMPI has been a source of some information on racial differences in personality disorder presentation. Less than 1% of the original MMPI standardization sample was Native American (less than national percentage). In one study, Native American groups had higher T scores on five validity and clinical scales (L, F, 4 [Pd], 8 [Sc], and 9 [Ma]), six content scales (DEP, HEA, ASP, CYN, BIZ, and TRT), and two supplementary scales (MAC–R and AAS) than the MMPI–2 normative group. (Robin, Albaugh, Greene, Caldwell & Goldman, 2003). In another study, 30 items from scales F, 1, 6, 8, and 9 had differential item functioning in endorsement rates for Native Americans versus the normative group (Hill, Pace & Robbins, 2010). Nine themes emerged from that study: core belief system, experiences of racism and discrimination, conflicting epistemologies, living in two worlds, community connectedness, responsibility and accountability to the community, traditional knowledge, stories as traditional knowledge, and language and historic loss. MMPI-2 may pathologize Indigenous world views, knowledge, beliefs, and behaviors rather than psychopathology. Regardless of diagnosis, the

Native American profiles were all similar across three additional studies, with significant elevations on scales F, 1, 4, 6, 8, 9 appearing most often (Dahlstrom, 1986; Robin, Greene, Albaugh, Caldwell, & Goldman, 2003; Pace et al., 2006). Ojibwa and Cree had elevations on all clinical scales resulting in a 50% misclassification rate.

Studies on African American responses to the MMPI-2 have yielded results indicating no significant bias compared to Caucasians when using a logistic regression approach (Castro, Gordon, Brown, Anestis & Joiner, 2008). This was also seen in a study by McNulty, Graham, Ben-Porath and Stein (1997). Timbrook and Graham (1994) also examined African Americans, and discovered a small difference in each group and for each gender. They controlled for factors of education, income and age and had no remaining difference. Unfortunately, factors such as SES and education are often tied to cultural differences. Dana and Whatley (1991) reviewed existing literature on MMPI use in African Americans, and indicated that even once sociodemographic variables are controlled for, item differences remain (though diminished in magnitude), particularly for scales F, 8 and 9. Monnot, Quirk, Hoerger and Brewer (2009) examined the MMPI-2 for bias within an African American veteran population. They discovered elevated scores across most scales, with significant elevations on three scales (2, 9, bipolar, ASP and ANG, with scales 1, 3, 8 with PTSD participants and 4 with gambling addiction participants showing the larges bias). Dana (1988) noted that MMPI items have shown differences in responding for African Americans in a large number of items across studies, from 22% to 39% of items, though these items were not always consistent, and 200 items with possible discrepancies. He cautioned that interpretation should include the client's history and cultural background. A regression analysis of the MMPI-2 for African Americans compared to Caucasians indicated several scales with potential bias, but that the magnitude of difference was small (Arbisi, Ben-

Porath & McNulty, 2002). Most scales indicated that African Americans were more likely to have underprediction of psychopathology, but they were more likely to experience hallucinations which elevated scale 8. Their review indicated other studies with group differences on scales 6, 8, 9 being cited most often. Overall, African American differences in responding on personality assessments have been inconsistent at best.

In general, examination of assessments for personality disorders has been extremely limited in Asian populations. Asians and Asian-Americans have been noted to have lower rates of personality disorders when compared to other ethnicities, though the reasons for these low rates and differences in the presentation of the disorders remain obscured (Ryder, Sun, Dere & Fung, 2014). They may need more sever presentations of some disorders before being considered pathological, such as with dependent personality disorder, in which criteria such as expressing disagreement and needing others to assume responsibility are considered more normative in the general population in a collectivist culture. Use of the MMPI in Asia has indicated a need for separate norms for the country it is to be used in (Butcher, Cheung & Lim, 2003). While most exist now with their translations, some do not.

Hispanic populations have also had limited research done on personality assessment disparities. A review by Malgady, Rogler and Costantino (1987) reported that Hispanics have significant differences on select MMPI scales and items, particularly those that deal with belief in spirits. Velasquez, Hallahan and Young (1993) attempted to control for differences in MMPI presentation by controlling for age, education and psychiatric diagnosis, in a discriminant functional analysis much like African American studies had before. Results did not favour this, indicating that differences remained after these moderators were controlled for. The MMPI was still able to discriminate between ethnic groups on scales L, 5, 6, 7, 8, and 9. A meta-analysis of

MMPI and MMPI-2 studies of African American and Hispanic participants indicated trivial but present differences between the groups when comparisons are made with Caucasian participants, with few results of a more robust effect size (Hall, Bansal & Lopez, 1999). It noted that male African Americans exhibited higher scores on seven scales, and lower scores on five scales, while females had higher scores on eight scales, and lower scores on four scales. Moderate effect sizes remained on scales L, F, 8 and 9 for African American males, and scales 5 and 9 for females. Latino/Hispanic Americans demonstrated higher scores on three scales and lower scores on ten scales. Of these, scales L and 5 remained with notable effect sizes. The authors reported that overall, the MMPI had a fair portrayal of both minorities.

Puente (1990) recommended not using the MMPI with minority clients at all, due to the lack of understanding of ethnic differences or how to apply that understanding to the interpretation of scores, unless they are highly acculturated and have primarily Western values. Marsella and Pedersen (1981) concluded that "attempts to adopt personality tests to diverse cultures [are] unhelpful since they are less than adequate even in their culture of origin" and recommended to avoid standardized personality tests and projective tests altogether. They suggested the use of symptom checklists and behavioural observations for diagnoses.

Diversity considerations within the DSM5

The APA's DSM series has a history of warning clinicians against under- or over-diagnosing personality disorders based upon stereotypes of gender roles and behaviours (APA, 2000, p. 688). However, within each disorder, little information on gender differences in the presentation of a disorder exists. The DSM-IV-TR contained prevalence data on gender differences in personality disorders, however, the DSM5 has shied away from this in this edition.

The DSM5 does not include information regarding differences in the prevalence rates of personality disorders by race as it does with gender. It does, however, indicate that "judgments about personality functioning must take into account the individual's ethnic, cultural, and social background. Personality disorders should not be confused with problems associated with acculturation following immigration or with the expression of habits, customs, or religious and political values professed by the individual's culture of origin" (APA, 2013, p. 648). For example, Paranoid Personality Disorder (Paranoid PD or PPD) may be mistaken in those who display guarded or defensive behaviours due to unfamiliarity, neglect or mutual mistrust. Clinical judgment should be used in addition to validated measures. However, this is not always taken into account, particularly in areas of research or test construction.

Cluster A personality disorders include Paranoid (PPD), Schizoid (SZPD), and Schizotypal (STPD) personality disorders. Cluster A disorders are noted to be predominantly odd or eccentric. They have a combined prevalence of 5.7%. Paranoid PD has an estimated prevalence of 2.3-4.4%. Part II of the National Comorbidity Survey Replication (NCSR) suggests a prevalence of 2.3%, while the National Epidemiologic Survey on Alcohol and Related Conditions (NESARC) data suggest a prevalence of paranoid personality disorder of 4.4%. Paranoid PD is marked by pervasive distrust and misinterpretation of motives of others as malevolent. The DSM5 (APA, 2013) indicates that there are many socio-cultural or life circumstances which must be taken into account in diagnosis, such as having a guarded or defensive behaviour because of unfamiliarity, language barriers, lack of knowledge of social rules, cultural behaviours which may appear or be misinterpreted as paranoid in nature, or from perceived neglect. It may be more commonly diagnosed in males. Schizoid PD is predominantly a disorder of detachment from social relationships and restricted emotions. NESARC suggests a

prevalence of 3.1%, while the NCSR suggests 4.9%, with more males receiving diagnosis and having more impairment. Cultural issues for interpersonal relations and defensive behaviours, as well as migration to a new environment (such as rural to metropolitan, or immigration to a new country) can result in emotional and interpersonal restriction to solitary activities, and being perceived as cold or hostile were indicated in the DSM5 (APA, 2013). Schizotypal PD is defined by a pattern of social and interpersonal difficulties, cognitive or perceptual distortions, and unusual behaviour. It may be slightly more common in males. Cultural issues of religious beliefs and rituals must be taken into consideration during evaluation. The DSM5 notes some rituals such as "voodoo, speaking in tongues, life beyond death, shamanism, mind reading, sixth sense, evil eye, magical beliefs related to health and illness" may appear to be schizotypal in nature if uninformed on these beliefs (APA, 2013). The 2004-2005 NESARC data indicates prevalence of 3.9 to 4.6% in the general population, while it is seen infrequently in clinical populations (0-1.9%). It has varying prevalence in other countries (0.6% in Norway).

Cluster B disorders are marked by their tendency to appear dramatic, emotional or erratic. Four personality disorders exist within this cluster, Antisocial PD, Borderline PD, Narcissistic PD, and Histrionic PD. Cluster B personality disorders have an overall prevalence of 1.5% within the general population according to the NESARC data. Antisocial PD is marked by patterns of disregard for the rights of others, with evidence that conduct disorder behaviours began prior to the age of 15. It is more common in males, though the DSM5 noted that this may be due to the emphasis on aggressive items in conduct disorder (APA, 2013). It is associated with lower socioeconomic status, and is more common in urban settings, and may be a part of a more protective survival strategy for at risk individuals. Overall prevalence rates are from 0.2 to 3.3% from previous DSMs, while 3.3% of the 2001-2002 NESARC data met criteria for Antisocial

PD. Borderline PD is characterized by unstable relationships, self-image, and affect, combined with impulsivity. It is primarily diagnosed in females. The DSM (APA, 2013) indicates prevalence estimates from 1.6 to 5.9% in the general population, with as high as 20% among psychiatric inpatients. Young adults and adolescents and those experiencing existential dilemmas or emotional decisions are more likely to display behaviours that resemble Borderline PD. The 2004-2005 NESARC data indicates a prevalence rate of 6.8%. Histrionic PD is defined as a prolonged and pervasive pattern of emotionality and attention seeking behaviours that has been more frequently diagnosed in females. The NESARC data has indicated a prevalence of 1.84%. Cultural considerations for this disorder must consider whether the behaviours cause distress or impairment, given that interpersonal behaviours, emotional expression and physical appearance vary so significantly across cultures (APA, 2013). Narcissistic PD has diagnostic features which indicates patterns of cognition and behaviours of grandiosity, minimal empathic ability, and a need for admiration. It is primarily diagnosed in males. Previous DSMs had prevalence estimates from 0 to 6.2%. The 2004-2005 NESARC data identified a prevalence of 7.6%. The DSM5 did not indicate specific cultural considerations.

Cluster C personality disorders are noted to be anxious or fearful. Cluster C personality disorders were noted within the NESRC data to have a combined prevalence of 6.0%. It consists of Avoidant PD, Dependent PD, and Obsessive-Compulsive PD. The NESARC data indicated that Avoidant PD has a prevalence of 2.4%. It is diagnosed equally in men and women. It is characterized by patterns of cognitive distortions of inadequacy and sensitivity to negative evaluation, and social inhibition. Cultural considerations noted in the DSM5 are acculturation after immigration and cultural differences in humility or shyness (APA, 2013). Dependent PD diagnosis consists of a need to be taken care of and related behaviours, and fear of separation

from those caregivers. Dependence varies greatly by culture, and should be considered within the context of that culture for distress or impairment. It is diagnosed more in women. Prevalence from the 2001-2002 NESARC data indicated a prevalence of 0.49%, while the NCSR indicated 0.6%. Obsessive-Compulsive PD in comparison, has a prevalence of 2.9 to 7.9%. The NESARC data indicated prevalence of 7.6%. It is diagnosed twice as often in men. It consists of patterns of order and perfection, a need for control and inflexibility. Those which place emphasis on work and productivity or other habits or interpersonal interactions within a cultural group should ensure that diagnosis is warranted for distress or impairment within that culture.

Gender Considerations

Gender is another factor in which personality disorder assessment may be biased. Many personality disorders involve maladaptive levels of gender-related traits, which may influence the presentation and prevalence of some disorders, and may introduce some bias within personality assessment items (Widiger & Samuel, 2005). This may lead to over- or under-diagnosis for some disorders. The DSM warns clinicians not to over- or under-diagnose personality disorders based on social stereotypes of gender roles/behaviours (APA, 2000, p. 688). However, within the diagnostic text, little information on gender differences in presentation is available. Costa, Terrracciano and McCrae (2001) noted that gender differences were most pronounced in European and American societies, and that the impact or magnitude of gender effects on personality disorder presentation varied across cultures. Some studies have shown differences in the presentation of personality disorders by gender, such as the presentation of narcissistic (O'Leary & Wright, 1986) and schizotypal PD (Bora & Arabaci, 2009). For schizotypal, women were more likely to endorse items involving social anxiety and odd beliefs, while men endorse negative and disorganized symptoms. Berk and Rhodes (2005) suggested that

women are more likely to be diagnosed with dependent PD (see also Crosby & Sprock, 2004), and that men are under-diagnosed. Over-diagnosis of women and under-diagnosis of men for borderline PD has been demonstrated, despite true gender differences in presentation of the PD (Bjorklund, 2006; Skodol & Bender, 2003). Many articles have examined the predominance of antisocial PD diagnosis in men, and of borderline PD in women (see Samuel & Widiger, 2009 for a thorough review; Crosby & Sprock, 2004; Trull, Jang, Tomko, Wood, & Sher, 2010). Issues in gender bias for histrionic PD were suggested as far back as 1978 by Warner, and were voiced strongly by Kaplan in 1983, when she suggested that "a healthy woman automatically earns the diagnosis of Histrionic Personality Disorder" (as cited in Samuel & Widiger, 2009). Other articles (e.g. Lynam & Widiger 2007) have also supported the theory of overdiagnosis of histrionic PD in women, beyond what would be expected from normal prevalence rates. Ford and Widiger (1989) examined gender bias within the diagnosis of antisocial and histrionic PDs, and noted that as far back as the DSM-II, gender bias has existed. They also found gender-biased items within the individual criteria for these two disorders.

With regards to specific personality assessments, the Million Clinical Multiaxial Inventory-III demonstrated that even when gender differences are controlled for by creating a standard score which differs depending on gender, gender differences still exist in the form of bias (Hynan, 2004). A study on gender bias in the diagnostic criteria of PDs using Item Response Theory (IRT) analyses to determine if the assessments may inherently contain gender bias in the construction of the test items was completed (Jane, Oltmanns, South, and Turkheimer, 2007). Results indicated that four items contained gender bias for men, where men were more likely to endorse the items, even though women contained the same level of the trait. Three of the four items were behavioural aspects within the criteria for antisocial PD. Two criteria were biased

towards women in the schizoid PD criteria, which the authors commented that these items made intuitive sense, as they were related to gender stereotypes. On average, men had higher scores on the schizoid, antisocial, narcissistic and avoidant PDs. Women scored higher on the paranoid, schizotypal, borderline, histrionic, dependent, and obsessive-compulsive PDs.

Gender and diversity in the NESARC test construction

The DSM5 states that "Data from the 2001-2002 National Epidemiologic Survey on Alcohol and Related Conditions suggest that approximately 15% of U.S. adults have at least one personality disorder" (APA, 2013, p. 646). Given that the DSM5 uses the NESARC data for prevalence statistics, it is important to ensure that the NESARC items and tests are not biased towards minority populations. Previous (unpublished undergraduate thesis) research by this author in May 2011 indicated that the NESARC data for the original seven personality disorders in the Wave 1 of data collection contained gender bias. While most item specific bias was constrained and did not affect the scale or test level diagnosis rates, the Dependent PD scale indicated a meaningful difference in women receiving a diagnosis when compared with men that was influenced by test bias. Over 17% of the mean difference in observed scores was attributable to differential test functioning (DTF), or test level bias. The remaining portion of variance was attributed to true differences in prevalence rates between men and women. It was noted then that because of the gender bias existing within the Dependent PD scale, that men and women not be compared on this scale and be interpreted with caution with women. Given these concerns, gender bias within the Wave 2 data collected, as well as a possibility of other biases, such as ethnic minority or cultural differences should be examined.

Concerns with NESARC overestimations of prevalence may be at least partially attributed to the construction of the scales requiring that only one item contain significant

distress or impairment in functioning, particularly when compared with other national surveys such as the National Comorbidity Survey Replication. The DSM clearly indicates a more pervasive impairment over more than one item. Another concern with NESARC prevalence estimation is gender bias within the construction of the items and overall scale scores. Cultural biases may also be present. These biases may inflate prevalence estimates in minority populations, and skew overall prevalence rates.

Cultural differences in the expression and social acceptance of personality traits may influence their expression in the more Western culture in which the NESARC data was collected. These differences in expression may appear deviant to an untrained or unfamiliar observer, and were not accounted for in the formulation of test items within the NESARC questionnaire. No questions exist for whether or not the behaviour is a deviation from the participant's culture, or maladaptive within their culture.

Item Response Theory

Item Response Theory is a mathematical model of evaluating items and scales (Morizot, Ainsworth, & Reise, 2007). IRT is based on the latent trait theory, which assumes that a trait or characteristic underlies and causes item responses. It describes the relationship between a person's response to an item, and their level of the latent trait being measured by the scale. It asks an essential question: given a person's score on the underlying construct or trait, what is that person's true score on item when they are viewed as a member of the focus group? What if they were a member of the reference group? If there is true measurement equivalence between the groups within the DFIT framework, then the true score differences would be equal to zero at the subscale and item levels. It estimates the probability of endorsing a response option based on the latent trait.

This relationship is expressed as Item Response Functions (IRFs) which contain an index of the relationship between the item and the measured construct (item discrimination; alpha) and an index of item difficulty (item threshold; beta). Individual differences on a test item are represented by a relative standing on the latent trait variable being tested. IRFs represent the item-trait regression functions for focus and reference groups. People from different groups with the same scores on the latent trait may respond to an item differently. How much the two IRFs differ from each other indicates amount of differential item functioning at the item level (or differential test functioning at the scale level).

Differential item functioning (DIF) occurs when persons with the same level of a latent trait respond differently to an item because of their group membership. Differential test functioning (DTF) has the same premise, but for the scale's total score. When differential item functioning or differential test functioning occurs, it can be inappropriate to compare groups on the item or scale in question, because the groups may naturally be different with regard to the latent trait being assessed. To compare them would be inaccurate and unfair to the groups. If there is no true difference between the groups, but they still have differences between their levels of the latent trait, it is possible that some form of test bias exists (whether it is gender or race bias, or some other kind of test bias).

DIF can be computed for both dichotomous and polytonomous scoring, and for uni- and multi-dimensional models, though is more accurate with unidimensional models (Oshima & Morris, 2008). Observed differences in personality disorder responses or scores may be due to item and test bias, or to true differences between the two populations (or to both). As only IRT (and not classical test theory) has the ability to distinguish between group differences and bias, the use of IRT analyses is essential to this study.

Differential functioning can be uniform or non-uniform in nature. Uniform differential functioning is when the probability of endorsing item in the coded direction is consistent for group at virtually all theta levels. Nonuniform is more flexible. It allows for group specific item response functions, varying based on level. For example, in groups of individuals with low trait levels, the reference group may endorse the item more frequently than focus group (and vice versa), but this may reverse at high trait levels. DIF is assessed by comparing the item a-parameter for non-uniform DIF, and b-parameter for uniform DIF (Elosua & Wells, 2013).

There are also differences in calculation for the contribution to the scale or differential test functioning (DTF). Compensatory differential item functioning (CDIF) is when an item has an additive contribution to a scale's DTF (no interactions). An item may have large chi-squared ($\chi 2$), but little CDIF if DIF was in opposite direction of other items. Non-compensatory differential item functioning (NCDIF) is the average squared difference between the expected item endorsement probabilities. For this approach, calculations of the probability of the focus group participant item *j* will be endorsed using the parameters (estimated trait levels) from the reference group. Then it is calculated the from the focus group. The difference between these two probabilities is then squared, and then the weighted average of the squared differences for all focus group participants in the sample is calculated. This approach incorporates products of individual item response probability curves. It should be noted that at the test level, particular item biases may cancel out at the scale level (DTF). This is referred to as DIF cancellation. Alternatively, amplification may occur, where an item's contribution increases the level of DTF.

IRT has several assumptions which must be examined prior to analyses. It assumes local independence (if the latent trait is held constant, there should be no association or correlation less than 0.1 among item responses), unidimensionality and model fit. It requires a sample size of 200

or 250 as a minimum, while 500 is ideal. This sample should be heterogeneous. There are several models (1PLM, 2PLM and 3PLM) to choose from based on the number of parameters used. The 1PLM model has the item and person on same trait continuum. It uses item difficulty and discrimination to determine how person will do on an item. However, it only contains items location, and is considered the most restrictive. In the 2PLM, items do not have to share a common slope (continuum). Item discrimination can vary across items to determine person's location on the trait, and it has item location and discrimination capacity. It is often used with dichotomous data. 3PLM models add in a guessing or chance parameter. It includes item location, discrimination capacity and a lower asymptote (pseudo-guessing for chance success on people with low theta).

There are several forms of IRT analyses and ways to determine differential item functioning. A thorough review was completed by Millsap and Everson on this topic in 1993. Likelihood ratios are capable of detecting bias at the scale level, but are unable to determine which items contributed to the bias without post hoc item-by-item analyses. Discriminant function analysis is best used for polytonomous item responses. The data set in use for this study is dichotomous in nature, however, and this is not the best fit. Confirmatory factor analysis is also often used to determine differential functioning. It does this by providing regression estimates of factor scores and examining the differences in item intercepts across populations. However, it does not differentiate between impact and bias. IRT analyses also make use of logistic regression and Mantel-Haenszel chi-square analyses, but linearly transform the models and examine the differences between the regression slopes discrimination and difficulty parameters to detect bias, assumes nonlinear relationship between the underlying/latent construct and the observed score at the item/subscale level. For dichotomous IRT analyses, the probability

of answering an item is expressed as a nonlinear logistic function within a logistic regression model (Raju, Laffitte & Byrne, 2002). This allows for standard error of measurement to vary from person to person. Item response functions are compared across the two populations for equality. It examines the interaction of person and item, plot probability of response given underlying characteristic based on the slope (discrimination) and item difficulty or 50% endorsement rate (inflexion point of ICC). IRT can look at test total scores for bias as well as individual items, and some forms of IRT can estimate the amount of bias contributing to the score (Raju's DFIT). Raju's DFIT analyses were unable to be completed as Census statisticians did not have access to this program. Raju's area under the curve model uses the area between the regression lines to determine differential item functioning. Mantel-Haenszel chi-square test for differential item functioning is used primarily to detect uniform bias only. Results are in a chi-squared format. This form is best used with simple models, as it can overestimate bias within complex models. Logistic regression IRT can detect for both uniform and non-uniform DIF, and also uses a chi-square significance test. Both of these methods are loglinear models, which do not work well with multiple parameters.

Item Response Theory in Research

Sharp, Goodyer and Croudace (2006) compared results from confirmatory factor analysis and 2PLM IRT on moods and feelings questionnaire; another example is a DIF IRT assessment of the PROMIS depression items for gender, age and education (Teresi et al. 2009) of polytomous data.

IRT analyses have been used effectively in other personality research. A study of normal and abnormal personality scales to determine if the 'abnormal' personalities were a maladaptive extreme of the five factor model indicated that the measures shared a common dimensional

structure (Samuel, Simms, Clark, Livesley, & Widiger, 2010). This study of multiple measures was made possible because of the ability to determine the underlying level of the personality trait in question, and compare across multiple test items. Their results lent support to the concept that personality disorders are a maladaptive or extreme level of the latent personality traits that exists within everyone. Jane, Oltmanns, South and Turkheimer (2007) completed a study of gender bias in diagnosis of personality disorders for the DSM-IV using IRT for polytomous data. IRT analyses of the NESARC data for older and younger adults indicates older adults more likely to receive obsessive-compulsive PD and schizoid PD and less likely to receive avoidant or dependent PDs (Balsis, Woods, Gleason & Oltmanns, 2007). IRT analysis has been used in a variety of studies on bias within personality disorder diagnosis. IRT analyses of the NESARC Wave 1 data for gender using Raju's DFIT model indicates the dependent scale has a bias towards women, and many items that contain bias, but cancel out at the scale level (Scoullar & O'Brien, 2011, unpublished Honour's Thesis). The DTF for the dependent scale accounted for 17% of the variance observed in women. The antisocial scale also had significant DIF and a significant DTF, however, with the measurement being used (Raju's DFIT model), a stringent cutoff is used to balance out the large sample size. This resulted in the antisocial scale being noted as having a negligible influence of bias for gender.

Item response theory has also been used to examine measurement bias within the Minnesota Multiphasic Personality Inventory (MMPI-2; Waller, Thompson, & Wenk, 2000). In this study, the MMPI uni- and multi-dimensional scales were assessed for gender bias at the item and test level using an IRT 2 parameter logistic model. They then performed differential item and test functioning analyses to determine that even though individual test items may contain gender bias, that the overall scale will not yield biased test scores. This study was essential for the

present study, as it showed that IRT, despite its assumption of unidimensionality, can be used for multidimensional scales of personality. The authors maintain that IRT provides the strongest method for detecting differential functioning within item and test level scores in group bias research.

Present study

Many studies have been completed using the NESARC data waves, though no completed studies could be found on validity of test construction for minority populations, or gender differences. This study will perform an examination of items for alternative response modalities or bias within test item and scale construction for race and/or gender using an IRT analysis. This can be used to inform prevalence rates, and build upon my previous research. This study hopes to shed light on any potential biases within the test construction of the personality disorders contained within section 10 of the NESARC Wave 1 and 2 questionnaires. It is anticipated that cultures which are more dependent in nature, such as collectivist cultures or participants with an Asian or Hispanic background, will have biased response patterns on items related to Dependent PD, while Native American populations will be more likely to respond to items positively with a long term spiritual belief in mystical or magical ideas. African American populations are anticipated to be more likely to endorse items of persecution. Are there differences in minority populations responding rates to mental health questions at the item or test level within the data when compared to Caucasian participants?

Method

Data was collected through the National Institute on Alcohol Abuse and Alcoholism's (NIAAA) National Epidemiologic Survey on Alcohol and Related Conditions (NESARC; Grant & Dawson, 2006). The survey contained three waves of data collection (2001-2002, 2004-2005,

and 2012-2013), with Wave 1 pertaining to participants lifestyles before the survey was conducted. The Wave 1 interview form contained information regarding seven personality disorders (avoidant, dependent, obsessive-compulsive, paranoid, schizoid, antisocial and histrionic personality disorders). Wave 2 focused on the period since Wave 1 had occurred and consisted of 34,653 participants. It contained items related to four personality disorders (borderline, schizotypal, narcissistic, and antisocial). Wave 3 consisted of an independent sample of 36,309 participants and used a different interview form, which contained information on only two personality disorders (borderline and schizotypal). Potential participants were first contacted in writing about the nature and statistical uses of the information they would be providing. Interviews were conducted face-to-face within the participant's home. Given that the criteria for personality disorders in the DSM-5 have not changed from those in DSM-IV, it was determined that items from the Wave 1 and 2 interviews were still appropriate and relevant today.

Participants

Wave 1 included 43,093 adult participants age 18 or older from across the United States. It consisted of non-institutionalized participants only. Of those who were contacted for interview, there was an 81% overall response rate. Ethnic minorities and young adults under the age of 24 were oversampled during the data collection phase. The data was later adjusted to reflect accurate sociodemographic information collected by the U.S. census in variables such as age, sex, and ethnicity. It included several ethnic minorities present in the population, with 8,600 (20.0%) of participants being African-American, and 8,308 (19.3%) Hispanic/Latino. It also contained 1,304 (3.0%) Native American or Native Alaskan and 363 (0.8%) Native Hawaiian. 1,334 (3.1%) were of Asian descent. The remaining 76.1% identified as Caucasian. 43% or 18,518 of respondents identified as male, while 57% or 24,575 identified as female.

Wave 2 consisted of 34,653 of the Wave 1 participants. Participant drop out was attributed to impairment or active duty military service (3,134 participants), and inability to locate or refusal to participate (5,306 participants). This yielded an 86.7% response rate from the original wave.

Measures

During Wave 1 and 2, the Alcohol Use Disorder and Associated Disability Interview Schedule – DSM-IV Version (AUDADIS-IV) was completed (Grant & Dawson, 2006). This structured interview consists of a series of yes/no questions. Although the primary purpose of the study was alcohol-related, participants were also interviewed about medication and drug use, family history, gambling, medical conditions, and psychological disorders (including depression, anxiety disorders, and personality disorders). Participants were interviewed using questions which utilized the DSM-IV criteria for personality disorders. Seven of the ten personality disorders described in the DSM-IV were contained in the interview. These were Paranoid, Schizoid, Antisocial, Histrionic, Avoidant, Dependent, and Obsessive-Compulsive. The remaining three, Borderline, Schizotypal, and Narcissistic, were completed in Wave 2 of data collection. The respondents were asked about how they felt or acted most of the time throughout their lives, regardless of situational factors, and were reminded not to include times when they were depressed, drinking, using drugs, or were physically ill. To receive a 'diagnosis,' participants had to endorse at least one symptom which caused significant social or occupational dysfunction, above and beyond the requisite number of positive responses. A subsample of 2657 respondents were contacted again to assess test-retest reliability 3 to 20 weeks after their original assessment, which was moderate (.40) for Histrionic PD to good (.67) for Antisocial PD (Cox, Sareen, Enns, Clara, & Grant, 2007).

The reliability of all of the personality disorders measured was excellent, with items contributing to only one latent trait (Balsis, Woods, Gleason, & Oltmanns, 2007). The reliability and validity of the study for both clinical and non-clinical samples has been shown to be good across international settings, of which several studies evidencing this are available on the NESARC website to review (Grant & Dawson, 2006; Cox, Sareen, Enns, Clara, & Grant, 2007). A study confirming the structure of the clinical interview used in the NESARC data showed support of the hierarchical organization of the personality disorders within the DSM-IV, meaning that the structured interview contained accurate diagnoses according to DSM criteria (Cox, Sareen, Enns, Clara, & Grant, 2007). However, one study suggested that the original NESARC criteria for diagnosis of PDs was too lenient, and did not assess a key component in PD diagnosis (Trull, Jahng, Tomko, Wood, & Sher, 2010). This component was the associated distress or impairment resulting from the PD. The original NESARC criteria requires only one item to cause significant distress or impairment. When this study took distress/impairment into consideration for every item, the NESARC data was suggested to overdiagnose PDs (9.1% prevalence, vs. the 21.5% suggested by the original NESARC analyses). These prevalence rates are closer to those found in other suggested prevalence rate studies (Trull, Jahng, Tomko, Wood, & Sher, 2010). The original diagnostic criteria as outlined above was used in analyses, as no further support for these findings has been suggested at the present time.

Procedures

IRT methods were used to determine if responses to individual items were indicative of the diagnosis which was received by the participant. In particular, the two parameter logistic model (2PLM) IRT analysis was used, as it is the most appropriate to measure personality scales

like the MMPI because of its ability to vary in both threshold and discrimination parameters, and as such will be used to assess the AUDADIS-IV (Waller, Thompson, & Wenk, 2000).

However, before the IRT analyses were calculated, checks for the assumption of unidimensionality had to be completed. The unidimensionality assumption posits that one underlying factor accounts for a person's responses to a question within a scale. Three statistical unidimensionality checks were completed, as no single statistical method ever provides sufficient information regarding unidimensionality. These checks were completed, for if the model was unidimensional, it would be possible to be more confident in the analyses using the 2PLM unidimensional tests; though IRT is robust to violations of this and a multidimensional model exists if needed.

First, a Modified Parallel Analysis (MPA) was completed, where the second eigenvalue of the data is compared to the second eigenvalues that occur in 100 randomly generated data sets (Monte Carlo simulation, 100 iterations). A non-significant p-value indicates unidimensionality. This test is sensitive to sample size and, therefore, Type 1 error; as such, is not considered strong evidence towards model fit. With large sample sizes, like the ones in the NESARC data set, even minimal differences between the observed and randomized samples are deemed significant. Drasgow & Lissak (1983) completed Monte Carlo simulations, and determined that this procedure is able to detect unidimensionality violations that interfere with parameter estimation, like those used in IRT. It was noted that the MPA procedure is good at detecting low (10%) and moderate (25%) levels of contamination, but poor with high (50%) contamination (Budescu et al., 1994).

Next, a maximum likelihood ratio test was completed, which evaluates the relative fits for a two-dimensional and one-dimensional models. A non-significant p-value indicates

unidimensionality. This test is also sensitive to sample size and prone to Type 1 error, and as such, is not considered strong evidence towards model fit. With large sample sizes, like the ones in the NESARC data set, even minimal differences are deemed significant. As such, multidimensionality is anticipated, with a two factor model providing better fit. The maximum likelihood ratio is often used in IRT analyses and other model fits (like confirmatory factor analysis), such as one completed by Chalmers & Flora (2014) which assessed its use with non-compensatory IRT models.

Finally, a comparison of the ratios of the first and second eigenvalues was completed. A ratio greater than 3.0 will indicate support for unidimensionality, as this indicates that there is one predominant factor accounting for the majority of the variance (Morizot, Ainsworth, & Reise, 2007). A ratio below 3.0 indicates more than one factor may be present, and that further assessment should proceed with caution for IRT analyses. As this test does not rely on p-levels and is not so easily influenced by sample size, it will provide a more accurate assessment of the dimensionality of the scale. This is particularly important when there has been multidimensionality present on other tests of model fit. This facet of Principal component analysis is widely used. If the first eigenvalue is substantially larger than the second eigenvalue, then it has been determined that the unidimensionality assumption is likely to hold. (Chou & Wang, 2010).

Once unidimensionality had been assessed and the 2PLM IRT analyses have been completed, the next step was to assess the differential functioning of items and tests (DFIT). In this procedure, item response functions (IRFs) are translated into a common metric. IRFs represent the item-trait regression functions for focus and reference groups. How much the two IRFs differ from each other indicates amount of DIF at the item level (or DTF at the scale level).

Should there be no DIF, the groups will have the same equated total at the same levels of the latent trait. If not, then DIF exists. Given that the NESARC data is presently managed by US Census, data analyses were run by a statistician at Census, Jahn Hakes. IRT analyses available for use by Census included Mantel-Haenszel and Logistic Regression approaches. Given the ability of the logistic regression IRT to calculate differences for both uniform and non-uniform data, this approach was utilized. It is noted, however, that both of these approaches use the total score as a measure of the latent trait, and that this approach may not work if the underlying trait being measured is multiparameter (Millsap & Everson, 1993). Both of these measures also use chi-squared ($\chi 2$), which are susceptible to sample size and can inflate the Type 1 error rate, however the logistic regression approach has a lower power for detecting non-uniform DIF (Elosua & Wells, 2013), which may balance this somewhat in large samples. Unlike other forms of regression, IRT logistic regression models use a nonlinear logistic function (Raju, Laffittee, & Byrne, 2002). The logistic regression approach to IRT uses scores on a latent trait to determine the differences between the groups. For IRT analysis of dichotomous data like the NESARC, logistic regression is used to determine the item response functions (IRFs) to allow for a standard error of measurement to vary for each participant (Raju, Laffitte, & Byrne, 2002). Differences were calculated using STATA's IRT analysis DIFLogistic. Differential item functioning between groups was assessed by performing a logistic regression analysis for each item using the reference group (Caucasian) and the focal group (minority). Each analysis provides an estimate for the constants and regression coefficients for the reference group and focal group, respectively. If the constants and regression coefficients are the same, then the predicted probability curves (ICCs and TCCs) are the same, and no DIF is present (de Ayala, 2009). DIF is determined if the constants or the regression coefficients are not equal. Uniform DIF is present if

the regression coefficients are equal, but the constants are not; this would result in parallel probability curves. Non-uniform DIF is present if the regression coefficients are not equal, but the constants are; this would result in probability curves which cross. Significance testing provided by this method is in chi-squared format.

Finally, Mantel-Haenszel (MH) differential item functioning was also completed using STATA's DIFMH to compare differential functioning to the logistic regression model. Odds ratios were also calculated. This allowed for an estimation of the magnitude of effect any significant test or item differential functioning existed (Monahan, McHorney, Stump & Perkins, 2007). It was examined to determine the impact or effect that this significance would have upon the total score of participants as a group being x times more or less likely to have a response in a certain direction. In cases of non-significance with notable odds ratios, true differences between the groups, or possible undetected differential functioning may be present. An additional measure of effect size was examined, Nagelkerke or Pseudo R2 (Zumbo, 1999). This measure of effect size examines changes in R2 due to insertion of variables of ethnicity or gender. Unfortunately, as this measure does not account for specific groups, but rather examines effect overall (much like an ANOVA without the pairwise comparisons), specific differences were not able to be checked for effect in this measure.

Results

It was expected that Asian, Hispanic, and Native American participants would have bias within item and test level responses in the dependent personality scale, and that this would be increasingly important for female participants, who already have a known bias within the dependent scale. It was expected that these issues will compound, and account for more than the 17% of variance previously discovered. It was also anticipated that African American

participants would have increased responding rates to paranoid personality disorder items of persecution, though these results were not anticipated to show within the overall scale results. It was suspected that those cultures with spiritual beliefs outside of Western Christianity would also have altered response patterns on items related to magical thinking, though it was not anticipated that the overall schizotypal scale be affected.

Weighted means and standard deviations for gender and ethnicity for each personality scale can be seen in Table 1. Number of participants for each group in the sample are presented in Tables 2a (ethnicity and gender) and 2b (ethnicity divided into gender). It should be noted that these numbers have been rounded following U.S. Census Bureau privacy policy, and as such, numbers are not exact. However, analyses were run using the exact numbers. Those groups with less than 15 participants have N/A indicated in the column, as per U.S. Census policy, and those analyses could not be run given privacy concerns.

Probability or percent of prevalence can be seen in Tables 2c (ethnicity and gender) and 2d (ethnicity divided into gender). Across all participants, obsessive-compulsive PD had the highest prevalence, at 7.88%, and dependent PD had the lowest, at just under 0.5%. For men, obsessive-compulsive and narcissistic personality disorders were most prevalent, at 7.87% and 7.69%, respectively. For women, obsessive-compulsive and borderline personality disorders were most prevalent, at 7.89 and 6.18%, respectively. For both men and women, dependent PD was least prevalent, at 0.37% and 0.61%. Prevalence data was not available for Native and Asian Americans for dependent PD, due to less than 15 participants endorsing the disorder, while African American and Hispanic participants had lower prevalence rates than the average across all participants. Caucasians had the highest rates for dependent PD, at 0.53%. Higher prevalence rates were seen across all remaining disorders for Native and African American participants,

Table 1

Weighted Means and Standard Deviations

Gender					
Personality Disorder	**Male N1=18500 N2=14500**		**Female N1=24500 N2=20000**		
Antisocial	0.0550 (0.2279)		0.0190 (0.1364)		
Avoidant	0.0191 (0.1369)		0.0277 (0.1640)		
Dependent	0.0037 (0.0606)		0.0061 (0.0778)		
Obsessive-Compulsive	0.0787 (0.2693)		0.0789 (0.2697)		
Paranoid	0.0382 (0.1917)		0.0497 (0.2173)		
Schizoid	0.0319 (0.1758)		0.0308 (0.1729)		
Histrionic	0.0189 (0.1363)		0.0181 (0.1332)		
Schizotypal	0.0424 (0.2015)		0.0364 (0.1873)		
Borderline	0.0559 (0.2297)		0.0618 (0.2409)		
Narcissistic	0.0769 (0.2665)		0.0477 (0.2132)		
Ethnicity					
Personality Disorder	**Native American (1) N1=850 N2=700**	**African American (3) N1=8100 N2=6400**	**Hispanic (5) N1=8300 N2=6400**	**Caucasian (4) N1=24500 N2=20000**	**Asian American (2) N1=1200 N2=900**
Antisocial	0.0909 (0.2876)	0.0370 (0.1887)	0.0331 (0.1789)	0.0362 (0.1867)	0.0124 (0.1107)
Avoidant	0.0353 (0.1846)	0.0198 (0.1394)	0.0198 (0.1394)	0.0244 (0.1543)	0.0227 (0.1490)
Dependent	N/A	0.0037 (0.0604)	0.0040 (0.0635)	0.0052 (0.0722)	N/A
Obsessive-Compulsive	0.0976 (0.2970)	0.0798 (0.2709)	0.0599 (0.2373)	0.0830 (0.2759)	0.0462 (0.2099)
Paranoid	0.1024 (0.3033)	0.0762 (0.2654)	0.0518 (0.2216)	0.0367 (0.1881)	0.0337 (0.1805)
Schizoid	0.0607 (0.2389)	0.0490 (0.2159)	0.0362 (0.1868)	0.0278 (0.1645)	0.0156 (0.1239)
Histrionic	0.0258 (0.1585)	0.0258 (0.1586)	0.0161 (0.1258)	0.0176 (0.1314)	0.0175 (0.1313)
Schizotypal	0.0672 (0.2506)	0.0678 (0.2514)	0.0390 (0.1936)	0.0352 (0.1843)	0.0183 (0.1342)
Borderline	0.1175 (0.3222)	0.0808 (0.2725)	0.0529 (0.2238)	0.0560 (0.2299)	0.0355 (0.1852)
Narcissistic	0.0754 (0.2642)	0.1243 (0.3299)	0.0754 (0.2640)	0.0501 (0.2181)	0.0520 (0.2222)

Note: N1 indicates number of participants for NESARC wave 1; N2 indicates number of participants for NESARC wave 2; standard deviations are presented in (). All Ns are rounded, as provided by Census. N/A indicates that data was not available as less than 15 participants obtained "yes" scores for scaled scores for this disorder, as indicated in the U.S. Census Bureau's privacy policy. Sources: CBDRB-FY19-422, CBDRB-FY19-448 and CBDRB-FY19-536.

Table 2a.

Descriptives: Number participants meeting criteria by having Total Scale Score of Yes for All participants, Gender and Ethnicity

All Participants and Gender					
Personality Disorder	**All Participants N1=43,000* N2=34,500^**		**Males N1=18,500 N2=14,500**		**Females N1=24,500 N2=20,000**
*Antisocial**	1400		950		450
*Avoidant**	1000		350		650
*Dependent**	200		60		150
*Obsessive-Compulsive**	3300		1400		1800
*Paranoid**	2100		750		1300
*Schizoid**	1400		600		800
*Histrionic**	800		350		450
*Schizotypal^*	1500		700		850
*Borderline^*	2200		900		1300
*Narcissistic^*	2400		1200		1200
Ethnicity					
Personality Disorder	**Native American N1=850 N2=700**	**African American N1=8100 N2=6400**	**Hispanic N1=8300 N2=6400**	**Caucasian N1=24500 N2=20000**	**Asian American N1=1200 N2=900**
Antisocial	70	250	250	850	20
Avoidant	30	150	150	600	30
Dependent	N/A	30	40	100	N/A
Obsessive-Compulsive	90	600	500	2000	70
Paranoid	90	600	450	900	50
Schizoid	50	350	300	700	20
Histrionic	30	200	150	450	20
Schizotypal	60	450	300	750	20
Borderline	90	500	400	1200	40
Narcissistic	60	750	500	1100	50

Note: N1 indicates number of participants for NESARC wave 1; N2 indicates number of participants for NESARC wave 2; standard deviations are presented in (). All Ns are rounded, as provided by Census. N/A indicates that data was not available as less than 15 participants obtained "yes" scores for scaled scores for this disorder, as indicated in the U.S. Census Bureau's privacy policy. Sources: CBDRB-FY19-422, CBDRB-FY19-448 and CBDRB-FY19-536.

Table 2b.

Descriptives: N for Ethnicity x Gender

Personality Disorder	**Native American Male**	**Native American Female**	**African American Male**	**African American Female**	**Hispanic Male**
Antisocial	50	20	150	100	150
Avoidant	N/A	20	50	100	70
Dependent	N/A	N/A	N/A	20	N/A
Obsessive-Compulsive	50	50	200	400	200
Paranoid	50	50	200	400	150
Schizoid	30	20	100	250	100
Histrionic	20	20	70	100	60
Schizotypal	30	30	150	250	100
Borderline	40	40	150	350	150
Narcissistic	40	20	300	500	250
Personality Disorder	**Hispanic Female**	**Caucasian Male**	**Caucasian Female**	**Asian American Female**	**Asian American Male**
Antisocial	90	600	250	N/A	20
Avoidant	100	200	400	20	N/A
Dependent	30	30	90	N/A	N/A
Obsessive-Compulsive	300	900	1100	40	40
Paranoid	300	350	550	30	20
Schizoid	150	300	350	N/A	N/A
Histrionic	70	200	250	N/A	N/A
Schizotypal	150	350	400	20	N/A
Borderline	200	500	700	20	20
Narcissistic	250	650	250	30	30

Note: N1 indicates number of participants for NESARC wave 1; N2 indicates number of participants for NESARC wave 2; standard deviations are presented in (). All Ns are rounded, as provided by Census. N/A indicates that data was not available as less than 15 participants obtained "yes" scores for scaled scores for this disorder, as indicated in the U.S. Census Bureau's privacy policy. Sources: CBDRB-FY19-422, CBDRB-FY19-448 and CBDRB-FY19-536.

Table 2c.

Probability (Prevalence Percentile) for group meeting criteria for a Personality Disorder ("Yes" in scale score) based on weighted means

Personality Disorder	**All Participants**	**Males**	**Females**
Antisocial	0.0362	0.0550	0.0190
Avoidant	0.0236	0.0191	0.0277
Dependent	0.0049	0.0037	0.0061
Obsessive-Compulsive	0.0788	0.0787	0.0789
Paranoid	0.0442	0.0382	0.0497
Schizoid	0.0314	0.0319	0.0308
Histrionic	0.0185	0.0189	0.0181
Schizotypal	0.0393	0.0424	0.0364
Borderline	0.0590	0.0559	0.0618
Narcissistic	0.0617	0.0769	0.0477

Personality Disorder	**Native American**	**African American**	**Hispanic**	**Caucasian**	**Asian American**
Antisocial	0.0909	0.0370	0.0331	0.0362	0.0124
Avoidant	0.0353	0.0198	0.0198	0.0244	0.0227
Dependent	N/A	0.0037	0.0040	0.0053	N/A
Obsessive-Compulsive	0.0976	0.0798	0.0599	0.0830	0.0462
Paranoid	0.1020	0.0762	0.0518	0.0367	0.0337
Schizoid	0.0607	0.0490	0.0362	0.0278	0.0156
Histrionic	0.0258	0.0258	0.0161	0.0176	0.0175
Schizotypal	0.0672	0.0678	0.0390	0.0352	0.0183
Borderline	0.1170	0.0808	0.0529	0.0560	0.0355
Narcissistic	0.0754	0.1240	0.0754	0.0501	0.0520

Note: N1 indicates number of participants for NESARC wave 1; N2 indicates number of participants for NESARC wave 2; standard deviations are presented in (). All Ns are rounded, as provided by Census. N/A indicates that data was not available as less than 15 participants obtained "yes" scores for scaled scores for this disorder, as indicated in the U.S. Census Bureau's privacy policy. Sources: CBDRB-FY19-422, CBDRB-FY19-448 and CBDRB-FY19-536.

Table 2d.

Probability (Prevalence Percentiles) for group meeting criteria for a Personality Disorder ("Yes" in scale score) based on weighted means

Personality Disorder	**Native American Male N1=400 N2=300**	**Native American Female N1=450 N2=400**	**African American Male N1=3000 N2=2300**	**African American Female N1=5100 N2=4200**	**Hispanic Male N1=3700 N2=2700**
Antisocial	0.1350	0.0506	0.0486	0.0279	0.0474
Avoidant	N/A	0.0418	0.0164	0.0225	0.0179
Dependent	N/A	N/A	N/A	0.0042	N/A
Obsessive-Compulsive	0.1090	0.0875	0.0731	0.0849	0.0559
Paranoid	0.1100	0.0957	0.0625	0.0869	0.0404
Schizoid	0.0775	0.0455	0.0430	0.0536	0.0361
Histrionic	0.0250	0.0265	0.0256	0.0260	0.0178
Schizotypal	0.0764	0.0594	0.0109	0.0660	0.0394
Borderline	0.1320	0.1050	0.0807	0.0809	0.0525
Narcissistic	0.1030	0.0516	0.1330	0.1180	0.0827
Personality Disorder	**Hispanic Female N1=4600 N2=3600**	**Caucasian Male N1=11000 N2=8900**	**Caucasian Female N1=13500 N2=11500**	**Asian American Male N1=550 N2=400**	**Asian American Female N1=700 N2=500**
Antisocial	0.0183	0.0563	0.0175	0.0232	N/A
Avoidant	0.0218	0.0191	0.0293	N/A	0.0209
Dependent	0.0050	0.0039	0.0065	N/A	N/A
Obsessive-Compulsive	0.0641	0.0840	0.0821	0.0517	0.0409
Paranoid	0.0636	0.0329	0.0403	0.0254	0.0415
Schizoid	0.0363	0.0290	0.0267	N/A	N/A
Histrionic	0.0143	0.0181	0.0171	N/A	N/A
Schizotypal	0.0385	0.0398	0.0310	N/A	0.0258
Borderline	0.0533	0.0514	0.0602	0.0420	0.0291
Narcissistic	0.0678	0.0684	0.0331	0.0593	0.0448

Note: N1 indicates number of participants for NESARC wave 1; N2 indicates number of participants for NESARC wave 2; standard deviations are presented in (). All Ns are rounded, as provided by Census. N/A indicates that data was not available as less than 15 participants obtained "yes" scores for scaled scores for this disorder, as indicated in the U.S. Census Bureau's privacy policy. Sources: CBDRB-FY19-422, CBDRB-FY19-448 and CBDRB-FY19-536.

while Asian American participants tended to have less probability of having a personality disorder than the average across all participants. In fact, Asian American participants had almost half the probability (up to 3% less) as the total population for antisocial, obsessive-compulsive, schizoid, schizotypal and borderline personality disorders, and a decrease of approximately 1% for narcissistic and paranoid personality disorders. Caucasians also had slightly less than the overall probability for paranoid and narcissistic personality disorders. Hispanic participants had lower probability for obsessive-compulsive PD, but higher probability for paranoid and narcissistic personality disorders. African American participants were approximately twice as likely to have paranoid, schizotypal or narcissistic personality disorders than the total population, and had a prevalence increase of at approximately 1% for borderline, histrionic and schizoid personality disorders. A 12.4% prevalence in narcissistic PD was observed, as well as a 6.78% in schizotypal and 7.62% in paranoid PDs. Native American participants had increased prevalence for almost all disorders examined compared to the total population prevalence. Over 1% increases in histrionic and narcissistic and over 2% increase in prevalence of obsessive-compulsive personality disorders were observed. Similar to African American participants, Native American participants were twice as likely to have a schizotypal (6.72%) or paranoid (10.2%) diagnosis. They were also almost twice as likely to have a borderline (11.7%) and schizoid (6.07%) diagnosis, and almost three times as likely to have an antisocial diagnosis (9.09%). Native American males had the highest probability of a diagnosis of all personality disorders except dependent (as no data was available) and histrionic (of which the highest was Native American females), with antisocial (13.5%) and borderline (13.2%) being the highest probability.

Mean differences were calculated and significance was assessed using a t-test for gender. A Wald F-test was used for calculating significant mean differences for ethnicity, as this allowed for the use of sample weights. These are presented in Table 3.

Dimensionality Tests

The results of the unidimensionality tests were relatively consistent across almost all scales and groups, indicating that more than one parameter was present, indicating that the unidimensionality assumption has not been met. Modified parallel analysis using a Monte Carlo simulation indicated multidimensionality for gender on all scales except for antisocial. Antisocial also indicated some unidimensionality for ethnicity, though two groups (Native American and Asian American) were not available due to sample size. Unidimensionality for these groups was noted across half of the personality scales. Results of this test can be seen in Table 4. As indicated previously, however, this test is subject to inconsistency with large sample sizes because of its reliance on significance testing. This may lead to results being significant at $p<0.05$ and $p<0.01$ even with very small differences. As such, additional measures were used.

Another significance test, the maximum likelihood ratio test, also indicated significant p-values ($p<0.001$) on all scales for all participants except the scales for which there were less than 15 participants in the group, which were not able to be calculated. As expected, significance testing indicated multidimensionality because of the large sample size (high N value), leading to small differences being identified as significant. These results may be seen in Table 5. These results indicate that the maximum likelihood ratio for a two-factor model provided significantly better fit than a one-factor model.

The comparison of first and second eigenvalues can be seen in Tables 6. These results indicated that all factors except for Native Americans antisocial scale were predominantly

Table 3.

Mean Differences. Significance estimates used t-tests (gender) and Wald F-tests (ethnicity)

Personality Disorder	**Male vs. Female**	**Native American (1) vs. Caucasian (4)**	**African American (3) vs. Caucasian (4)**	**Hispanic (5) vs. Caucasian (4)**	**Asian American (2) vs Caucasian (4)**
Antisocial	15.0***	19.16***	0.06	0.88	35.65***
Avoidant	4.69***	2.57	3.93*	3.71	0.11
Dependent	2.78**	N/A	2.55	1.59	N/A
Obsessive-Compulsive	0.071	1.64	0.55	31.73***	30.43***
Paranoid	4.66***	29.73***	87.99***	18.35***	0.25
Schizoid	-0.532	10.59**	40.22***	7.12**	8.65**
Histrionic	-0.541	2.35	10.33**	0.50	0.00
Schizotypal	-2.37*	9.69**	55.88***	1.13	10.75**
Borderline	1.97*	19.28***	26.74***	0.59	9.04**
Narcissistic	-9.31 ***	4.96*	174.2***	27.87***	0.05

Personality Disorder	**Native American (1) vs. African American (3)**	**Native American (1) vs. Asian American (2)**	**Native American (1) vs. Hispanic (5)**	**Hispanic (5) vs. African American (3)**	**Hispanic (5) vs. Asian American (2)**	**African American (3) vs. Asian American (2)**
Antisocial	17.91***	36.67***	20.51***	0.89	19.30***	27.57***
Avoidant	4.90*	2.30	4.87*	0.00	0.29	0.29
Dependent	0.46	N/A	N/A	0.12	N/A	N/A
Obsessive-Compulsive	2.27	15.91***	10.26***	14.28***	3.57	20.38***
Paranoid	4.31*	26.51***	16.68***	22.92***	7.24**	35.67***
Schizoid	1.24	17.48***	5.50*	9.09**	17.57***	43.82***
Histrionic	0.00	1.53	3.02	10.42**	0.10	3.06
Schizotypal	0.00	18.71***	7.00**	30.60***	12.33***	59.50***
Borderline	6.34*	28.48***	20.28***	23.90***	5.41*	32.69***
Narcissistic	15.43***	2.70	0.00	49.47***	5.67*	49.77***

Note: "*" significant at $p<0.05$, "**" $p<0.01$, and "***" $p<0.001$. Sources: CBDRB-FY19-422, CBDRB-FY19-448 and CBDRB-FY19-536.

Table 4.

Unidimensionality test: modified parallel analysis.

All Participants and Gender			
Personality Disorder	**All Participants**	**Males**	**Females**
Antisocial	1.572 ** (0.6178)	1.608 * (0.8531)	1.564 (0.9572)
Avoidant	0.4479 ** (0.2394)	0.4481 ** (0.2495)	0.4483 ** (0.2437)
Dependent	0.5986 ** (0.2665)	0.6195 ** (0.2972)	0.5915 ** (0.2659)
Obsessive-Compulsive	0.7504 ** (0.2838)	0.7135 ** (0.2881)	0.7792 ** (0.2998)
Paranoid	0.7239 ** (0.3627)	0.7066 ** (0.3578)	0.7392 ** (0.3716)
Schizoid	0.5984 ** (0.3452)	0.6245 ** (0.3546)	0.5923 ** (0.3406)
Histrionic	1.129 ** (0.3884)	1.129 ** (0.3930)	1.130 ** (0.4034)
Schizotypal	1.692 ** (0.5099)	1.692 ** (0.5267)	1.696 ** (0.5197)
Borderline	1.176 ** (0.3336)	1.254 ** (0.3584)	1.126 ** (0.361)
Narcissistic	0.9853 ** (0.2725)	0.9866 ** (0.2966)	0.9878 ** (0.2976)

Ethnicity					
Personality Disorder	**Native American**	**African American**	**Hispanic**	**Caucasian**	**Asian American**
Antisocial	N/A	1.787 (1.353)	1.985 (1.903)	1.566 (0.849)	N/A
Avoidant	0.7157 * (0.4791)	0.4253 ** (0.2630)	0.4246 ** (0.2487)	0.4632 ** (0.247)	0.5031 (0.3996)
Dependent	N/A	0.6283 ** (0.3049)	0.5264 ** (0.3079)	0.6019 ** (0.2819)	N/A
Obsessive-Compulsive	0.782 ** (0.532)	0.7869 ** (0.3344)	0.7746 ** (0.3336)	0.7471 ** (0.2876)	0.8642 ** (0.4995)
Paranoid	0.7277 ** (0.4987)	0.7717 ** (0.3687)	0.7119 ** (0.3833)	0.7255 ** (0.3659)	0.6345 * (0.4835)
Schizoid	0.6904 (0.548)	0.5635 ** (0.3598)	0.6216 ** (0.3733)	0.6154 ** (0.344)	0.6472 (0.5463)
Histrionic	1.155 (0.9997)	1.215 ** (0.401)	1.166 ** (0.4259)	1.103 ** (0.4144)	1.129 (0.8561)
Schizotypal	1.873 (1.417)	1.752 ** (0.5741)	1.748 ** (0.5403)	1.671 ** (0.5164)	1.689 (1.057)
Borderline	1.715 (1.494)	1.144 ** (0.4427)	1.220 ** (0.4477)	1.186 ** (0.3543)	1.366 * (0.9208)
Narcissistic	1.196 ** (0.8171)	0.9496 ** (0.3712)	1.184 ** (0.3639)	0.9527 ** (0.2854)	1.264 ** (0.7559)

Note. Montecarlo (100) random data parallel analysis tests in (); ** indicates p<0.01; "***" indicates p<0.001. Sources: CBDRB-FY19-422, CBDRB-FY19-448 and CBDRB-FY19-536.

Table 5.

Maximum likelihood ratio test indicates that the Likelihood Ratio for a two-factor model provided significantly better fit than a one-factor model.

All Participants and Gender			
Personality Disorder	**All Participants**	**Males**	**Females**
Antisocial	4689 ***	2053 ***	2650 ***
Avoidant	600.3 ***	262.6 ***	345.4 ***
Dependent	414.9 ***	182.3 ***	241.9 ***
Obsessive-Compulsive	1538 ***	585.2 ***	961.5 ***
Paranoid	2597 ***	1145 ***	1461 ***
Schizoid	1108 ***	526.8 ***	592.8 ***
Histrionic	2758 ***	1201 ***	15650 ***
Schizotypal	5284 ***	2364 ***	2931 ***
Borderline	2640 ***	1230 ***	1424 ***
Narcissistic	2586 ***	1093 ***	1521 ***

Ethnicity					
Personality Disorder	**Native American**	**African American**	**Hispanic**	**Caucasian**	**Asian American**
Antisocial	N/A	1025 ***	852.5 ***	2702 ***	N/A
Avoidant	26.34 ***	83.86 ***	98.56 ***	408.5 ***	26.66 ***
Dependent	N/A	92.34 ***	57.46 ***	246.0 ***	N/A
Obsessive-Compulsive	30.71 ***	344.7 ***	286.4 ***	892.5 ***	54.91 ***
Paranoid	54.02 ***	565.4 ***	490.4 ***	1492 ***	60.14 ***
Schizoid	38.93 ***	187.5 ***	196.9 ***	729.1 ***	24.99 **
Histrionic	70.06 ***	610.0 ***	485.5 ***	1603 ***	78.68 ***
Schizotypal	121.9 ***	1081 ***	892.9 ***	3187 ***	131.1 ***
Borderline	91.36 ***	478.3 ***	414.5 ***	1714 ***	77.20 ***
Narcissistic	69.48 ***	456.5 ***	588.2 ***	1500 ***	98.49 ***

Note: "*" significant at $p<0.05$, "**" $p<0.01$, and "***" $p<0.001$. Sources: CBDRB-FY19-422, CBDRB-FY19-448 and CBDRB-FY19-536.

Table 6.

Unidimensionality tests: Ratio of first to second eigenvalue

All Participants and Gender					
Personality Disorder	**All Participants**		**Males**		**Females**
Antisocial	9.186		8.969		9.220
Avoidant	9.511		9.419		9.568
Dependent	7.390		7.061		7.542
Obsessive-Compulsive	5.207		5.471		5.019
Paranoid	6.795		6.951		6.663
Schizoid	6.871		6.626		6.929
Histrionic	3.771		3.757		3.775
Schizotypal	4.217		4.208		4.212
Borderline	7.950		7.444		8.303
Narcissistic	6.919		6.829		6.960
Ethnicity					
Personality Disorder	**Native American**	**Hispanic**	**African American**	**Caucasian**	**Asian American**
Antisocial	2.771*	7.178	8.153	9.107	4.292
Avoidant	5.818	10.22	9.941	9.140	8.826
Dependent	N/A	8.488	7.039	7.338	N/A
Obsessive-Compulsive	5.443	5.236	4.990	5.156	4.449
Paranoid	7.336	6.927	6.355	6.750	8.115
Schizoid	5.677	6.784	7.327	6.650	6.158
Histrionic	3.504	3.707	3.465	3.855	3.754
Schizotypal	4.038	4.056	4.014	4.284	4.362
Borderline	4.764	7.271	8.268	7.945	7.061
Narcissistic	5.156	5.619	7.208	7.209	5.123

Note. "*" indicates that the ratio of first to second eigenvalues was less than 3.0. Sources: CBDRB-FY19-422, CBDRB-FY19-448 and CBDRB-FY19-536.

explained by one factor, as they all had ratios greater than 3.0. The scale in exception had a ratio of 2.771, which was close to the cutoff. It is possible that this scale contained more than one underlying factor structure. It is possible that the antisocial scale contains a multidimensional structure for Native Americans, and its IRT analyses should be examined with caution, although these analyses can be robust to violations of the unidimensionality assumption.

Despite the suggestion of multidimensionality, because the presence of a dominant first dimension, it suggests that IRT analyses can still be performed adequately. The IRT analyses are robust to violations of the unidimensionality assumption so long as one dominant latent trait or dimension is present.

Differential Item Functioning and Differential Test Functioning

2PLM IRT analyses were then run on the scales for all participants, then on each group. The item parameters for each comparison were then equated as part of the DIF and DTF analyses.

ANOVA results indicated significant results for all groups except dependent for ethnicity, and all except obsessive-compulsive, schizoid and histrionic for gender (Table 7). Partial Eta Squared were all trivial. Goodness of fit for logistic regression Pseudo R2 was also calculated for gender and ethnicity compared to a constant only (item only) model. Results were also trivial. However, it is noted that both of these measures do not separate the ethnicities out, and do not account for bias that may be inherent in the test that may influence these comparisons.

Cluster A Personality Disorders were examined first. For the Paranoid Personality Disorder scale, a significant ANOVA was found for both gender and ethnicity (see Table 7). Pairwise comparisons (Table 8) indicated that mean scale differences between all ethnicities

Table 7.

ANOVA results and goodness of fit for all personality disorder scales

Gender			
Personality Disorder	**F**	**p-level**	**R2 Partial Eta Squared**
Antisocial	402.0***	<0.001	0.0093
Avoidant	34.33***	<0.001	0.0008
Dependent	12.64***	0.0004	0.0003
Obsessive-Compulsive	0.01	0.9295	0.0000
Paranoid	33.34***	<0.001	0.0008
Schizoid	0.43	0.5118	0.0000
Histrionic	0.45	0.5040	0.0000
Schizotypal	8.16**	0.0043	0.0002
Borderline	5.49*	0.0191	0.0002
Narcissistic	127.5***	<0.001	0.0037
Ethnicity			
Personality Disorder	**F**	**p-level**	**R2 Partial Eta Squared**
Antisocial	14.58***	<0.001	0.0027
Avoidant	2.42*	0.0463	0.0003
Dependent	0.83	0.5046	0.0001
Obsessive-Compulsive	14.72***	<0.001	0.0014
Paranoid	31.06***	<0.001	0.0057
Schizoid	16.81***	<0.001	0.0025
Histrionic	3.51**	0.0071	0.0005
Schizotypal	20.16***	<0.001	0.0036
Borderline	14.93***	<0.001	0.0030
Narcissistic	47.79***	<0.001	0.0095

Goodness of fit for logistic regression Pseudo R2					
Personality Disorder	**Pseudo-R2 (Constant)**	**Pseudo-R2 (Gender)**	**Gender-Constant**	**Pseudo-R2 (Ethnicity)**	**Ethnicity-Constant**
Antisocial	0.5584	0.5617	0.0033	0.5590	0.0006
Avoidant	0.7442	0.7442	0.0000	0.7446	0.0004
Dependent	0.8760	0.8761	0.0001	0.8781	0.0021
Obsessive-Compulsive	0.5949	0.5951	0.0002	0.5951	0.0002
Paranoid	0.7392	0.7401	0.0009	0.7412	0.0020
Schizoid	0.3847	0.3850	0.0003	0.3868	0.0021
Histrionic	0.6857	0.6857	0.0000	0.6869	0.0012
Schizotypal	0.6008	0.6010	0.0002	0.6017	0.0009
Borderline	0.7983	0.7984	0.0001	0.7991	0.0008
Narcissistic	0.5611	0.5617	0.0005	0.5622	0.0011

Note: Goodness of fit for logistic regressions of PD Dx (0/1) using Wald test, on (a) its items constant only model, (b) items + female, and (c) items plus 4 race indicators (vs. Caucasian). "*" significant at p<0.05, "**" p<0.01, and "***" p<0.001. Sources: CBDRB-FY19-422, CBDRB-FY19-448 and CBDRB-FY19-536.

Table 8.

Pairwise comparisons for Ethnicity for Cluster A Personality Disorders.

Ethnicity Pairwise Comparisons: Paranoid			
(I) vs. (J)	**Mean Difference**	**Standard Error**	**Significance (p<.05)**
Native American/African American	0.0261***	0.0075	0.0005
Native American/Hispanic	0.0506***	0.0066	0.0000
Native American/Caucasian	0.0656***	0.0069	0.0000
Native American/Asian American	0.0687***	0.0106	0.0000
African American/Hispanic	0.0244***	0.0038	0.0000
African American/Caucasian	0.0395***	0.0033	0.0000
African American/Asian American	0.0425***	0.0057	0.0000
Hispanic/Asian American	0.0181***	0.0050	0.0003
Hispanic/Caucasian	0.0151***	0.0031	0.0000
Asian American/Caucasian	0.0031	0.0052	0.557
Ethnicity Pairwise Comparisons: Schizoid			
(I) vs. (J)	**Mean Difference**	**Standard Error**	**Significance (p<.05)**
Native American/African American	0.0117	0.0061	0.055
Native American/Hispanic	0.0245***	0.0055	0.0000
Native American/Caucasian	0.0329***	0.0060	0.0000
Native American/Asian American	0.0451***	0.0080	0.0000
African American/Hispanic	0.0128***	0.0032	0.0000
African American/Caucasian	0.0212***	0.0028	0.0000
African American/Asian American	0.0334***	0.0046	0.0000
Hispanic/Asian American	0.0206***	0.0041	0.0000
Hispanic/Caucasian	0.0084**	0.0027	0.0017
Asian American/Caucasian	0.0122**	0.0045	0.0066
Ethnicity Pairwise Comparisons: Schizotypal			
(I) vs. (J)	**Mean Difference**	**Standard Error**	**Significance (p<.05)**
Native American/African American	0.0003	0.0077	0.968
Native American/Hispanic	0.0302***	0.0063	0.0000
Native American/Caucasian	0.0333***	0.0074	0.0000
Native American/Asian American	0.0504***	0.0099	0.0000
African American/Hispanic	0.0299***	0.0040	0.0000
African American/Caucasian	0.0330***	0.0036	0.0000
African American/Asian American	0.0501***	0.0062	0.0000
Hispanic/Asian American	0.0202***	0.0049	0.0000
Hispanic/Caucasian	0.0031	0.0035	0.371
Asian American/Caucasian	0.0171**	0.0060	0.0045

Note: "*" significant at p<0.05, "**" p<0.01, and "***" p<0.001. Sources: CBDRB-FY19-422, CBDRB-FY19-448 and CBDRB-FY19-536.

except for Asian American compared to Caucasian participants were significant. At the scale level, females had more frequent response rates than men, and Native Americans, African Americans and Hispanics were more likely to respond positively to the scale compared to Caucasians. Native Americans were more likely than other minorities to respond yes, and Hispanics were more likely to respond yes than African American or Asian American participants. There was no notable difference between Asian Americans and Caucasians, a trend which was found in many of the scales to follow. Non-compensatory differential item functioning (NC-DIF) was examined next, to determine if these results were due to true differences between the populations, or if bias had been introduced into the items. Three items within the paranoid scale had significant chi-squared for gender (see Table 9a). Like the significance tests for the unidimensionality tests previously mentioned, this significance test is also prone to giving significant results when only minor differences exist in the population, and should be interpreted with caution when large samples like the NESARC data is used. As such, examination of odds ratios and test level functioning is warranted. Native Americans had a similar pattern of items containing bias when compared with Caucasians, with four items containing bias (see Table 9b for ethnicity analyses). When compared with Caucasians, Hispanic participants had significant differences in item functioning in six out of the nine items. African Americans compared with Caucasians had bias present in seven of the nine items in the paranoid scale. There was no differential functioning for items for comparisons between Caucasians and Asian Americans. Common items between groups which contained bias for many of the comparisons was item S10Q1A30, which stated “The kind of person who takes a long time to forgive people who have insulted/slighted you,” and item S10Q1A31, “Many people you can't forgive because they said/did something long ago.” This indicates that the items with differential

Table 9a.

Paranoid Personality Disorder responses and differential functioning of items and tests: Logistic Regression, Mantel-Haenszel (MH) Chi2 and Odds Ratios for gender

Scale Item	**Male vs. Female DIF**			
	CHI2	**Sig.**		**Odds Ratio**
S10Q1A26: Often have to keep an eye out to keep people from using/hurting/lying to you	0.24	0.6276		1.135*
S10Q1A27: Spend a lot of time wondering if can trust friends/people you work with	2.89	0.0893		0.811*
S10Q1A28: Find it is best not to let others know much about you because they will use it against you	8.45**	0.0037		0.863*
S10Q1A29: Detect hidden threats or insults in things people say or do	0.16	0.6865		0.834*
S10Q1A30: The kind of person who takes a long time to forgive people who have insulted/slighted you	37.12***	0.0000		0.784*
S10Q1A31: Many people you can't forgive because they said/did something long ago	5.23*	0.0222		0.844*
S10Q1A32: Often get angry or lash out when someone criticizes or insults you	0.07	0.7864		0.956
S10Q1A33: Often suspected that your spouse or partner has been unfaithful	0.99	0.3200		2.127*
S10Q1A35: When around people, often feel that you are being watched or stared at	1.04	0.3078		1.120*
	Male vs. Female DTF			
Scale Score	**CHI2**	**Sig.**	**MH CHI2**	**Odds Ratio**
PARADX2: Paranoid Personality Disorder	0.42	0.5157	5.89*	1.303*

Note: "*" significant at p<0.05, "**" p<0.01, and "***" p<0.001. Sources: CBDRB-FY19-422, CBDRB-FY19-448 and CBDRB-FY19-536.

Table 9b.

Paranoid Personality Disorder differential functioning of items and tests: Logistic Regression, Mantel-Haenszel (MH) Chi2 and Odds Ratios for ethnicity

Scale Item	**Native American Vs. Caucasian DIF**			**Hispanic vs. Caucasian DIF**		
	CHI2	**Sig.**	**Odds Ratio**	**CHI2**	**Sig.**	**Odds Ratio**
S10Q1A26	1.66	0.1970	1.539*	22.52***	0.0000	0.728*
S10Q1A27	0.03	0.8690	0.948	12.03***	0.0005	1.249*
S10Q1A28	4.30*	0.0380	1.410*	2.21	0.1368	1.408*
S10Q1A29	0.32	0.5713	1.144	17.93***	0.0000	1.055
S10Q1A30	28.83***	0.0000	0.537*	24.48***	0.0000	0.771*
S10Q1A31	13.63***	0.0002	0.739*	5.87*	0.0154	0.878*
S10Q1A32	2.05	0.1524	0.798	0.03	0.8542	1.266*
S10Q1A33	6.73**	0.0095	1.221	3.97*	0.0464	1.001
S10Q1A35	0.14	0.7053	1.206	0.54	0.4625	1.224*
	Native American Vs. Caucasian DTF			**Hispanic vs. Caucasian DTF**		
Scale Score	**CHI2**	**Sig.**	**MH CHI2**	**Odds Ratio**	**CHI2**	**Sig.**	**MH CHI2**	**Odds Ratio**
PARADX2	0.00	0.9689	2.54	1.792	4.48*	0.0343	2.60	0.789

Scale Item	**African American vs. Caucasian DIF**			**Asian American vs. Caucasian DIF**		
	CHI2	**Sig.**	**Odds Ratio**	**CHI2**	**Sig.**	**Odds Ratio**
S10Q1A26	0.03	0.8624	1.385*	0.01	0.9070	0.984
S10Q1A27	12.58***	0.0004	1.136*	0.41	0.5206	1.593*
S10Q1A28	0.06	0.8066	2.858*	0.01	0.9115	1.105
S10Q1A29	28.34***	0.0000	1.232*	1.50	0.2214	0.954
S10Q1A30	142.8***	0.0000	0.455*	0.86	0.3543	0.826
S10Q1A31	32.61***	0.0000	0.495*	0.13	0.7159	1.210
S10Q1A32	7.64**	0.0057	0.563*	1.42	0.2333	1.215
S10Q1A33	4.17*	0.0411	1.290*	0.56	0.4558	0.572*
S10Q1A35	7.15**	0.0075	1.467*	0.35	0.5549	0.962
	African American vs. Caucasian DTF			**Asian American vs. Caucasian DTF**		
Scale Score	**CHI2**	**Sig.**	**MH CHI2**	**Odds Ratio**	**CHI2**	**Sig.**	**MH CHI2**	**Odds Ratio**
PARADX2	22.79***	0.0000	19.2***	0.582*	0.58	0.4453	0.58	0.736

Note: Odds ratios are in the form of Minority:Caucasian. This indicates that for every unit increase in minority, Caucasian would go up by the indicated ratio number. "*" significant at $p<0.05$, "**" $p<0.01$, and "***" $p<0.001$. Sources: CBDRB-FY19-422, CBDRB-FY19-448 and CBDRB-FY19-536.

functioning may be contributing somewhat to the observed mean differences. This indicates that these scale items are introducing bias into the diagnostic criteria for this scale. Differential Test Functioning indicated that these differences were nonsignificant at the overall paranoid scale level for gender and for the comparisons to Caucasians for Native Americans. Hispanic and African Americans obtained significant chi-squared results, indicating the presence of differential functioning, or bias, for these two minorities at the scale level. It should be noted again that in these situations where some DIF results in significant DTF while others do not, that the individual items may all be biased in the same direction, or may be in opposing directions. This would lead to amplification of the DIF to the test level when most items are biased or the items with more influence on the scale are biased in the same direction. In the case of it DIF not leading to DTF, the items may have bias in opposing directions (e.g., bias towards men for one item and women for another), leading to the bias effectively cancelling each other out. The odds ratios beside each significance check of DIF helps to identify exactly how each item may be canceling or amplifying these results by giving a direction to each DIF result. Mantel-Haenszel (MH) differential item functioning was also completed to obtain odds ratios and to compare overall scale functioning. Odds ratios (OR) were examined to determine the impact or effect that this would have upon the total score of participants. Note that ORs presented are in the format of minority:Caucasian, and as such, the numbers presented indicate that for every 1 minority person, the Caucasian person was x times as likely to have a response of yes. For gender analysis, the ORs are presented in the format of male:female, and as such, the numbers presented indicate that for every 1 male participant, the female participant was x times as likely to have a response of yes. Reversed odds ratios are available in Table 21 at the end of the analyses. The odds ratio was significant for African Americans, with an odds ratio of 0.582, with African

Americans answering yes more frequently than Caucasians. Interestingly, the MH analysis also indicated a significant scale level difference in functioning for gender, which had a significant odds ratio of 1.303 (with women answering yes more frequently than men). Overall, these results indicated differential test functioning, or scale level bias, with a notable impact for African Americans, with this minority tending to be scored higher than Caucasians.

For the Schizoid Personality Disorder scale, a significant ANOVA was found for ethnicity but not for gender (see Table 7). Pairwise comparisons (Table 8) indicated that mean scale differences between all ethnicities except for Native American compared to African American participants were significant. At the scale level, females had similar response rates than men. Native Americans, African Americans and Hispanics were more likely to respond positively to the scale compared to Caucasians. Asian Americans were less likely than Caucasians to respond positively. Native Americans were more likely than other minorities to respond yes, and African Americans were more likely to respond yes than Hispanics or Asian American participants. There was no notable difference between African Americans and Native Americans. Differential item functioning was examined next, to determine if these results were due to true differences between the populations, or if bias had been introduced into the items. Five items within the schizoid scale had significant chi-squared for gender, despite not having a significant difference in overall response rate mean differences (see Table 10a). Native Americans had only two items contributing bias to the significant difference in means when compared with Caucasians: items S10Q1A49 “Rarely show much emotion,” and S10Q1A51 “Rarely react to praise or criticism” (see Table 10b). These items were common to most of the group comparisons. When compared with Caucasians, Hispanic participants had significant differences in item functioning in six out

Table 10a.

Schizoid Personality Disorder responses and differential functioning of items and tests: Logistic Regression, Mantel-Haenszel (MH) Chi2 and Odds Ratios for gender

Scale Item	**Male vs. Female DIF**			
	CHI2	**Sig.**		**Odds Ratio**
S10Q1A43: Are there very few people you're really close to outside of immediate family	0.01	0.9405		0.970
S10Q1A45: Would be just happy without having any close relationship	18.57***	0.0000		1.091*
S10Q1A46: Take little pleasure in being with others	9.41**	0.0022		0.916*
S10Q1A47: Have almost always preferred to do things alone rather than with others	2.95	0.0856		0.812*
S10Q1A48: Could be content without ever being sexually involved with anyone	180.8***	0.0000		3.297*
S10Q1A49: Rarely show much emotion	9.86**	0.0017		0.491*
S10Q1A50: Very few things that give you pleasure	5.37*	0.0205		0.846*
S10Q1A51: Rarely react to praise or criticism	0.83	0.3611		0.729*
S10Q1A52: The sort of person who doesn't care about what people think of you	3.65	0.0562		0.779*
S10Q1A53: Find nothing makes you very happy or sad	0.14	0.7069		0.728*
	Male vs. Female DTF			
Scale Score	**CHI2**	**Sig.**	**MH CHI2**	**Odds Ratio**
SCHIZDX2: Schizoid Personality Disorder	1.26	0.2614	0.02	0.989

Note: "*" significant probability at p<0.05, "**" p<0.01, and "***" p<0.001. Sources: CBDRB-FY19-422, CBDRB-FY19-448 and CBDRB-FY19-536.

Table 10b.

Schizoid Personality Disorder differential functioning of items and tests: Logistic Regression, Mantel-Haenszel (MH) Chi2 and Odds Ratios for ethnicity

Scale Item	Native American Vs. Caucasian DIF			Hispanic vs. Caucasian DIF		
	CHI2	Sig.	Odds Ratio	CHI2	Sig.	Odds Ratio
S10Q1A43	0.53	0.4678	1.054	12.82***	0.0003	0.744*
S10Q1A45	0.29	0.5917	1.021	1.68	0.1946	1.764*
S10Q1A46	0.91	0.3392	1.178	5.93*	0.0149	1.265*
S10Q1A47	0.30	0.5825	0.857	5.88*	0.0153	1.550*
S10Q1A48	2.02	0.1552	1.246	0.00	0.9505	0.764*
S10Q1A49	6.10*	0.0135	1.075	4.25*	0.0394	0.806*
S10Q1A50	0.01	0.9369	0.621*	3.18	0.0745	1.488*
S10Q1A51	4.32*	0.0377	1.206	6.15*	0.0131	0.775*
S10Q1A52	2.52	0.1122	0.873	0.20	0.6508	1.006
S10Q1A53	0.01	0.9433	1.158	6.25*	0.0124	0.964
	Native American Vs. Caucasian DTF			Hispanic vs. Caucasian DTF		
Scale Score	CHI2	Sig.	MH CHI2	Odds Ratio	CHI2	Sig.	MH CHI2	Odds Ratio
SCHIZDX2	0.02	0.8789	0.02	1.054	2.23	0.1353	2.11	0.872

Scale Item	African American vs. Caucasian DIF			Asian American vs. Caucasian DIF		
	CHI2	Sig.	Odds Ratio	CHI2	Sig.	Odds Ratio
S10Q1A43	39.84***	0.0000	0.797*	0.20	0.6583	0.848
S10Q1A45	6.87**	0.0087	1.833*	1.85	0.1732	1.437*
S10Q1A46	2.61	0.1064	1.247*	3.50	0.0614	1.327*
S10Q1A47	1.78	0.1827	1.326*	0.12	0.7250	1.444*
S10Q1A48	24.21***	0.0000	1.092*	0.00	0.9977	0.645*
S10Q1A49	43.15***	0.0000	0.717*	1.57	0.2103	1.050
S10Q1A50	0.03	0.8543	1.335*	0.08	0.7747	1.647*
S10Q1A51	57.55***	0.0000	0.652*	0.00	0.9802	0.991
S10Q1A52	18.57***	0.0000	0.991	0.01	0.9239	0.840*
S10Q1A53	32.37***	0.0000	0.806*	0.80	0.3711	1.313
	African American vs. Caucasian DTF			Asian American vs. Caucasian DTF		
Scale Score	CHI2	Sig.	MH CHI2	Odds Ratio	CHI2	Sig.	MH CHI2	Odds Ratio
SCHIZDX2	19.44***	0.0000	36.2***	0.605*	0.93	0.3355	6.60*	0.490*

Note: "*" significant probability at p<0.05, "**" p<0.01, and "***" p<0.001. Sources: CBDRB-FY19-422, CBDRB-FY19-448 and CBDRB-FY19-536.

of the ten items. African Americans compared with Caucasians had bias present in seven of the ten items in the schizoid scale. There was no differential functioning for items for comparisons between Caucasians and Asian Americans. This was consistent with mean differences and ANOVA pairwise comparisons. This indicates that the items with differential functioning may be contributing somewhat to the observed mean differences. This indicates that these scale items are introducing bias into the diagnostic criteria for this scale. To determine if these items were contributing to the overall score mean differences, test level functioning was examined next. Differential Test Functioning indicated that these differences were only significant at the overall schizoid scale level for African Americans when compared to Caucasians obtaining a significant chi-squared result. This indicates the presence of differential functioning, or bias, for this minority at the scale level. MH differential item functioning was also completed to obtain odds ratios and to compare overall scale functioning. ORs were examined to determine the impact or effect that this would have upon the total score of participants. The OR was significant for African Americans, with an OR of 0.605, with Caucasians answering yes less frequently than African Americans. Interestingly, the MH analysis also indicated a significant scale level difference in functioning for Asian Americans compared to Caucasians, which had a significant OR of 0.490. Overall, these results indicated differential test functioning, with a notable impact for African Americans, with this minority tending to be scored higher than Caucasians.

For the Schizotypal Personality Disorder scale, a significant ANOVA was found for both gender and ethnicity (see Table 7). Pairwise comparisons (Table 8) indicated that mean scale differences between most ethnicities were significant, except for Native American compared to African American and Hispanic compared to Caucasian. At the scale level, males were more likely to respond yes than females. Native Americans were more likely to respond positively than

Caucasians, Asian Americans or Hispanics. African Americans responded positively more often than Asian Americans, Caucasians, and Hispanics. Caucasians responded positively more frequently than Asian Americans. Differential item functioning was examined next, to determine if these results were due to true differences between the populations, or if bias had been introduced into the items. The schizotypal scale contained sixteen items. Five items within the schizoid scale had significant chi-squared for gender (see Table 11a). Native Americans had ten items contributing bias to the significant difference in means when compared with Caucasians (see Table 11b). When compared with Caucasians, Hispanic participants had significant differences in item functioning in seven out of the sixteen items. African Americans compared with Caucasians had bias present in thirteen of the sixteen items in the schizotypal scale. There was no differential functioning for items for comparisons between Caucasians and Asian Americans. A common item between groups which contained bias for many of the comparisons was item W2S10Q1A49, “Have you had trouble expressing your emotions and feelings.” This indicates that the items with differential functioning may be contributing somewhat to the observed mean differences. This indicates that these scale items are introducing bias into the diagnostic criteria for this scale. To determine if these items were contributing to the overall score mean differences, test level functioning was examined next. Differential Test Functioning indicated that these differences were significant at the overall schizotypal scale level for African Americans and Native Americans when compared to Caucasians, obtaining a significant chi-squared result. This indicates the presence of differential functioning, or bias, for this minority at the scale level. MH and ORs were examined to determine the impact or effect that this would have upon the total score of participants. The OR was significant for African Americans, with an odds ratio of 0.607, with Caucasians answering yes less frequently than African Americans. It

Table 11a.

Schizotypal Personality Disorder responses and differential functioning of items and tests: Logistic Regression, Mantel-Haenszel (MH) Chi2 and Odds Ratios for gender

Scale Item	**Male vs. Female DIF**			
	CHI2	**Sig.**		**Odds Ratio**
W2S10Q1A38: Have you often had the feeling that things that have no special meaning to most people are really meant to give you a message	0.96	0.3269		1.310*
W2S10Q1A39: Have you often had the feeling of being watched or stared at, when around people	0.07	0.7976		1.466*
W2S10Q1A40: Have you ever felt that you could make things happen just by making a wish or thinking	2.22	0.1365		1.090
W2S10Q1A41: Have you had personal experiences with the supernatural	12.45***	0.0004		1.473*
W2S10Q1A42: Have you believed that you have a "sixth sense" that allows you to know and predict things that others can't	11.45***	0.0007		1.705*
W2S10Q1A43: Have you had the sense that some force is around you, even though you cannot see anyone	0.99	0.3187		1.357*
W2S10Q1A44: Have you often seen auras or energy fields around people	0.18	0.6695		1.517*
W2S10Q1A45: Have people thought you are odd, eccentric or strange	1.89	0.1692		0.715*
W2S10Q1A46: Have there been very few people that you're really close to outside of your immediate family	7.08**	0.0078		1.109*
W2S10Q1A47: Often felt nervous when with other people even whom you have known for a while	0.68	0.4103		1.573*
W2S10Q1A48: Have you rarely shown emotion	3.68	0.0550		0.447*
W2S10Q1A49: Have you had trouble expressing your emotions and feelings	0.31	0.5754		0.737*
W2S10Q1A50: Have felt suspicious of people, even if you have known them for a while	6.73**	0.0095		1.160*
W2S10Q1A51: Have people thought you have strange ideas	9.65**	0.0019		0.596*
W2S10Q1A52: Have people thought you act strangely	0.05	0.8177		0.646*
W2S10Q1A53: Have you often thought that objects or shadows are really people or animals, or that noises are actually people's voices	0.79	0.3752		1.335*
	Male vs. Female DTF			
Scale Score	**CHI2**	**Sig.**	**MH CHI2**	**Odds Ratio**
SKPDX: Schizotypal Personality Disorder	2.89	0.0890	1.83	1.124

Note: "*" significant at p<0.05, "**" p<0.01, and "***" p<0.001. Sources: CBDRB-FY19-422, CBDRB-FY19-448 and CBDRB-FY19-536.

Table 11b.

Schizotypal Personality Disorder differential functioning of items and tests: Logistic Regression, Mantel-Haenszel (MH) Chi2 and Odds Ratios for ethnicity

Scale Item	Native American Vs. Caucasian DIF			Hispanic vs. Caucasian DIF		
	CHI2	Sig.	Odds Ratio	CHI2	Sig.	Odds Ratio
W2S10Q1A38	0.30	0.5844	0.997	0.52	0.4708	1.711*
W2S10Q1A39	7.07**	0.0078	1.429*	4.93*	0.0263	1.570*
W2S10Q1A40	4.53*	0.0333	1.142	0.04	0.8393	1.447*
W2S10Q1A41	0.01	0.9352	1.291*	1.53	0.2154	0.686*
W2S10Q1A42	0.03	0.8544	1.271	0.33	0.5649	1.314*
W2S10Q1A43	0.00	0.9573	1.165	0.65	0.4215	0.729*
W2S10Q1A44	3.86*	0.0494	1.415	1.20	0.2735	1.380*
W2S10Q1A45	2.16	0.1413	0.881	20.21***	0.0000	0.551*
W2S10Q1A46	15.15***	0.0001	0.915	0.03	0.8589	1.209*
W2S10Q1A47	15.45***	0.0001	0.944	7.78**	0.0053	0.855*
W2S10Q1A48	9.79**	0.0018	0.914	10.48**	0.0012	1.218*
W2S10Q1A49	18.12***	0.0000	0.627*	5.65*	0.0175	0.826*
W2S10Q1A50	13.50***	0.0002	0.929	0.77	0.3817	1.623*
W2S10Q1A51	5.18*	0.0229	0.910	27.95***	0.0000	0.492*
W2S10Q1A52	7.28**	0.0070	0.915	22.23***	0.0000	0.476*
W2S10Q1A53	2.53	0.1117	1.796*	3.75	0.0527	1.659*
	Native American Vs. Caucasian DTF			Hispanic vs. Caucasian DTF		
Scale Score	CHI2	Sig.	MH CHI2	Odds Ratio	CHI2	Sig.	MH CHI2	Odds Ratio
SKPDX	10.63**	0.0011	3.80	0.629*	3.16	0.0754	8.66**	0.704*

Scale Item	African American vs. Caucasian DIF			Asian American vs. Caucasian DIF		
	CHI2	Sig.	Odds Ratio	CHI2	Sig.	Odds Ratio
W2S10Q1A38	1.28	0.2579	1.800*	3.17	0.0750	1.289
W2S10Q1A39	4.43*	0.0352	2.213*	0.72	0.3972	1.075
W2S10Q1A40	13.67***	0.0002	1.292*	0.24	0.6226	2.281*
W2S10Q1A41	31.29***	0.0000	0.648*	1.43	0.2324	0.872
W2S10Q1A42	4.83*	0.0279	1.329*	0.03	0.8735	1.260
W2S10Q1A43	39.71***	0.0000	0.848*	2.93	0.0871	0.771*
W2S10Q1A44	8.33**	0.0039	0.884	0.29	0.5910	1.303
W2S10Q1A45	18.77***	0.0000	0.594*	0.56	0.4541	0.580*
W2S10Q1A46	12.29***	0.0005	1.301*	0.00	0.9963	1.051
W2S10Q1A47	3.43	0.0638	0.880*	0.53	0.4651	0.741
W2S10Q1A48	12.83***	0.0003	0.956	1.37	0.2414	1.261*
W2S10Q1A49	50.40***	0.0000	0.533*	2.24	0.1344	0.801
W2S10Q1A50	0.69	0.4060	1.833*	0.47	0.4946	1.187
W2S10Q1A51	39.90***	0.0000	0.573*	0.42	0.5165	0.909
W2S10Q1A52	13.02***	0.0003	0.777*	0.04	0.8491	0.617*
W2S10Q1A53	4.35*	0.0370	0.995	0.15	0.6938	1.266
	African American vs. Caucasian DTF			Asian American vs. Caucasian DTF		
Scale Score	CHI2	Sig.	MH CHI2	Odds Ratio	CHI2	Sig.	MH CHI2	Odds Ratio
SKPDX	16.92***	0.0000	25.0***	0.607*	0.26	0.6103	2.51	0.572

Note: "*" significant at p<0.05, "**" p<0.01, and "***" p<0.001. Sources: CBDRB-FY19-422, CBDRB-FY19-448 and CBDRB-FY19-536.

was also significant for Native Americans, with an OR of 0.629, but no significant MH chi-squared. The MH analysis also indicated a significant scale level difference in functioning for Hispanics compared to Caucasians, which had a significant odds ratio of 0.704. Overall, these results indicated differential test functioning, or scale level bias, with a notable impact for Native Americans and African Americans, with both of these minorities tending to be scored higher than Caucasians.

Cluster B Personality Disorder scales were examined next. For the Antisocial Personality Disorder scale, a significant ANOVA was found for both gender and ethnicity (see Table 7) Pairwise comparisons (Table 12) indicated that mean scale differences between most ethnicities were significant, except for African American compared to Caucasians and Hispanics, and Hispanics compared to Caucasians. Native Americans responded positively more frequently than all other groups, and Asian Americans responded positively less often than other groups. Men responded positively more often than women. Differential item functioning was examined next, to determine if these results were due to true differences between the populations, or if bias had been introduced into the items. The antisocial scale contained thirty items. Twenty two items within the antisocial scale had significant chi-squared for gender (see Table 13a). Native Americans had eighteen items contributing bias to the significant difference in means when compared with Caucasians (see Table 13b). When compared with Caucasians, Hispanic participants had significant differences in item functioning in eleven out of the thirty items. African Americans compared with Caucasians also had bias present in eleven of the items in the antisocial scale. Differential functioning for items for comparisons between Caucasians and Asian Americans indicated bias within sixteen of the items. A common item between groups which contained bias for many of the comparisons was item S11A1A28, “Ever get into a fight

Table 12.

Pairwise comparisons for Ethnicity for Cluster B Personality Disorders.

Ethnicity Pairwise Comparisons: Antisocial			
Ethnicity (I) vs. (J)	**Mean Difference**	**Standard Error**	**Significance (p<.05)**
Native American/African American	0.0539***	0.0058	0.0000
Native American/Hispanic	0.0578***	0.0056	0.0000
Native American/Caucasian	0.0547***	0.0068	0.0000
Native American/Asian American	0.0785***	0.0089	0.0000
African American/Hispanic	0.0038	0.0029	0.181
African American/Caucasian	0.0008	0.0031	0.794
African American/Asian American	0.0246***	0.0040	0.0000
Hispanic/Asian American	0.0207**	0.0038	0.0000
Hispanic/Caucasian	0.0031	0.0030	0.302
Asian American/Caucasian	0.0238***	0.0051	0.0000
Ethnicity Pairwise Comparisons: Borderline			
(I) vs. (J)	**Mean Difference**	**Standard Error**	**Significance (p<.05)**
Native American/African American	0.0390***	0.0086	0.0000
Native American/Hispanic	0.0686***	0.0075	0.0000
Native American/Caucasian	0.0631***	0.0093	0.0000
Native American/Asian American	0.0844***	0.0129	0.0000
African American/Hispanic	0.0295***	0.0044	0.0000
African American/Caucasian	0.0241***	0.0044	0.0000
African American/Asian American	0.0454***	0.0068	0.0000
Hispanic/Asian American	0.0158**	0.0028	0.0059
Hispanic/Caucasian	0.0055	0.0043	0.201
Asian American/Caucasian	0.0213**	0.0075	0.0046
Ethnicity Pairwise Comparisons: Narcissistic			
(I) vs. (J)	**Mean Difference**	**Standard Error**	**Significance (p<.05)**
Native American/African American	0.0517***	0.0097	0.0000
Native American/Hispanic	0.0005	0.0080	0.953
Native American/Caucasian	0.0232**	0.0087	0.0074
Native American/Asian American	0.0205	0.0123	0.096
African American/Hispanic	0.0513***	0.0053	0.0000
African American/Caucasian	0.0750***	0.0044	0.0000
African American/Asian American	0.0723***	0.0083	0.0000
Hispanic/Asian American	0.0210**	0.0068	0.0021
Hispanic/Caucasian	0.0237***	0.0042	0.0000
Asian American/Caucasian	0.0027	0.0072	0.703

Table 12 cont.

Ethnicity Pairwise Comparisons: Histrionic			
(I) vs. (J)	**Mean Difference**	**Standard Error**	**Significance (p<.05)**
Native American/African American	0.0000	0.0044	0.992
Native American/Hispanic	0.0097**	0.0037	0.0086
Native American/Caucasian	0.0082	0.0047	0.083
Native American/Asian American	0.0082	0.0065	0.203
African American/Hispanic	0.0097***	0.0022	0.0000
African American/Caucasian	0.0082***	0.0022	0.0002
African American/Asian American	0.0083*	0.0035	0.019
Hispanic/Asian American	0.0015	0.0030	0.625
Hispanic/Caucasian	0.0015	0.0021	0.474
Asian American/Caucasian	0.0000	0.0036	0.993

Note: "*" significant at p<0.05, "**" p<0.01, and "***" p<0.001. Sources: CBDRB-FY19-422, CBDRB-FY19-448 and CBDRB-FY19-536.

Table 13a.

Antisocial Personality Disorder responses and differential functioning of items and tests from Wave1: Logistic Regression, Mantel-Haenszel (MH) Chi2 and Odds Ratios for gender

Scale Item	**Male vs. Female DIF**		
	CHI2	**Sig.**	**Odds Ratio**
S11A1A1: Often cut class, not go to class or go to school and leave without permission	227.9***	0.0000	1.152*
S11A1A5: Ever have a time when often absent from school, other than when caring for someone who was sick	105.2***	0.0000	1.846*
S11A1A6: More than once quit a job without knowing where would find another one	88.96***	0.0000	1.387*
S11A1A7: More than once quit a school program without knowing what would do next	21.45***	0.0000	1.359*
S11A1A8: Travel from place to place for 1+ months without advance plans or without knowing how long would be gone or where would work	34.20***	0.0000	0.862*
S11A1A9: Ever have time lasting 1+ months when had no regular place to live	33.91***	0.0000	1.280*
S11A1A10: Ever have time lasting 1+ months when lived with others because did not have own place to live	155.1***	0.0000	1.863*
S11A1A11: Ever have time when you lied a lot, other than to avoid being hurt	84.58***	0.0000	1.990*
S11A1A12: Ever use a false or made-up name or alias	39.91***	0.0000	1.529*
S11A1A13: Ever scam or con someone for money, to avoid responsibility or just for fun	16.07***	0.0001	1.008
S11A1A14: Ever do things that could easily have hurt you or someone else, like speeding or driving after having too much to drink	6.33*	0.0119	0.534*
S11A1A15: Ever get more than 3 tickets for reckless/careless driving, speeding, or causing an accident	4.69	0.0304	0.426*
S11A1A16: Ever have driver's license suspended or revoked for moving violations	11.05***	0.0009	0.366*
S11A1A17: Ever destroy/break/vandalize someone else's property (car, home, etc.)	0.40	0.5272	0.544*
S11A1A18: Ever start fire on purpose to destroy someone else's property or just to see it burn	1.49	0.2227	0.505*
S11A1A19: Ever fail to pay off debts -- like moving to avoid rent, not making payments on loan or mortgage, failing to pay alimony or child support or filing bankruptcy	15.13***	0.0001	1.525*
S11A1A20: Ever steal something from someone/someplace when no one was around	30.94***	0.0000	1.088
S11A1A21: Ever forge someone else's signature, like on a legal document or check	19.41***	0.0000	2.122*
S11A1A22: Ever shoplift	97.76***	0.0000	1.292*
S11A1A23: Ever rob or mug someone or snatch a purse	0.87	0.3517	0.481*
S11A1A24: Ever make money illegally, like selling stolen property or selling drugs	2.10	0.1475	0.473*

Table 13a cont.

S11A1A25: Ever do something you could have been arrested for, regardless of whether you were caught or not	0.00	0.9696		0.557*
S11A1A26: Ever force someone to have sex with you against their will	1.76	0.1849		1.497
S11A1A27: Ever get into a lot of fights that you started	23.48***	0.0000		1.050
S11A1A28: Ever get into a fight that came to swapping blows with someone like a husband, wife, boyfriend or girlfriend	352.8***	0.0000		3.506*
S11A1A29: Ever use a weapon like a stick, knife or gun in a fight	24.05***	0.0000		1.107
S11A1A30: Ever hit someone so hard that you injured them or they had to see a doctor	3.44	0.0635		0.402*
S11A1A31: Ever harass, threaten or blackmail someone	12.43***	0.0004		1.860*
S11A1A32: Ever physically hurt another person in any way on purpose	27.10***	0.0000		0.751*
S11A1A33: Ever hurt an animal or pet on purpose	0.14	0.7106		0.385*
	Male vs. Female DTF			
Scale Score	**CHI2**	**Sig.**	**MH CHI2**	**Odds Ratio**
ANTISOX2: Antisocial Personality Disorder	7.48**	0.0062	3.75	0.857*

Note: "*" significant at p<0.05, "**" p<0.01, and "***" p<0. Sources: CBDRB-FY19-422, CBDRB-FY19-448 and CBDRB-FY19-536.

.

Table 13b.

Antisocial Personality Disorder differential functioning of items and tests from Wave1: Logistic Regression, Mantel-Haenszel (MH) Chi2 and Odds Ratios for ethnicity

Scale Item	**Native American Vs. Caucasian DIF**					**Hispanic vs. Caucasian DIF**			
	CHI2	**Sig.**		**Odds Ratio**		**CHI2**	**Sig.**		**Odds Ratio**
S11A1A1	8.10**	0.0044		0.947		85.10***	0.0000		1.665*
S11A1A5	6.78**	0.0092		1.151		17.66***	0.0000		1.647*
S11A1A6	15.92***	0.0001		1.084		0.00	0.9478		0.855*
S11A1A7	5.77*	0.0163		0.902		0.08	0.7713		1.561*
S11A1A8	9.39**	0.0022		1.235		4.20*	0.0403		0.983
S11A1A9	3.35	0.0673		1.686*		7.36**	0.0067		1.015
S11A1A10	0.59	0.4412		1.407*		6.13*	0.0133		1.206*
S11A1A11	9.77**	0.0018		1.016		0.34	0.5571		1.744*
S11A1A12	4.19*	0.0406		1.380		0.00	0.9897		2.038*
S11A1A13	4.38*	0.0364		0.850		1.23	0.2675		1.140
S11A1A14	16.84***	0.0000		0.621*		5.95*	0.0147		0.428*
S11A1A15	8.20**	0.0042		0.694*		1.10	0.2948		0.774*
S11A1A16	18.52***	0.0000		1.024		4.00*	0.0456		1.224*
S11A1A17	5.67*	0.0173		0.871		8.28**	0.0040		0.841
S11A1A18	8.66**	0.0033		0.566		0.24	0.6237		0.717*
S11A1A19	0.47	0.4942		1.177		0.81	0.3679		0.894
S11A1A20	7.46**	0.0063		0.923		3.69	0.0546		0.786*
S11A1A21	0.01	0.9105		0.578*		3.37	0.0664		1.066
S11A1A22	8.25**	0.0041		0.809		3.34	0.0677		0.725*
S11A1A23	1.25	0.3641		1.458		2.38	0.1227		2.156*
S11A1A24	0.91	0.3394		0.991		6.48*	0.0109		0.896
S11A1A25	5.46*	0.0194		0.755*		1.66	0.1972		0.475*
S11A1A26	0.13	0.7169		0.813		0.14	0.7049		2.018*
S11A1A27	0.69	0.4076		0.949		1.22	0.2686		1.480*
S11A1A28	4.14*	0.0420		1.523*		8.68**	0.0032		1.064
S11A1A29	0.49	0.4822		1.558*		3.13	0.0770		1.421*
S11A1A30	3.39	0.0656		1.293		1.40	0.2374		0.967
S11A1A31	0.70	0.4019		0.955		0.18	0.6719		1.081
S11A1A32	1.30	0.2551		0.848		9.45**	0.0021		0.766*
S11A1A33	9.08**	0.0026		1.363		1.10	0.2943		0.943
	Native American Vs. Caucasian DTF					**Hispanic vs. Caucasian DTF**			
Scale Score	**CHI2**	**Sig.**	**MH CHI2**	**Odds Ratio**		**CHI2**	**Sig.**	**MH CHI2**	**Odds Ratio**
ANTISOX2	1.57	0.2100	0.08	1.078		4.30*	0.0381	7.25**	1.332*

Table 13b cont.

Scale Item	African American vs. Caucasian DIF				Asian American vs. Caucasian DIF			
	CHI2	Sig.		Odds Ratio	CHI2	Sig.		Odds Ratio
S11A1A1	0.71	0.3993		1.177*	42.19***	0.0000		1.633*
S11A1A5	0.03	0.8669		1.085	12.21***	0.0005		1.830*
S11A1A6	0.05	0.8257		1.050	7.46**	0.0063		1.238
S11A1A7	1.48	0.2235		1.308*	6.28*	0.0122		2.478*
S11A1A8	2.16	0.1420		0.703*	1.03	0.3111		1.460
S11A1A9	9.29**	0.0023		0.886	1.93	0.1646		1.117
S11A1A10	0.52	0.4712		1.352*	1.45	0.2291		1.136
S11A1A11	2.91	0.0882		1.488*	3.91*	0.0479		1.801*
S11A1A12	6.65**	0.0099		2.012*	6.34*	0.0118		1.821*
S11A1A13	1.38	0.2394		1.811*	0.03	0.8642		1.589
S11A1A14	88.28***	0.0000		0.265*	0.44	0.5063		0.430*
S11A1A15	7.38**	0.0066		0.648*	2.30	0.1297		0.711*
S11A1A16	16.45***	0.0001		1.419*	1.20	0.2732		0.712
S11A1A17	0.84	0.3590		0.564*	7.78**	0.0053		0.825
S11A1A18	0.66	0.4167		0.774	0.08	0.7804		0.509
S11A1A19	2.85	0.0913		1.269*	5.71*	0.0168		1.054
S11A1A20	2.70	0.1004		0.815*	8.34**	0.0039		1.373
S11A1A21	6.34*	0.0118		0.672*	4.89*	0.0270		1.544
S11A1A22	18.91***	0.0000		0.592*	8.85**	0.0029		0.684*
S11A1A23	4.70*	0.0302		2.990*	3.52	0.0606		0.841
S11A1A24	0.00	0.9457		1.103	0.10	0.7502		0.736
S11A1A25	6.11*	0.0134		0.472*	0.28	0.5967		0.414*
S11A1A26	0.07	0.7876		2.582*	0.06	0.8134		3.290
S11A1A27	1.23	0.2681		0.991	3.89*	0.0485		0.779
S11A1A28	17.78***	0.0000		2.249*	5.13*	0.0236		0.403*
S11A1A29	16.95***	0.0000		3.925*	3.76	0.0525		0.934
S11A1A30	0.00	0.9589		1.378*	6.10*	0.0135		0.568*
S11A1A31	0.02	0.8866		1.328*	4.99*	0.0255		1.559
S11A1A32	0.00	0.9602		0.960	7.98**	0.0047		0.965
S11A1A33	0.90	0.3440		1.324*	2.78	0.0953		1.292
	African American vs. Caucasian DTF				Asian American vs. Caucasian DTF			
Scale Score	CHI2	Sig.	MH CHI2	Odds Ratio	CHI2	Sig.	MH CHI2	Odds Ratio
ANTISOX2	4.89*	0.0270	0.96	0.902	6.09*	0.0136	0.04	1.134

Note: "*" significant at p<0.05, "**" p<0.01, and "***" p<0.001. Sources: CBDRB-FY19-422, CBDRB-FY19-448 and CBDRB-FY19-536.

that came to swapping blows with someone like a husband, wife, boyfriend or girlfriend." This was consistent with mean differences and ANOVA pairwise comparisons. This indicates that the items with differential functioning may be contributing somewhat to the observed mean differences. This indicates that these scale items are introducing bias into the diagnostic criteria for this scale. To determine if these items were contributing to the overall score mean differences, test level functioning was examined next. Differential Test Functioning indicated that these differences were significant at the overall antisocial scale level for all comparisons, except for Native Americans compared to Caucasians, obtaining a significant chi-squared results. This indicates the presence of differential functioning, or bias, for these minorities at the scale level, as well as for gender. MH and ORs were examined to determine the impact or effect that this would have upon the total score of participants. The OR was significant for Hispanics compared to Caucasians, with an odds ratio of 1.332. It was also significant for gender, with an OR of 0.857, but no significant MH chi-squared. Overall, these results indicated differential test functioning, with a notable impact for gender, with men tending to be scored higher than women, and for Hispanics, with this minority tending to score higher than Caucasians.

Borderline Personality Disorder scale was examined next. A significant ANOVA was found for both gender and ethnicity (see Table 7). Pairwise comparisons (Table 12) indicated that mean scale differences between most ethnicities were significant, except for Hispanics compared to Caucasians. Native Americans responded positively more frequently than all other groups, and Asian Americans responded positively less often than other groups. Women responded positively more often than men. Differential item functioning was examined next, to determine if these results were due to true differences between the populations, or if bias had been introduced into the items. The borderline scale contained eighteen items. Eight items within the borderline scale

had significant chi-squared for gender (see Table 14a). Native Americans had five items contributing bias to the significant difference in means when compared with Caucasians (see Table 14b). When compared with Caucasians, Hispanic participants had significant differences in item functioning in five out of the eighteen items. African Americans compared with Caucasians also had bias present in seven of the items in the borderline scale. Differential functioning for items for comparisons between Caucasians and Asian Americans indicated bias within two of the items. Some common items between groups which contained bias were item W2S10Q1A9, "Has it been very important to you that people pay attention to you or admire you in some way," and W2S10Q1A14, "Have you often expected other people to do what you ask without question because of who you are." This was generally consistent with mean differences and ANOVA pairwise comparisons. This indicates that the items with differential functioning may be contributing somewhat to the observed mean differences. This indicates that these scale items are introducing bias into the diagnostic criteria for this scale. To determine if these items were contributing to the overall score mean differences, test level functioning was examined next. Differential Test Functioning indicated that the only differences that remained significant at the overall borderline scale level was for gender and for Native Americans compared to Caucasians. These comparisons obtained significant chi-squared results. This indicates the presence of differential functioning, or bias, for this minority at the scale level, as well as for gender. MH and ORs were examined to determine the impact or effect that this would have upon the total score of participants. The OR was significant for gender, with an OR of 1.396, with a significant MH chi-squared. The OR for Native Americans was not significant. Interestingly, the MH and OR were significant for African Americans compared to Caucasians. Overall, these results indicated

Table 14a.

Narcissistic Personality Disorder responses and differential functioning of items and tests: Logistic Regression, Mantel-Haenszel (MH) Chi2 and Odds Ratios for gender

Scale Item	**Male vs. Female DIF**			
	CHI2	**Sig.**		**Odds Ratio**
W2S10Q1A2: Have people often failed to appreciate your very special talents or accomplishments	17.54***	0.0000		1.387*
W2S10Q1A3: Have people told you that you have too high an opinion of yourself	4.75*	0.0294		0.798*
W2S10Q1A4: Have you thought a lot about the power, fame, or recognition that will be yours someday	21.78***	0.0000		0.926
W2S10Q1A5: Have you thought a lot about the perfect romance that will be yours someday	52.45***	0.0000		1.466*
W2S10Q1A6: Have you almost always insisted on seeing the top person when you have a problem	28.31***	0.0000		1.084*
W2S10Q1A7: Have you felt it was important to spend time with people who are important or influential	56.02***	0.0000		0.931*
W2S10Q1A8: Have you often found yourself losing interest in people after they have served their purpose	5.33*	0.0209		0.962
W2S10Q1A9: Has it been very important to you that people pay attention to you or admire you in some way	41.49***	0.0000		1.049
W2S10Q1A10: Have you thought that you could ignore certain rules or social conventions when they get in your way	1.87	0.1717		0.748*
W2S10Q1A11: Have you felt that you were the kind of person who deserves special treatment	3.93*	0.0474		1.777*
W2S10Q1A12: Have you often found it necessary to step on a few toes to get what you want	9.74**	0.0018		0.704*
W2S10Q1A13: Have you often put your needs above other people's	6.29*	0.0121		1.138*
W2S10Q1A14: Have you often expected other people to do what you ask without question because of who you are	7.00**	0.0082		0.739*
W2S10Q1A15: Have other people's problems or feelings failed to interest you	2.03	0.1541		0.691*
W2S10Q1A16: Have people complained to you that you don't listen to them or care about their feelings	5.44*	0.0197		0.480*
W2S10Q1A17: Have you often been envious of others	0.63	0.4266		1.179*
W2S10Q1A18: Have you felt that others are often envious of you	23.32***	0.0000		1.222*
W2S10Q1A19: Have you found that there are very few people who are worth your time and attention	36.92***	0.0000		1.200*
	Male vs. Female DTF			
Scale Score	**CHI2**	**Sig.**	**MH CHI2**	**Odds Ratio**
NARCDX: Narcissistic Personality Disorder	5.85*	0.0156	10.2**	1.239*

Note: "*" significant at p<0.05, "**" p<0.01, and "***" p<0.001. Sources: CBDRB-FY19-422, CBDRB-FY19-448 and CBDRB-FY19-536.

Table 14b.

Narcissistic Personality Disorder differential functioning of items and tests: Logistic Regression, Mantel-Haenszel (MH) Chi2 and Odds Ratios for ethnicity

Scale Item	**Native American Vs. Caucasian DIF**					**Hispanic vs. Caucasian DIF**			
	CHI2	**Sig.**		**Odds Ratio**		**CHI2**	**Sig.**		**Odds Ratio**
W2S10Q1A2	0.05	0.8162		1.390*		13.95***	0.0002		1.305*
W2S10Q1A3	1.43	0.2320		1.495*		13.75***	0.0002		1.588*
W2S10Q1A4	0.27	0.6028		1.877*		0.01	0.9208		1.996*
W2S10Q1A5	0.05	0.8272		1.152		0.60	0.4367		1.101*
W2S10Q1A6	0.13	0.7154		1.049		2.56	0.1095		0.953
W2S10Q1A7	3.11	0.0778		0.951		0.13	0.7141		1.306*
W2S10Q1A8	0.50	0.4794		0.676		2.29	0.1303		0.793*
W2S10Q1A9	12.70***	0.0004		0670*		12.39***	0.0004		1.113*
W2S10Q1A10	4.21*	0.0401		1.013		16.54***	0.0000		0.710*
W2S10Q1A11	0.99	0.3198		1.167		18.68***	0.0000		1.825*
W2S10Q1A12	6.21*	0.0127		0.817		5.12*	0.0236		0.604*
W2S10Q1A13	0.14	0.7079		0.653*		16.78***	0.0000		0.661*
W2S10Q1A14	3.92*	0.0476		0.753*		7.47**	0.0063		0.442*
W2S10Q1A15	0.95	0.3299		0.690*		31.32***	0.0000		1.221*
W2S10Q1A16	0.36	0.5501		1.145		17.77***	0.0000		0.664*
W2S10Q1A17	6.24*	0.0125		0.763*		28.67***	0.0000		0.307*
W2S10Q1A18	0.04	0.8439		1.113		0.15	0.7000		1.351*
W2S10Q1A19	0.18	0.6726		1.511*		6.11*	0.0135		2.018*
	Native American Vs. Caucasian DTF					**Hispanic vs. Caucasian DTF**			
Scale Score	**CHI2**	**Sig.**	**MH CHI2**	**Odds Ratio**		**CHI2**	**Sig.**	**MH CHI2**	**Odds Ratio**
NARCDX	0.04	0.8390	0.01	1.051		7.89**	0.0050	14.6***	0.704*

Table 14b cont.

Scale Item	African American vs. Caucasian DIF					Asian American vs. Caucasian DIF			
	CHI2	Sig.		Odds Ratio		CHI2	Sig.		Odds Ratio
W2S10Q1A2	29.39***	0.0000		1.282*		5.26*	0.0219		1.103
W2S10Q1A3	42.32***	0.0000		1.398*		4.68*	0.0305		0.975
W2S10Q1A4	14.28***	0.0002		2.906*		0.03	0.8629		2.835*
W2S10Q1A5	4.85*	0.0277		1.589*		1.37	0.2421		1.480*
W2S10Q1A6	34.06***	0.0000		1.067		0.03	0.8528		0.596*
W2S10Q1A7	17.16***	0.0000		1.467*		2.46	0.1168		1.375*
W2S10Q1A8	24.59***	0.0000		0.972		0.92	0.3377		1.225
W2S10Q1A9	66.01***	0.0000		0.726*		1.08	0.2995		1.282*
W2S10Q1A10	37.38***	0.0000		0.624*		0.31	0.5802		0.845
W2S10Q1A11	25.04***	0.0000		2.032*		0.02	0.8812		2.227*
W2S10Q1A12	50.88***	0.0000		0.483*		1.95	0.1626		0.611*
W2S10Q1A13	86.49***	0.0000		0.812*		0.08	0.7806		0.679*
W2S10Q1A14	45.01***	0.0000		0.540*		0.42	0.5178		0.587*
W2S10Q1A15	62.39***	0.0000		0.648*		0.26	0.6129		0.776
W2S10Q1A16	23.02***	0.0000		0.793*		2.47	0.1161		0.692*
W2S10Q1A17	41.91***	0.0000		0.186*		6.13*	0.0133		0.651*
W2S10Q1A18	3.96*	0.0465		1.517*		0.00	0.9687		1.003
W2S10Q1A19	0.37	0.5418		1.888*		1.39	0.2386		1.629*
	African American vs. Caucasian DTF					Asian American vs. Caucasian DTF			
Scale Score	CHI2	Sig.	MH CHI2	Odds Ratio		CHI2	Sig.	MH CHI2	Odds Ratio
NARCDX	78.38***	0.0000	59.3***	0.541*		0.05	0.8308	6.18*	0.565*

Note: "*" significant at p<0.05, "**" p<0.01, and "***" p<0.001. Sources: CBDRB-FY19-422, CBDRB-FY19-448 and CBDRB-FY19-536.

differential test functioning, or scale level bias, with a notable impact for gender, with women tending to be scored higher than men.

Narcissistic Personality Disorder scale was then examined. A significant ANOVA was found for both gender and ethnicity (see Table 7). Pairwise comparisons (Table 12) indicated that mean scale differences between most ethnicities were significant, except for Native Americans compared to Hispanic or Asian American participants, and Asian Americans compared to Caucasians. Men responded positively more often than women. Differential item functioning was examined next, to determine if these results were due to true differences between the populations, or if bias had been introduced into the items. The narcissistic scale contained eighteen items. Fifteen items within the narcissistic scale had significant chi-squared for gender (see Table 15a). Native Americans had five items contributing bias to the significant difference in means when compared with Caucasians (see Table 15b). When compared with Caucasians, Hispanic participants had significant differences in item functioning in twelve out of the eighteen items. African Americans compared with Caucasians also had bias present in seventeen of the items in the narcissistic scale. Differential functioning for items for comparisons between Caucasians and Asian Americans indicated bias within three of the items. Some common items between groups which contained bias were item W2S10Q1A20, “Have you ever gotten into sexual relation quickly or without thinking about the consequences,” and item W2S10Q1A28, “Have you often done things impulsively.” The items with differential functioning may be contributing somewhat to the observed mean differences, and this was examined next to determine if the items the scale belong to are introducing bias into the diagnostic criteria for this scale and to determine the impact these item biases had on overall score mean differences. Differential Test Functioning (DTF) indicated that the gender discrepancies for items remained at

Table 15a.

Borderline Personality Disorder responses and differential functioning of items and tests: Logistic Regression, Mantel-Haenszel (MH) Chi2 and Odds Ratios for gender

Scale Item	**Male vs. Female DIF**		
	CHI2	**Sig.**	**Odds Ratio**
W2S10Q1A20: Have you ever gotten into sexual relation quickly or without thinking about the consequences	13.55***	0.0002	0.433*
W2S10Q1A21: Have you had a problem with gambling or spending too much money	12.37***	0.0004	0.849*
W2S10Q1A22: Have you often become frantic when you thought that someone you really cared about was going to leave you	5.36*	0.0206	1.029
W2S10Q1A23: Have your relationships with people you really care about had lots of extreme ups and downs	1.56	0.2116	1.204*
W2S10Q1A24: Have you all of a sudden changed your sense of who you are and where you are headed	1.47	0.2248	1.133*
W2S10Q1A25: Your sense of who you are has often changed depending on the situation or whom you are with	0.14	0.7066	1.064
W2S10Q1A26: Have you been so different with different people or in different situations that you sometimes don't know who you really are	0.12	0.7265	0.990
W2S10Q1A27: Have there been lots of sudden changes in your personal goals, career plans, religious beliefs, or other important aspects of your life	3.03	0.0816	1.077
W2S10Q1A28: Have you often done things impulsively	37.07***	0.0000	0.761*
W2S10Q1A29: Have you tried to hurt or kill yourself, or threatened to do so	0.24	0.6248	1.729*
W2S10Q1A30: Have you ever cut, burned, or scratched yourself on purpose	0.33	0.5652	1.166
W2S10Q1A31: Have you had a lot of sudden mood changes	2.77	0.0962	1.878*
W2S10Q1A32: Have you gone to extremes to keep people from leaving you	1.38	0.2399	0.800*
W2S10Q1A33: Have you often felt empty inside	5.34*	0.0209	1.705*
W2S10Q1A34: Have you often had temper outbursts or gotten so angry that you lose control	4.99*	0.0255	1.027
W2S10Q1A35: Have you hit people or thrown things when you got angry	5.32*	0.0211	1.173*
W2S10Q1A36: Have even little things made you angry or have you had difficulty controlling your anger	1.24	0.2654	0.923
W2S10Q1A37: Have you gotten suspicious of other people or felt spaced out under a lot of stress	9.61**	0.0019	1.068
	Male vs. Female DTF		
Scale Score	**CHI2**	**Sig.**	**MH CHI2**	**Odds Ratio**
BPDDX: Borderline Personality Disorder	7.19**	0.0073	9.57**	1.396*

Note: "*" significant at p<0.05, "**" p<0.01, and "***" p<0.001. Sources: CBDRB-FY19-422, CBDRB-FY19-448 and CBDRB-FY19-536.

Table 15b.

Borderline Personality Disorder differential functioning of items and tests: Logistic Regression, Mantel-Haenszel (MH) Chi2 and Odds Ratios for ethnicity

Scale Item	**Native American Vs. Caucasian DIF**				**Hispanic vs. Caucasian DIF**			
	CHI2	**Sig.**		**Odds Ratio**	**CHI2**	**Sig.**		**Odds Ratio**
W2S10Q1A20	7.47**	0.0063		0.961	9.68**	0.0019		0.648*
W2S10Q1A21	0.25	0.6201		1.448*	1.59	0.2080		0.821*
W2S10Q1A22	4.93*	0.0264		0.936	6.11*	0.0135		1.031
W2S10Q1A23	8.94**	0.0028		1.134	7.00**	0.0082		1.321*
W2S10Q1A24	4.43	0.0353		0.980	3.94*	0.0471		1.209*
W2S10Q1A25	0.10	0.7490		0.980	0.23	0.6292		1.143
W2S10Q1A26	1.57	0.2101		1.049	3.44	0.0638		1.415*
W2S10Q1A27	1.13	0.2881		0.942	0.16	0.6892		1.567*
W2S10Q1A28	0.97	0.3247		0.902	1.56	0.2110		0.758*
W2S10Q1A29	0.19	0.6651		1.407	0.12	0.7315		0.778*
W2S10Q1A30	1.18	0.2767		1.112	0.00	0.9535		0.765*
W2S10Q1A31	0.36	0.5491		1.103	0.86	0.3530		1.434*
W2S10Q1A32	0.30	0.5859		0.920	0.13	0.7188		0.894
W2S10Q1A33	2.10	0.1473		0.997	1.82	0.1772		1.235*
W2S10Q1A34	3.68	0.0550		0.938	0.00	0.9470		1.021
W2S10Q1A35	10.32**	0.0013		0.878	0.37	0.5431		0.876*
W2S10Q1A36	0.04	0.8456		0.736	3.87*	0.0490		0.838*
W2S10Q1A37	0.03	0.8685		1.031	0.56	0.4535		0.987
	Native American Vs. Caucasian DTF				**Hispanic vs. Caucasian DTF**			
Scale Score	**CHI2**	**Sig.**	**MH CHI2**	**Odds Ratio**	**CHI2**	**Sig.**	**MH CHI2**	**Odds Ratio**
BPDDX	5.77*	0.0163	0.00	0.962*	3.50	0.0612	2.17	0.797

Table 15b cont.

Scale Item	African American vs. Caucasian DIF					Asian American vs. Caucasian DIF			
	CHI2	Sig.		Odds Ratio		CHI2	Sig.		Odds Ratio
W2S10Q1A20	20.38***	0.0000		0.853*		2.00	0.1576		0.520*
W2S10Q1A21	0.08	0.7752		1.552*		5.89*	0.0153		1.119
W2S10Q1A22	2.25	0.1335		0.946		1.93	0.1650		1.567*
W2S10Q1A23	1.06	0.3042		1.778*		3.92*	0.0477		0.962
W2S10Q1A24	0.88	0.3477		1.485*		2.93	0.0868		1.034
W2S10Q1A25	2.94	0.0863		1.544*		3.37	0.0664		2.062*
W2S10Q1A26	0.21	0.6503		1.279*		0.55	0.4577		1.944*
W2S10Q1A27	0.16	0.6881		1.434*		1.22	0.2684		0.893
W2S10Q1A28	20.05***	0.0000		0.694*		0.64	0.4241		0.870
W2S10Q1A29	5.90*	0.0151		0.438*		1.66	0.1970		0.857
W2S10Q1A30	0.47	0.4938		0.338*		0.02	0.8901		0.850
W2S10Q1A31	6.28*	0.0122		0.961		0.48	0.4875		1.174
W2S10Q1A32	0.41	0.5244		0.873		0.01	0.9120		1.348
W2S10Q1A33	3.32	0.0684		0.801*		0.69	0.4047		0.835
W2S10Q1A34	7.81**	0.0052		0.878		0.90	0.3428		1.197
W2S10Q1A35	3.69	0.0548		0.867*		0.02	0.8842		0.656*
W2S10Q1A36	3.28	0.0702		0.639*		0.35	0.5526		0.870
W2S10Q1A37	0.34	0.5605		0.862*		0.03	0.8703		1.181
	African American vs. Caucasian DTF					Asian American vs. Caucasian DTF			
Scale Score	CHI2	Sig.	MH CHI2	Odds Ratio		CHI2	Sig.	MH CHI2	Odds Ratio
BPDDX	3.50	0.0614	8.87**	0.669*		0.00	0.9626	0.27	1.344

Note: "*" significant at p<0.05, "**" p<0.01, and "***" p<0.001. Sources: CBDRB-FY19-422, CBDRB-FY19-448 and CBDRB-FY19-536.

the test level. MH Chi-square and ORs for gender were also significant. Interestingly, women were noted to be the recipients of the differential test functioning, with an OR of 1.239. Native American compared to Caucasian DTF was not significant, nor was the MH Chi-square nor the OR. Hispanic DTF remained significant, as was the MH Chi-square and the OR (0.704). African American compared to Caucasian DTF Logistic Regression Chi-square was significant, as was MH Chi-square and the OR (0.541). Asian American compared to Caucasian DTF was not significant, however a significant MH and OR were observed despite this. Overall, these results indicated differential test functioning (scale level bias) with a notable impact for gender and ethnicity, with women tending to be scored higher than men despite the higher weighted mean and prevalence for men, and African Americans and Hispanic populations tending to score higher than Caucasians.

Histrionic Personality Disorder scale was examined next. A significant ANOVA was found for ethnicity but not gender (see Table 7). Pairwise comparisons (Table 12) indicated that mean scale differences between most ethnicities were not significant. Significant comparisons existed between Native American and Hispanic, and between African American and Asian American, Caucasian, and Hispanic populations, with both populations tending to respond positively more frequently than their comparison groups. Differential item functioning was examined next, to determine if these results were due to true differences between the populations, or if bias had been introduced into the items. The histrionic scale contained eleven items. Five items within the histrionic scale had significant chi-squared for gender (see Table 16a). Native Americans had three items contributing bias to the significant difference in means when compared with Caucasians (see Table 16b). When compared with Caucasians, Hispanic participants had significant differences in item functioning in three out of the eleven items.

Table 16a.

Histrionic Personality Disorder responses and differential functioning of items and tests: Logistic Regression, Mantel-Haenszel (MH) Chi2 and Odds Ratios for gender

Scale Item	**Male vs. Female DIF**			
	CHI2	**Sig.**		**Odds Ratio**
S10Q1A54: Like to be the center of attention	0.34	0.5611		0.775*
S10Q1A55: Feelings often change very suddenly/unexpectedly, sometimes for no reason	42.56***	0.0000		2.048*
S10Q1A56: Feel uncomfortable if not center of attention	0.33	0.5673		1.158
S10Q1A57: Ever discovered that people aren't as close as you thought they were	31.65***	0.0000		1.335*
S10Q1A58: Flirt a lot	1.29	0.2557		0.434*
S10Q1A59: Display emotions in obvious/dramatic ways so people always know how you feel	5.02*	0.0251		1.373*
S10Q1A60: Often find yourself "coming on" to people	4.99*	0.0255		0.353*
S10Q1A61: Try to draw attention to yourself by way you dress or look	1.34	0.2473		0.762*
S10Q1A62: Often make a point of being dramatic and colorful	0.00	0.9900		1.101
S10Q1A63: Change mind about things depending on people you're with or what read or saw on TV	27.18***	0.0000		1.337*
S10Q1A64: Often express self using generalities and very little detail	1.13	0.2876		0.887*
	Male vs. Female DTF			
Scale Score	**CHI2**	**Sig.**	**MH CHI2**	**Odds Ratio**
HISTDX2: Histrionic Personality Disorder	0.53	0.4655	2.79*	1.395*

Note: "*" significant at p<0.05, "**" p<0.01, and "***" p<0.001. Sources: CBDRB-FY19-422, CBDRB-FY19-448 and CBDRB-FY19-536.

Table 16b.

Histrionic Personality Disorder differential functioning of items and tests: Logistic Regression, Mantel-Haenszel (MH) Chi2 and Odds Ratios for ethnicity

Scale Item	**Native American Vs. Caucasian DIF**				**Hispanic vs. Caucasian DIF**			
	CHI2	**Sig.**		**Odds Ratio**	**CHI2**	**Sig.**		**Odds Ratio**
S10Q1A54	0.00	0.9954		0.685*	2.92	0.0872		0.883*
S10Q1A55	9.01**	0.0027		1.247	1.01	0.3138		1.283*
S10Q1A56	1.54	0.2145		0.726	0.01	0.9085		1.747*
S10Q1A57	3.95*	0.0469		1.191	0.89	0.3461		0.847*
S10Q1A58	2.03	0.1540		1.116	0.09	0.7615		1.280*
S10Q1A59	0.77	0.3810		1.186	1.34	0.2475		0.763*
S10Q1A60	0.96	0.3270		1.178	9.19**	0.0024		2.492*
S10Q1A61	0.14	0.7095		0.822	0.02	0.8840		1.580*
S10Q1A62	7.47**	0.0063		1.081	5.03*	0.0250		0.656*
S10Q1A63	1.10	0.2943		0.867	1.26	0.2610		0.874*
S10Q1A64	2.47	0.1163		0.823	6.41*	0.0114		0.904*
	Native American Vs. Caucasian DTF				**Hispanic vs. Caucasian DTF**			
Scale Score	**CHI2**	**Sig.**	**MH CHI2**	**Odds Ratio**	**CHI2**	**Sig.**	**MH CHI2**	**Odds Ratio**
HISTDX2	10.58**	0.0011	0.01	1.036	3.01	0.0826	1.72	0.769

Scale Item	**African American vs. Caucasian DIF**				**Asian American vs. Caucasian DIF**			
	CHI2	**Sig.**		**Odds Ratio**	**CHI2**	**Sig.**		**Odds Ratio**
S10Q1A54	23.45***	0.0000		0.554*	4.57*	0.0325		1.523*
S10Q1A55	1.10	0.2933		1.202*	0.10	0.7465		0.833
S10Q1A56	1.66	0.1975		1.141	028	0.5982		2.006*
S10Q1A57	9.80**	0.0017		1.583*	0.52	0.4705		0.835
S10Q1A58	16.01***	0.0001		0.988	0.01	0.9037		0.551*
S10Q1A59	1.66	0.1981		0.857*	0.19	0.6653		0.619*
S10Q1A60	7.97	0.0048		1.662*	4.19*	0.0407		1.334
S10Q1A61	12.31***	0.0005		1.341*	3.61	0.0575		3.187*
S10Q1A62	202	0.1551		0.903	126	0.2618		0.622*
S10Q1A63	0.19	0.6591		1.016	0.22	0.6378		1.073
S10Q1A64	0.33	0.5666		0.774*	2.48	0.1155		1.235*
	African American vs. Caucasian DTF				**Asian American vs. Caucasian DTF**			
Scale Score	**CHI2**	**Sig.**	**MH CHI2**	**Odds Ratio**	**CHI2**	**Sig.**	**MH CHI2**	**Odds Ratio**
HISTDX2	10.86***	0.0010	13.3***	0.553*	0.56	0.4530	0.67	0.653

Note: "*" significant at p<0.05, "**" p<0.01, and "***" p<0.001. Sources: CBDRB-FY19-422, CBDRB-FY19-448 and CBDRB-FY19-536.

African Americans compared with Caucasians also had bias present in four of the items in the histrionic scale. Differential functioning for items for comparisons between Caucasians and Asian Americans indicated bias within two of the items. A common item between groups which contained differential item functioning was item S10Q1A57, “Ever discovered that people aren't as close as you thought they were.” This indicates that the items with differential functioning may be contributing somewhat to the observed mean differences, and may provide evidence that these scale items are introducing bias into the diagnostic criteria for this scale. To determine if these items were contributing to the overall score mean differences, test level functioning was examined next. Differential Test Functioning indicated that the only difference that remained significant at the overall histrionic scale level was for Native and African Americans compared to Caucasians. This comparison obtained significant logistic regression chi-squared results. This indicates the presence of differential functioning, or bias, for this minority at the scale level. MH and ORs were examined to determine the impact or effect that this would have upon the total score of participants. The OR was significant for both African Americans, with an OR of 0.553, with a significant MH chi-squared. The MH and OR for Native Americans was not significant. Interestingly, the MH and OR were significant for gender. Overall, these results indicated differential test functioning, or scale level bias, with a notable impact for African Americans compared to Caucasians, with African Americans responding positively and tending to score higher than Caucasians.

Finally, Cluster C personality disorders were examined for their differential functioning at the item and test level. Avoidant Personality Disorder scale was examined first. A significant ANOVA was found for both ethnicity and gender (see Table 7). Pairwise comparisons (Table 17) indicated that mean scale differences between most ethnicities were not significant. Significant

Table 17.

Pairwise comparisons for Ethnicity for Cluster C Personality Disorders.

Ethnicity Pairwise Comparisons: Avoidant			
(I) vs. (J)	**Mean Difference**	**Standard Error**	**Significance (p<.05)**
Native American/African American	0.0155***	0.0041	0.0002
Native American/Hispanic	0.0155***	0.0041	0.0002
Native American/Caucasian	0.0109*	0.0056	0.050
Native American/Asian American	0.0126	0.0074	0.090
African American/Hispanic	0.000	0.0022	0.999
African American/Caucasian	0.0046	0.0025	0.066
African American/Asian American	0.0029	0.0033	0.387
Hispanic/Asian American	0.0029	0.0033	0.388
Hispanic/Caucasian	0.0046	0.0024	0.059
Asian American/Caucasian	0.0017	0.0043	0.689
Ethnicity Pairwise Comparisons: Dependent			
(I) vs. (J)	**Mean Difference**	**Standard Error**	**Significance (p<.05)**
Native American/African American	N/A	N/A	N/A
Native American/Hispanic	N/A	N/A	N/A
Native American/Caucasian	N/A	N/A	N/A
Native American/Asian American	N/A	N/A	N/A
African American/Hispanic	0.0004	0.0010	0.695
African American/Caucasian	0.0016	0.0012	0.172
African American/Asian American	N/A	N/A	N/A
Hispanic/Asian American	N/A	N/A	N/A
Hispanic/Caucasian	0.0012	0.0011	0.287
Asian American/Caucasian	N/A	N/A	N/A
Ethnicity Pairwise Comparisons: Obsessive-Compulsive			
(I) vs. (J)	**Mean Difference**	**Standard Error**	**Significance (p<.05)**
Native American/African American	0.0179*	0.0076	0.019
Native American/Hispanic	0.0377***	0.0069	0.0000
Native American/Caucasian	0.0146	0.0099	0.138
Native American/Asian American	0.0515***	0.0112	0.0000
African American/Hispanic	0.0198***	0.0040	0.0000
African American/Caucasian	0.0033	0.0045	0.471
African American/Asian American	0.0336***	0.0060	0.0000
Hispanic/Asian American	0.0137*	0.0054	0.011
Hispanic/Caucasian	0.0231***	0.0043	0.0000
Asian American/Caucasian	0.0368***	0.0076	0.0000

Note: "*" significant at p<0.05, "**" p<0.01, and "***" p<0.001 Sources: CBDRB-FY19-422, CBDRB-FY19-448 and CBDRB-FY19-536.

comparisons existed for Native American participants compared to Hispanic and African American participants, with this group responding more positively than their comparison groups. Women responded positively more frequently than men. Differential item functioning was examined next, to determine if item bias had been introduced. The histrionic scale contained seven items. Four items within the avoidant scale had significant chi-squared for gender (see Table 18a). Native Americans had no items contributing bias to the significant difference in means when compared with Caucasians (see Table 18b). When compared with Caucasians, Hispanic participants had significant differences in item functioning in three out of the seven items. African Americans compared with Caucasians also had bias present in one of the items in the histrionic scale. Differential functioning for items for comparisons between Caucasians and Asian Americans indicated no bias within the items. A common item between groups which contained differential item functioning was item S10Q1A3 “Find it hard to be "open" even with people you are close to.” To determine if these items were contributing to the overall score mean differences, test level functioning was examined next. Differential Test Functioning indicated no significant differences at the overall histrionic scale level for any of the five comparisons made. This indicates the that there appears to be no differential functioning, or bias, for this scale for either minority compared to Caucasians or gender. MH and ORs were examined to determine any possible remaining impact or effect upon the total score of participants. Both MH Chi-squares and ORs were nonsignificant for all groups tested. Overall, these results indicated no differential test functioning, or scale level bias.

Dependent Personality Disorder scale was examined next. A significant ANOVA was found for and gender but not for ethnicity (see Table 7). Women responded positively more frequently than men. Pairwise comparisons (Table 17) indicated that mean scale differences

Table 18a.

Avoidant Personality Disorder responses and differential functioning of items and tests: Logistic Regression, Mantel-Haenszel (MH) Chi2 and Odds Ratios for gender

Scale Item	**Male vs. Female DIF**			
	CHI2	**Sig.**		**Odds Ratio**
S10Q1A1: Avoid jobs or tasks that deal with a lot of people	52.60***	0.0000		0.619*
S10Q1A2: Avoid getting involved with people unless certain they will like you	33.49***	0.0000		0.689*
S10Q1A3: Find it hard to be "open" even with people you are close to	68.72***	0.0000		0.757*
S10Q1A4: Often worry about being criticized or rejected in social situations	0.009	0.7581		1.241*
S10Q1A5: Believe that you are not as good, as smart, or as attractive as most people	17.22***	0.0000		1.642*
S10Q1A6: Usually quiet or have very little to say when meeting new people because you believe they are better than you are	0.82	0.3648		0.912
S10Q1A7: Afraid of trying new things or doing things outside usual routine because afraid of being embarrassed	2.77	0.0963		1.604*
	Male vs. Female DTF			
Scale Score	**CHI2**	**Sig.**	**MH CHI2**	**Odds Ratio**
AVODPDX2: Avoidant Personality Disorder	0.00	0.9973	0.01	1.032

Note: "*" significant at p<0.05, "**" p<0.01, and "***" p<0.001. Sources: CBDRB-FY19-422, CBDRB-FY19-448 and CBDRB-FY19-536.

Table 18b.

Avoidant Personality Disorder differential functioning of items and tests: Logistic Regression, Mantel-Haenszel (MH) Chi2 and Odds Ratios for ethnicity

Scale Item	**Native American Vs. Caucasian DIF**				**Hispanic vs. Caucasian DIF**			
	CHI2	**Sig.**		**Odds Ratio**	**CHI2**	**Sig.**		**Odds Ratio**
S10Q1A1	0.21	0.6496		0.90	0.02	0.8983		0.799*
S10Q1A2	0.33	0.5642		2.47	1.25	0.2639		1.730*
S10Q1A3	2.94	0.0865		0.17	4.78*	0.0288		0.975
S10Q1A4	0.00	0.9528		4.01*	6.09*	0.0136		0.764*
S10Q1A5	0.27	0.6011		0.00	0.93	0.3338		0.732*
S10Q1A6	0.02	0.8787		0.54	4.31*	0.0379		1.381*
S10Q1A7	0.15	0.6977		0.38	0.94	0.3315		1.198*
	Native American Vs. Caucasian DTF				**Hispanic vs. Caucasian DTF**			
Scale Score	**CHI2**	**Sig.**	**MH CHI2**	**Odds Ratio**	**CHI2**	**Sig.**	**MH CHI2**	**Odds Ratio**
AVODPDX2	3.88	0.0490	0.43	1.832	0.57	0.4518	1.22	0.779

Scale Item	**African American vs. Caucasian DIF**				**Asian American vs. Caucasian DIF**			
	CHI2	**Sig.**		**Odds Ratio**	**CHI2**	**Sig.**		**Odds Ratio**
S10Q1A1	0.04	0.8393		1.069	2.50	0.1140		0.811
S10Q1A2	0.27	0.6059		2.357*	0.50	0.4808		2.122*
S10Q1A3	22.64***	0.0000		1.355*	1.58	0.2090		0.670*
S10Q1A4	2.14	0.1431		0.528*	0.06	0.8035		0.677*
S10Q1A5	0.36	0.5482		0.757*	1.41	0.2353		0.843
S10Q1A6	0.02	0.9014		1.100	0.06	0.8093		1.365*
S10Q1A7	0.93	0.3360		0.700*	0.17	0.6821		1.016*
	African American vs. Caucasian DTF				**Asian American vs. Caucasian DTF**			
Scale Score	**CHI2**	**Sig.**	**MH CHI2**	**Odds Ratio**	**CHI2**	**Sig.**	**MH CHI2**	**Odds Ratio**
AVODPDX2	1.04	0.3085	2.89	0.700	0.38	0.5350	0.87	1.835

Note: "*" significant at p<0.05, "**" p<0.01, and "***" p<0.001. Sources: CBDRB-FY19-422, CBDRB-FY19-448 and CBDRB-FY19-536.

between ethnicities were not significant, however, statistics were not able to be reported for Native and Asian American populations due to privacy policies of Census. Significant comparisons may exist for these groups, but are not able to be examined. Differential item functioning was examined next, to determine if item bias had been introduced. The dependent scale contained eight items. Five items within the dependent scale had significant chi-squared for gender (see Table 19a). Interestingly, Census computers were unable to calculate a scale level comparison for differential test functioning analysis using a logistic regression model. MH Chi-square and odds ratios were not significant. Native and Asian American statistics were not available, as mentioned previously. When compared with Caucasians, Hispanic participants had significant differences in item functioning in five out of the eight items (Table 19b). African Americans compared with Caucasians had bias present in one of the items in the dependent scale. A common item between groups which contained differential item functioning was item S10Q1A9, “Depend on others to handle important areas in life.” To determine if these items were contributing to the overall score mean differences, test level functioning was examined next. Differential Test Functioning indicated no significant differences at the overall dependent scale level for any comparisons made. This indicates the that there appears to be no differential functioning, or bias, for this scale for either minority compared to Caucasians or gender. MH and ORs were examined to determine any possible remaining impact or effect upon the total score of participants. Both MH Chi-squares and ORs were nonsignificant for all groups tested. Overall, these results indicated no differential test functioning, or scale level bias.

Obsessive-Compulsive Personality Disorder scale was examined last. A significant ANOVA was found for ethnicity but not for gender (see Table 7). Pairwise comparisons (Table 17) indicated that mean scale differences between Caucasians with Hispanics and Asian

Table 19a.

Dependent Personality Disorder responses and differential functioning of items and tests: Logistic Regression, Mantel-Haenszel (MH) Chi2 and Odds Ratios for gender

Scale Item	**Male vs. Female DIF**			
	CHI2	**Sig.**		**Odds Ratio**
S10Q1A8: Need a lot of reassurance from others before making everyday decisions	2.62	0.1058		1.223*
S10Q1A9: Depend on others to handle important areas in life	33.62***	0.0000		0.732*
S10Q1A10: Find it hard to disagree with people even when think they are wrong, because fear losing their support or approval	4.87*	0.0273		1.472*
S10Q1A11: Find it hard to start or work on tasks when there is no one to help	8.53**	0.0035		1.077
S10Q1A12: Ever volunteered to do unpleasant things to get other people to like you	66.62***	0.0000		0.666*
S10Q1A13: Usually feel uncomfortable when alone because afraid can't take care of self	15.09***	0.0001		1.108
S10Q1A14: When close relationship ends, feel you have to immediately find someone else to take care of you	2.99	0.0836		0.585*
S10Q1A15: Worry a lot about being left alone to take care of self	0.81	0.3676		1.887*
	Male vs. Female DTF			
Scale Score	**CHI2**	**Sig.**	**MH CHI2**	**Odds Ratio**
DEPPDDX2: Dependent Personality Disorder	N/A	N/A	0.07	0.947

Note: "*" significant at p<0.05, "**" p<0.01, and "***" p<0.001. Sources: CBDRB-FY19-422, CBDRB-FY19-448 and CBDRB-FY19-536.

Table 19b.

Dependent Personality Disorder differential functioning of items and tests: Logistic Regression, Mantel-Haenszel (MH) Chi2 and Odds Ratios for ethnicity

Scale Item	**Native American Vs. Caucasian DIF**					**Hispanic vs. Caucasian DIF**			
	CHI2	**Sig.**		**Odds Ratio**		**CHI2**	**Sig.**		**Odds Ratio**
S10Q1A8	N/A	N/A		N/A		0.08	0.7761		1.674*
S10Q1A9	N/A	N/A		N/A		7.82**	0.0052		0.855*
S10Q1A10	N/A	N/A		N/A		18.78***	0.0000		0.706*
S10Q1A11	N/A	N/A		N/A		5.12*	0.0237		0.997
S10Q1A12	N/A	N/A		N/A		15.65***	0.0001		0.883
S10Q1A13	N/A	N/A		N/A		6.15*	0.0131		1.375*
S10Q1A14	N/A	N/A		N/A		0.14	0.7081		1.466*
S10Q1A15	N/A	N/A		N/A		0.30	0.5865		1.185*
	Native American Vs. Caucasian DTF					**Hispanic vs. Caucasian DTF**			
Scale Score	**CHI2**	**Sig.**	**MH CHI2**	**Odds Ratio**		**CHI2**	**Sig.**	**MH CHI2**	**Odds Ratio**
DEPPDDX2	N/A	N/A	N/A	N/A		2.43	0.1188	0.05	0.910

Scale Item	**African American vs. Caucasian DIF**					**Asian American vs. Caucasian DIF**			
	CHI2	**Sig.**		**Odds Ratio**		**CHI2**	**Sig.**		**Odds Ratio**
S10Q1A8	1.20	0.2729		1.257		N/A	N/A		N/A
S10Q1A9	1.60	0.2055		0.998		N/A	N/A		N/A
S10Q1A10	1.39	0.2388		0.825*		N/A	N/A		N/A
S10Q1A11	1.33	0.2489		1.245*		N/A	N/A		N/A
S10Q1A12	0.62	0.4311		0.757*		N/A	N/A		N/A
S10Q1A13	11.59***	0.0007		1.272		N/A	N/A		N/A
S10Q1A14	0.69	0.4059		1.182		N/A	N/A		N/A
S10Q1A15	0.29	0.5905		1.092		N/A	N/A		N/A
	African American vs. Caucasian DTF					**Asian American vs. Caucasian DTF**			
Scale Score	**CHI2**	**Sig.**	**MH CHI2**	**Odds Ratio**		**CHI2**	**Sig.**	**MH CHI2**	**Odds Ratio**
DEPPDDX2	2.78	0.0955	0.02	1.482		N/A	N/A	N/A	N/A

Note: N/A indicates that data was not available as less than 15 participants obtained "yes" scores for scaled scores for this disorder, as indicated in the U.S. Census Bureau's privacy policy. "*" significant at $p<0.05$, "**" $p<0.01$, and "***" $p<0.001$. Sources: CBDRB-FY19-422, CBDRB-FY19-448 and CBDRB-FY19-536.

Americans, with Caucasians responding more positively than their comparison groups. Native American participants also tended to respond yes more frequently than these two groups. There was no significant difference between Caucasians when compared to Native or African Americans. Differential item functioning was examined next, to determine if item bias had been introduced. The obsessive-compulsive scale contained ten items. Three items within the obsessive-compulsive scale had significant chi-squared for gender (see Table 20a). Native Americans had one item contributing bias when compared with Caucasians (see Table 20b). When compared with Caucasians, Hispanic participants had significant differences in item functioning in two out of the ten items. African Americans compared with Caucasians also had bias present in three of the items in the obsessive-compulsive scale. Differential functioning for items for comparisons between Caucasians and Asian Americans indicated bias within one of the items. A common item between groups which contained differential item functioning was item S10Q1A16, "The kind of person who focuses on details/order/organization or likes to make lists and schedules." To determine if these items were contributing to mean differences, test level functioning was examined next. Differential Test Functioning indicated no significant differences at the overall obsessive-compulsive scale level for any of the five comparisons made. This indicates the that there appears to be no differential functioning, or bias, for this scale for either minority compared to Caucasians or gender. MH and ORs were examined to determine any possible remaining impact or effect upon the total score of participants. Both MH Chi-squares and ORs were nonsignificant for all groups tested. Overall, these results indicated no differential test functioning, or scale level bias.

Overall, results indicated that significant item level bias existed on all measures of personality. These items may represent differences in the way that a disorder is presented in a

Table 20a.

Obsessive-Compulsive Personality Disorder responses and differential functioning of items and tests: Logistic Regression, Mantel-Haenszel (MH) Chi2 and Odds Ratios for gender

Scale Item	**Male vs. Female DIF**			
	CHI2	**Sig.**		**Odds Ratio**
S10Q1A16: The kind of person who focuses on details/order/organization or likes to make lists and schedules	8.28**	0.0040		1.338*
S10Q1A17: Sometimes get so caught up with details/schedules/organization that you lose sight of what wanted to accomplish	8.31**	0.0039		1.130*
S10Q1A18: Have trouble finishing jobs because spend so much time trying to get things exactly right	1.89	0.1690		0.851*
S10Q1A19: You or others feel you are so devoted to work/school you have no time left for anyone else or just having fun	1.48	0.2241		0.699*
S10Q1A20: Others think you have unreasonably high standards/morals/ideas about right and wrong	2.03	0.1545		1.139*
S10Q1A21: Have trouble throwing out worn-out/worthless things even if have no sentimental value	5.60*	0.0179		0.987
S10Q1A22: Hard to let others help if they don't agree to do things exactly the way you want	3.53	0.0601		1.384*
S10Q1A23: Hard to spend money on self/others even when have enough	0.96	0.3275		1.101*
S10Q1A24: Often so sure you are right that doesn't matter what others say	2.51	0.1131		0.694*
S10Q1A25: Have others told you that you are stubborn or rigid	1.89	0.1692		0.790*
	Male vs. Female DTF			
Scale Score	**CHI2**	**Sig.**	**MH CHI2**	**Odds Ratio**
OBCOMDX2: Obsessive-Compulsive Personality Disorder	0.01	0.9236	1.20	1.083

Note: "*" significant at p<0.05, "**" p<0.01, and "***" p<0.001 Sources: CBDRB-FY19-422, CBDRB-FY19-448 and CBDRB-FY19-536.

Table 20b.

Obsessive-Compulsive Personality Disorder differential functioning of items and tests: Logistic Regression, Mantel-Haenszel (MH) Chi2 and Odds Ratios for ethnicity

Scale Item	**Native American Vs. Caucasian DIF**				**Hispanic vs. Caucasian DIF**			
	CHI2	**Sig.**		**Odds Ratio**	**CHI2**	**Sig.**		**Odds Ratio**
S10Q1A16	0.15	0.6973		0.703*	47.38***	0.0000		0.798*
S10Q1A17	1.50	0.2207		0.887	0.15	0.6988		1.451*
S10Q1A18	0.00	0.9857		0.963	0.00	0.9490		1.208*
S10Q1A19	0.00	0.9732		1.094	1.29	0.2565		1.223*
S10Q1A20	0.58	0.4444		1.127	0.91	0.3407		1.066
S10Q1A21	0.13	0.7149		1.114	1.95	0.1626		0.718*
S10Q1A22	11.94***	0.0005		1.064	1.93	0.1646		0.973
S10Q1A23	0.28	0.5970		1.150	0.65	0.4194		1.040
S10Q1A24	1.57	0.2105		1.001	12.91***	0.0003		1.783*
S10Q1A25	1.90	0.1681		1.104	1.56	0.2120		0.794*
	Native American Vs. Caucasian DTF				**Hispanic vs. Caucasian DTF**			
Scale Score	**CHI2**	**Sig.**	**MH CHI2**	**Odds Ratio**	**CHI2**	**Sig.**	**MH CHI2**	**Odds Ratio**
OBCOMDX2	3.23	0.0724	0.05	0.935	0.30	0.5834	1.21	0.892

Scale Item	**African American vs. Caucasian DIF**				**Asian American vs. Caucasian DIF**			
	CHI2	**Sig.**		**Odds Ratio**	**CHI2**	**Sig.**		**Odds Ratio**
S10Q1A16	8.23**	0.0041		0.723*	18.78***	0.0000		1.242*
S10Q1A17	0.30	0.5825		1.007	0.27	0.6033		1.863*
S10Q1A18	0.55	0.4600		0.790*	0.00	0.9598		1.186
S10Q1A19	6.69	0.0097		1.199*	3.21	0.0731		1.341*
S10Q1A20	18.83***	0.0000		1.958*	1.65	0.1991		0.904
S10Q1A21	0.53	0.4677		0.922*	1.21	0.2722		0.585*
S10Q1A22	0.10	0.7466		0.925	1.15	0.2834		0.947
S10Q1A23	0.24	0.6224		0.919	0.30	0.5850		0.710*
S10Q1A24	0.73	0.3920		1.119*	0.39	0.5344		1.767*
S10Q1A25	5.72*	0.0168		0.926*	2.41	0.1206		0.685*
	African American vs. Caucasian DTF				**Asian American vs. Caucasian DTF**			
Scale Score	**CHI2**	**Sig.**	**MH CHI2**	**Odds Ratio**	**CHI2**	**Sig.**	**MH CHI2**	**Odds Ratio**
OBCOMDX2	0.00	0.9631	0.39	0.940	1.22	0.2700	0.22	0.879

Note: "*" significant at p<0.05, "**" p<0.01, and "***" p<0.001. Sources: CBDRB-FY19-422, CBDRB-FY19-448 and CBDRB-FY19-536.

population, or may represent bias in the items. Odds ratios for items give us the ability to determine the direction and impact of each of these items, and whether the items differences represent true differences in presentation of a disorder or item bias. In many cases, this bias was minimal in terms of impact, or was balanced with other items which were biased in the opposite direction. However, this was not always the case, and this led to test level bias.

Bias was observed in logistic regression differential test functioning (LR-DTF) with a notable impact (odds ratio; OR of 0.857) for males in the antisocial scale. This resulted in a 16.7% (see reversed OR, Table 21) increase in likelihood of having the disorder. Overall, men were observed to have an antisocial PD prevalence of 5.5%. Women had LR-DTF bias with a notable OR impact in the borderline (1.396) and narcissistic (1.239) scales, increasing the likelihood of diagnosis of these disorders by 39.6% and 23.9% respectively. Women had a prevalence of 6.18% for borderline PD, and 4.77% for narcissistic PD. These rates may be inflated due to the bias present. This is particularly interesting, since men had higher prevalence for narcissistic PD, at 7.69%. Additional possible areas of bias for women were identified in the paranoid and histrionic scales, with a significant Mantel-Haenszel (MH) DTF and notable OR (1.303 and 1.395) impact. These differences made women 30.3% and 39.5% more likely to have the diagnosis criteria met. Women had overall prevalence rates of 4.97% for paranoid PD and 1.81% for histrionic PD, which may be inflated due to the bias. It should again be noted that the MH test is less reliable for this data analysis, and as such, these results are noted as being possible for bias, but may also represent true group differences in these areas (women may have a true difference in the presentation or amount of people with paranoid or histrionic PDs).

Native American participants compared to Caucasians had significant LR-DTF for histrionic and borderline scales, however, only the borderline scale indicated a significant OR

Table 21.

Reversed Odds Ratios and significant findings

	Male vs Female		**Native American vs Caucasian**		**Hispanic vs Caucasian**		**African American vs Caucasian**		**Asian American vs Caucasian**	
Scale	**OR-F (M)**	**OR-R (F)**	**OR-F (N)**	**OR-R (C)**	**OR-F (H)**	**OR-R (C)**	**OR-F (Af)**	**OR-R (C)**	**OR-F (As)**	**OR-R (C)**
Avoidant	0.969	1.032	0.546	1.832	1.284	0.779	1.429	0.700	0.545	1.835
Dependent	1.056	0.947	N/A	N/A	1.010	0.910	0.675	1.482	N/A	N/A
Obsessive-Compulsive	0.923	1.083	1.070	0.935	1.121	0.892	1.064	0.940	1.138	0.879
Paranoid	0.767	1.303*,^	0.558	1.792	1.267	0.789#	1.718	0.582*,^,#	1.359	0.736
Schizoid	1.011	0.989	0.949	1.054	1.147	0.872	1.653	0.605*,^,#	2.041	0.490*,^
Histrionic	0.717	1.395*,^	0.965	1.036#	1.300	0.769	1.808	0.553*,^,#	1.531	0.653
Antisocial	1.167	0.857*,#	0.927	1.078	0.751	1.332*,^,#	1.109	0.902#	0.882	1.134#
Borderline	0.716	1.396*,^,#	1.040	0.962*,#	1.255	0.797	1.495	0.669*,^	0.744	1.344
Narcissistic	0.807	1.239*,^,#	0.951	1.051	1.420	0.704*,^,#	1.848	0.541*,^,#	1.770	0.565*,^
Schizotypal	0.890	1.124	1.590	0.629*	1.420	0.704*,^	1.647	0.607*,^,#	1.748	0.572

Note: OR-R is the odds ratio reported in DIF calculations, which focused on the reference group (Caucasian). OR-F indicates the recalculated odds ratio for the focal group (minority) for every unit increase in the Caucasian reference group. For gender, the OR-F indicates odds for the focal group (males), while OR-R indicates odds for the reference group (females). These indications are also in brackets for ease of reading. "*****" indicates that OR results were significant. "^" indicates that Mantel-Haenszel Chi-square was significant for DTF. "#" indicates that logistic regression Chi-square was significant for DTF. Sources: CBDRB-FY19-422, CBDRB-FY19-448 and CBDRB-FY19-536.

impact of bias. This impact was still small (0.962), and so Natives were only 4% more likely to have met the criteria. They had an overall prevalence of 11.7%. Of note, the schizotypal scale had a significant odds ratio of 0.629, with significant LR-DTF. This resulted in Native Americans being 59% more likely to have the diagnosis during this assessment. Native Americans had a prevalence rate of 6.72% for schizotypal, which was almost twice the prevalence rate across all participants. Bias may be considered to be influencing the overall prevalence of this disorder.

For Caucasians, LR-DTF was found with the histrionic scale when comparing to Native Americans, though this OR remained nonsignificant, indicating that despite bias being present at the scale level, Caucasians were only 3.6% more likely to have the disorder. Test bias was present for Caucasians when compared to both Hispanic and Asian American participants for the antisocial scale. Though the Asian American comparison did not have a significant OR (1.134) or MH, Caucasians were 13.4% more likely to have a positive score. The Hispanic comparison did have significant MH and OR (1.332), indicating a significant impact of bias, with Caucasians 33.2% more likely to have a positive scale score. Caucasians had an antisocial prevalence of 3.62%, of which test bias may account for some level. Interestingly, several odds ratios were quite large for Caucasians in this sample, despite lack of test bias. Caucasians were notably more likely than Native Americans and Asian Americans to have positive scale scores on the avoidant scale (ORs of 1.832, 1.835; 1.83 times or 83% more likely), and more likely than Native Americans to have a positive diagnosis on the paranoid scale (OR of 1.792). Given the lack of test bias present on those scales, those may be considered to be true differences in responding rates. Despite this, the paranoid scale had prevalence rates for Native Americans at 10.2%, while Caucasians had a rate of only 3.67%. This may be, in part, due to the extremely large sample size of Caucasian participants. Similar results were also seen in the ORs of the dependent scale

compared to African Americans (1.482) and borderline scale compared to Asian Americans (1.344), though Caucasians had a higher prevalence rate for both of these (0.5% and 5.6%).

Hispanic participants had significant LR-DTF on the narcissistic scale, with significant MH and OR (0.704), indicating the presence of test level bias, with 42% increased likelihood of having a positive score. They had an overall prevalence rate of 7.54%. They also had a significant MH and OR (also 0.704) for the schizotypal scale, but did not have a significant LR-DTF. This indicates a possible presence of bias impacting the prevalence rate of 3.9%, but again, is less reliable than the LR-DTF. It is also possible that this is not bias, and is instead detecting a true difference in the culture. The paranoid scale had a significant LR-DTF, however, this was not paired with a significant OR or MH. Despite it not being significant, it did still indicate that Hispanics were 26.7% more likely to have positive scores on the paranoid scale than Caucasians. It is also possible that this is due to test level bias. Three scales had notable, but not significant ORs, avoidant, borderline and histrionic. These scales had odds ratios from 1.255 to 1.300, indicating that they were 25 to 30% more likely to have positive scores on these scales. These results are considered to be true differences in the populations, given that no test level bias was detected.

Asian American participants had significant ORs and MH for the schizoid (OR=.490) and narcissistic (OR=0.565) scales when compared to Caucasians, but these did not have significant LR-DTF. This indicated possible bias in these areas, with 104.1% and 77.0% increased likelihood of positive results on these scales. Interestingly, despite the increased response tendency, Asian Americans had lower prevalence rates on both these scales than the average across all participants (1.56% and 5.2%, respectively). In fact, the schizoid prevalence was less than half of average. They also had several odds ratios which did not indicate significant bias,

but did indicate notable response differences. Paranoid, histrionic and schizotypal scales all had large ORs (0.572 to 0.736), corresponding to increased likelihood of positive scale scores from 35.9% to 74.8%.

Compared to Caucasians, African Americans were noted to have significant LR-DTF, MH and OR for the paranoid (OR=0.582; 71.8% more likely to have a positive score), schizoid (OR=0.605; 65.3% more likely), histrionic (OR=0.553; 80.8% more likely), narcissistic (OR=0.541; 84.8% more likely) and schizotypal (OR=0.607; 64.7% more likely) scales. African American prevalence rates were approximately twice the total of all participants for the narcissistic scale (12.4%), schizotypal scale (6.78%), and paranoid scale (7.62%). They also had higher schizoid (4.9%), histrionic (2.58%) and borderline (8.08%) than the average prevalence. Some of this difference is certainly attributable to bias. They had significant LR-DTF, but not OR or MH for the antisocial scale, indicating that while it contained bias, that the bias did not have a large impact on the results. It is also possible that the bias actually brought the ORs closer together. They had an overall prevalence rate for antisocial PD of 3.7%. An additional source of possible bias was on the borderline scale, which had a significant OR and MH, but not LR-DTF. This indicated that African Americans were 49.5% more likely to have met the borderline criteria. They had an overall prevalence for this disorder of 8.08%. A large but nonsignificant OR (0.700) was also seen on the avoidant scale, with no discernable bias present. This may be considered a true difference in the responding on this scale, with African Americans 42.9% more likely to have a positive total score.

To sum, many differences were found throughout the personality disorder criteria assessed. These differences may represent cultural disparities in the way that a disorder is presented in a population, or may represent bias in the items and scales.

Discussion

Differential item functioning was found throughout the AUDADIS-IV scale which the NESARC data is based upon. A majority of items for most scales did show significant chi-squared results, however, these items are not cause for concern. The statistical tests were significant because of the very large sample. Much of this bias was balanced out at the test level for the personality scales, but some remained. The avoidant, dependent and obsessive-compulsive scales may be used for both gender and ethnic comparisons, as they did not have overall test bias present. Considering the lack of test level bias, and the remaining differences in the means between groups on these scales, it can be assumed that these differences are attributable to true differences between the groups. Those scales with bias and notable impact present for gender included the antisocial scale for men, and the borderline and narcissistic scales for women. Scales with bias and notable impact for ethnicity included schizotypal and borderline for Native Americans, antisocial for Caucasians (when compared to Hispanics), narcissistic for Hispanics, and paranoid, schizoid, narcissistic and schizotypal for African Americans. These scales should be interpreted with caution with these groups.

The original hypotheses included one of disparities in functioning for the dependent scale for Hispanic, Asian American, and Native American participants. Unfortunately, due to Census privacy restrictions, data was not available for the Native and Asian American populations. No difference in test level functioning was observed for Hispanic participants compared to Caucasians. It is unclear if test level bias would have been present at this time. It was also hypothesized that African American groups would have disparities in their response rates for the paranoid scale. This was confirmed throughout our analyses. It was anticipated that cultures with spiritual beliefs outside of Western Christianity would differ on the schizotypal items. This

certainly was the case. However, this went beyond the anticipated few items related to unusual or magical beliefs. Both African American and Native American groups had significant, meaningful test bias present. Overall, the significant difference in African American functioning throughout the NESARC data's AUDADIS-IV scale was surprising, with differential test functioning present for this minority on five of the ten scales. Prevalence rates, assessment, and general knowledge about how this minority is perceived in relation to personality disorder presentation and diagnosis should be examined for individual items and extreme caution should be taken when reviewing results. Clinicians should be aware of these differences during interview as well.

Many clinicians rely upon objective measures when assessing for a personality disorder or other mental health concern. Given the distinct possibility of test items and scales introducing bias which can influence test results for minority populations, or for a specific gender, Martinkova et al. (2017) suggested that differential item functioning should be completed for all measures during development. This concern for bias towards certain groups has been seen throughout personality assessment, such as the MMPI-2, as well as the national study examined here. This is highly concerning for those who use tools such as the MMPI-2 or the DSM5, which do not indicate specific differences in minorities for reporting rates or specific diagnostic criteria for which to watch for examiner bias.

Many clinicians are underinformed of minority client spiritual beliefs and cultural practices. This lack of understanding has been seen in research into Native American mental health care. Thomason (1999) identified many sources of bias within the administration and interpretation of standardized tests. While his work pertained to Native Americans, it certainly holds true for other minorities. These sources included: the test may not be designed to yield valid information with minorities; the examiner may be biased/not knowledgeable of cultural

differences of the minority (for example, a client who avoids eye contact, displays little emotion, is not talkative, and behaves modestly may be displaying respect/humility rather than resistance); testing may be foreign to traditional cultures, and classifying people on quantitative scales may be contrary to their values (for example, equality and the emphasis on the group over the individual or clients from remote rural areas may need an orientation to what testing is and how it can help); testing may be biased if it emphasizes factors that conflict with basic cultural values (such as timed tests penalizing clients who are not accustomed to rushing through a task); tests may be used improperly (such as performing an assessment without first gauging the client's acculturation level and reading level, or making invalid interpretations because of important cultural factors); consumers of reports may not understand them and may make inaccurate judgments about the client, especially if they are unaware of the cultural differences involved.

A thorough examination of their cultural and spiritual beliefs should be undertaken prior to diagnosing a minority population (see Hodge & Limb, 2010 and 2011). A review completed by Monod et al. in 2011 reviewed several of these measures. Minority clients undergoing personality assessment should ensure that clinicians are cognizant of any spiritual beliefs, rituals or cultural differences compared to more European American traditions and culture. Acculturation should also be assessed, to determine the use of a measure’s appropriateness and impact of any cultural differences that a minority client may have compared to the standardization sample. Without these considerations, many minority clients may be assumed to meet criteria for a personality disorder item, and may meet the overall criteria for a disorder when it is not otherwise indicated. It is the hope of this study that clinicians should be highly cautious when diagnosing minority clientele, and to take into consideration that men and women

or minority clients may differ on how they respond to questions related to personality disorder diagnosis.

A review of items and understanding of clients should also be undertaken. In some cases, language barriers may be present. If a client's primary language is not English, or if they may not have been exposed to certain ideas or words, the client may have a different understanding of the questions being asked. Preliminary analyses in test construction should assess for consistent understanding of items, regardless of a participant's background. Thomason (1999) completed a very thorough review of sources of bias and accommodations for Native Americans which clinicians should be made aware of when working with minority clients or when assessing disorders that have been discovered as having bias within the design. Issues that pertain to those areas include: ensuring that the examiner avoid confounding culture and personality, recognizing examiner bias (e.g., stereotyping, Western world view, not knowledgeable about the minority), acknowledging that the major personality theories were developed by Caucasian males and are not necessarily applicable across cultures, recognizing that standard interpretations of the results of personality tests cannot be applied to other cultural groups, understanding that translation of materials are not always accurate and can be difficult because meanings are not always retained (e.g., there is no word for "if" in the Hopi language, and no "if-then" linguistic structure in the Navajo language), and recognizing culture-specific response sets (e.g., the "Yes" set of many people in collectivist cultures, cultural differences in self-disclosure, the length of time taken to answer a question) may influence results or flag as a response set on validity indices.

Limitations of the study come from the nature of the data set used. The test construction fails to take into consideration the importance of significant distress or impairment of the individual answering the items. The DSM stipulates the requirement of some items causing

significant issues within the society, and this requirement is not properly addressed within the AUDADIS-IV. Although the scale score does require that one item contain an indicator of significant distress or impairment in functioning, the DSM criteria clearly calls for more than one item to contain this.

Future analyses should require a more stringent rule on more than one item containing significant stress to the individual. Also, follow-up analyses on adjusting prevalence rates to account for bias within the diagnosis of the disorders should be undertaken, to examine the true impact these psychological disorders can have on society.

References

Allen, J. (2002). Assessment training for practice in American Indian and Alaskan Native settings. *Journal of Personality Assessment, 79*, 216-225.

American Psychiatric Association: *Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition.* Arlington, VA, American Psychiatric Association, 2013.

Arbisi, P., Ben-Porath, Y. & McNulty, J. (2002). A comparison of MMPI-2 validity in African American and Caucasian psychiatric inpatients. Psychological Assessment, 14, 3-15.

Ascoli, M., Lee, T., Warfa, N., Mairura, J., Persaud, A., & Bhui, K. (2011). Race, culture, ethnicity and personality disorder: Group Careif position paper. *World Cultural Psychiatry Research Review, 6, 52-60.*

Asnaani, A., Richey, J.A., Dimaite, R., Hinton, D.E., & Hofmann, S.G. (2010). A cross-ethnic comparison of lifetime prevalence rates of anxiety disorders. *J Nerv Ment Dis., 198(8), 551–555*.

Balsis, S., Woods, C. M., Gleason, M. E. J., & Oltmanns, T. F. (2007). Overdiagnosis and underdiagnosis of personality disorders in older adults. *American Journal of Geriatric Psychiatry, 15(9),* 742-753.

Butcher, J., Cheung, F & Lim, J. (2003). Use of the MMPI-2 with Asian populations. *Psychological Assessment, 15*, 248-256.

C'de Baca,J., Castillo, D.T., Mackaronis, J.E., & Qualls, C. (2014). Ethnic differences in personality disorder patterns among women veterans diagnosed with PTSD. *Behav. Sci., 4, 72-86.*

Castro, Y., Gordon, K., Brown, J., Anestis, J & Joiner, T. (2008). Examination of racial differences on the MMPI-2 clinical and restructured clinical scales in an outpatient sample. *Assessment, 15*, 277-286.

Costa, P., Terracciano, A. & McCrae, R. (2001). Gender differences in personality traits across cultures: Robust and surprising findings. *Journal of Personality and Social Psychology, 81*, 322-331.

Cox, B. J., Sareen, J., Enns, M. W., Clara, I., & Grant, B. F. (2007). The fundamental structure of Axis II personality disorders assessed in the National Epidemiologic Survey on Alcohol and Related Conditions. *Journal of Clinical Psychiatry, 68,* 1913-1920.

Dana, R. (1988). Culturally diverse groups and the MMPI interpretation. *Professional Psychology: Research and Practice, 19,* 490-495.

Dana, R., Hornby, R & Hoffman, T. (1983). *Personality assessment of Rosebud Sioux: A comparison of Rorschach, Millon Clinical Multiaxial Inventory, and 16PF reports*. U.S. Department of Education.

Dana, R. & Whatley, P. (1991). When does a difference make a difference? MMPI scores and African-Americans. *Journal of Clinical Psychology, 47*, 400-406.

Elosua, P. & Wells, C. (2013). Detecting DIF in Polytomous Items Using MACS, IRT and Ordinal Logistic Regression. *Psicológica, 34,* 327-342.

Grant, B. F., & Dawson, D. A. (2006). Introduction to the National Epidemiologic Survey on Alcohol and Related Conditions. *Alcohol Research & Health, 29(2)*, 74-78.

Hill, J., Pace, T. & Robbins, R. (2010). Decolonizing personality assessment and honoring indigenous voices: A critical examination of the MMPI-2. *Cultural Diversity and Ethnic Minority Psychology,16*, 16–25.

Glass, M., Bieber, S., & Tkachuk, M. (1996). Personality styles and dynamics of Alaskan Native and nonnative incarcerated men. Journal of Personality Assessment, 66, 583-603.

Gray, J. S., Brionez, J., Petros, T., & Gonzaga, K. T. (2018, January 18). Psychometric evaluation of depression measures with Northern Plains Indians. *American Journal of Orthopsychiatry. Advance online publication*. http://dx.doi.org/10.1037/ort0000309

Hall, G., Bansal, A. & Lopez, I. (1999). Ethnicity and psychopathology: A meta-analytic review of 31 years of comparative MMPI/MMPI-2 research. *Psychological Assessment, 11,* 186-197.

Hidalgo, M. D. & Lopez-Pina, J. (2004). Differential item functioning detection and effect size: A comparison between logistic regression and Mantel-Haenszel procedures. *Educational and Psychological Measurement, 64,* 903-915.

Hodge, D. & Limb, G. (2010). A Native American perspective on spiritual assessment: The strengths and limitations of a complementary set of assessment tools. *Health & Social Work, 35(2),* 121-131.

Hodge, D. & Limb, G. (2011). Spiritual assessment and Native Americans: Establishing the social validity of a complementary set of assessment tools. *Social Work,* 213-223.

Lewis-Fernandez, R. & Kleinman, A. (1994). Culture, personality, and psychopathology. *Journal of Abnormal Psychology, 103*, 67-71.

Malgady, R., Rogler, L & Constantino, G. (1987). Ethnocultural and linguistic bias in mental health evaluation of Hispanics. *American Psychologist, 42*, 228-234.

Martinkova, P., Drabinova, A., Liaw, Y., Sanders, E., McFarland, J & Price, R. (2017). Checking equity: Why differential item functioning analysis should be a routine part of developing conceptual assessments. *Cell Biology Education CBE Life Sciences Education, 16,* 1-13.

McGilloway, A., Hall, R.E., Lee, T. & Bhui, K.S. (2010). Research article review of personality disorder, race and ethnicity: prevalence, aetiology and treatment. *BMC Psychiatry, 10, 33*.

McNulty, J., Graham, J., Ben-Porath, Y. & Stein, L. (1997). Comparative validity of the MMPI-2 scores of African American and Caucasian mental health center clients. *Psychological Assessment, 9*, 464-470.

Millsap, R. & Everson, H. (1993). Methodology review: Statistical approaches for assessing measurement bias. *Applied Psychological Measurement, 17,* 297-334.

Monahan, P., Mchorney, C., Stump, T. & Perkins, A. (2007). Odds ratio, delta, ETF classification and standardization measures of DIF magnitude for binary logistic regression. *Journal of Educational and Behavioral Statistics, 32,* 92-109.

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

Monod, S. Brennan, M., Rochat, E., Martin, E., Rochat, S. & Bula, C. (2011). Instruments measuring spirituality in clinical research: A systematic review. *J Gen Intern Med, 26,* 1345–1357.

Morizot, J., Ainsworth, A. T., & Reise, S. P. (2007). Toward modern psychometrics: Application of item response theory models in personality research. In Robins, R. W., Fraley, R. C., & Krueger, R. F. (Eds.), *Handbook of research methods in personality psychology* (pp. 407-423). New York, NY, US: Guilford Press.

Oshima, T.C., & Morris, S.B. (2008). Raju's differential functioning of items and tests (DFIT). *Educational Measurement: Issues and Practice, 3*, 43-50.

Pace, T., Robbins, R., Choney, S., Hill, J., Lacey, K., & Blair ,G.(2006) A cultural-contextual perspective on the validity of the MMPI-2 with American Indians. *Cultural Diversity and Ethnic Minority Psychology, 12,* 320 –333.

Pignolo, C., Fulcheri, M., Mazzeschi, C., Nuovo, S., Lis, A. & Zennaro, A. (2018). Psychometric properties of the Italian Version of the Personality Assessment Inventory (PAI). *Psychological Assessment, 30*, 1226–1236.

Pulay, A.J., Stinson, F.S., Dawson, D.A., Goldstein, R.B., Chou, S.P., Huang, B., … Grant, B.F. (2009). Prevalence, correlates, disability, and comorbidity of DSM-IV schizotypal personality disorder: Results from the Wave 2 National Epidemiologic Survey on Alcohol and Related Conditions. *Prim Care Companion J Clin Psychiatry,11(2),53–67*.

Raju, N., Laffitte, L. & Byrne, B. (2002). Measurement equivalence: A comparison of methods based on confirmatory factor analysis and item response theory. *Journal of Applied Psychology, 87*, 517–529

Ryder, A., Sun, J., Dere, J & Fung, K. (2014). Personality disorders in Asians: Summary, and a call for cultural research. *Asian Journal of Psychiatry, 7,* 86-88.

Samuel, D., Simms, L., Clark, L. A., Livesley, W. J., & Widiger, T. A. (2010). An item response theory integration of normal and abnormal personality scales. *Personality Disorders: Theory, Research and Treatment, 1(1)*, 5-21.

Sharp, C., Goodyer, I. & Croudace, T. (2006). The short mood and feelings questionnaire (SMFQ): A unidimensional item response theory and categorical data factor analysis of self-report ratings from a community sample of 7-through 11-year-old children. *Journal of Abnormal Child Psychology, 34*, 379–391.

Stover, J., Solano, A. & Liporace, M. (2015). Personality Assessment Inventory: Psychometric analysis of its Argentinean Version. *Psychological Reports: Measures & Statistics, 117,* 799-823.

Sue, S. (1996). Measurement, testing, and ethnic bias: Can solutions be found? *Multicultural Assessment in Counseling and Clinical Psychology, 4.*

Thomason, T. (1999). *Psychological and vocational assessment of Native Americans*. Flagstaff, Arizona: National Institute on Disability and Rehabilitation Research, U.S. Department of Education . ISBN 1-888557-83-4

Timbrook, R. & Graham, J. (1994). Ethnic differences on the MMPI-2? *Psychological Assessment, 6*, 212-217.

Triandis, H. & Suh, E. Cultural influences on personality. Annu. Rev. Psychol., 53, 133-160.

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.

Velasquez, R., Callahan, W. & Young, R. (1993). Hispanic-White MMPI comparisons: Does psychiatric diagnosis make a difference? *Journal of Clinical Psychology, 49*, 528-534.

Vesga-López, O., Schneier, F., Wang, S., Heimberg, R., Liu, S., Hasin, D.S., & Blanco, C. (2008). Gender differences in generalized anxiety disorder: Results from the National Epidemiologic Survey on Alcohol and Related Conditions (NESARC). *J Clin Psychiatry, 69(10), 1606–1616.*

Waller, N. G., Thompson, J. S., & Wenk, E. (2000). Using IRT to separate measurement bias from true group differences on homogeneous and heterogeneous scales: An illustration with the MMPI. *Psychological Methods, 5(1),* 125-146.

Zumbo, B. D. (1999). A handbook on the theory and methods of differential item functioning (DIF): Logistic regression modeling as a unitary framework for binary and likert-type (ordinal) item scores. Ottawa, ON: Directorate of Human Resources Research and Evaluation, Department of National Defense.