



Published in final edited form as:

Psychol Assess. 2023 September ; 35(9): 721–728. doi:10.1037/pas0001255.

Assessing the Measurement Invariance of the Personality Inventory for *DSM-5* across Black and White Americans

Colin D. Freilich¹, Isabella M. Palumbo², Robert D. Latzman³, Robert F. Krueger¹

¹Department of Psychology, University of Minnesota

²Department of Psychology, Georgia State University

³Takeda Pharmaceuticals

Abstract

The Personality Inventory for *DSM-5* (PID-5) is the primary tool for assessing maladaptive personality traits within the *DSM-5* Alternative Model for Personality Disorders (AMPD). Evidence has begun to accumulate on the replicability and measurement invariance of its five-domain factor structure across countries, clinical and community populations, and sex, but its equivalency across racial groups within a given country is largely unstudied. Attempting to replicate the evidence of non-invariance demonstrated by Bagby et al. (2022), we examined the factor structure of the PID-5 across White Americans (WA; $n = 612$) and Black Americans (BA; $n = 613$) within the U.S. The five-domain structure emerged across both samples with reasonably congruent factor loadings. Therefore, we tested for measurement invariance using the 13-step framework advocated by Marsh et al. (2009) for personality data. We found support for the PID-5's comparability across racial groups, offering some preliminary backing for its use with Black Americans, though additional evidence is needed to clarify the conflicting results and further validate the instrument.

Keywords

Personality Disorder; PID-5; AMPD; Measurement Invariance; Race

In formal diagnostic manuals, personality disorders (PDs) have traditionally been conceptualized as categorical syndromes, determined by polythetic symptom lists. Limitations of this approach, including frequent PD comorbidity, heterogeneity within syndromes, and lack of treatment specificity (Kupfer et al., 2002; Tyrer et al., 2007) led to the development of the Alternative Model for Personality Disorders (AMPD; Krueger et al., 2012). The AMPD was included in Section III (Emerging Models and Measures) of *the Diagnostic and Statistical Manual of Mental Disorders (DSM-5; American Psychiatric*

Correspondence concerning this article, including information about the data collected or code for the analyses run can be sent to Colin Freilich at freil016@umn.edu.

Hypotheses and design were not preregistered. C.D.F. had a lead role in conceptualization, methodology, formal analysis, and writing (original draft). I.M.P. had lead role in data curation and writing (review and editing) and a supporting role in conceptualization. R.D.L. had a lead role in funding acquisition and investigation, and a supporting role in conceptualization, supervision, and writing (review and editing). R.F.K. had a lead role in conceptualization, project administration, and supervision, and a supporting role in methodology and writing (review and editing).

Association [APA], 2013) and defines PDs as clinically significant difficulties in personality functioning expressed in high levels of at least one maladaptive personality trait.

Consequently, maladaptive personality traits are the descriptive core of the AMPD and the *Personality Inventory for DSM-5* (PID-5) was developed alongside the AMPD to measure said traits. In particular, the PID-5 measures 25 lower-order trait facets and defines an algorithm for averaging a subset of the scores to index the five higher-order domains of Negative Affectivity, Detachment, Antagonism, Disinhibition, and Psychoticism. The scoring algorithm was informed by the empirically derived hierarchical structure of the traits, with the 25 lower-order facets loading with varying weights onto five higher-order domains. This structure was initially derived using psychometric approaches involving item response theory and exploratory factor analysis in three U.S. samples that were constructed to be representative of the U.S. population in terms of race and other major demographic variables (including e.g., ~71% White Americans [WA] and ~12% Black Americans [BA]; Krueger et al., 2012).

PID-5 Factor Structure Replicability

The original structure proposed by Krueger et al. (2012) has been replicated across numerous samples (meta-analyses: Somma et al., 2019; Watters & Bagby, 2018; reviews: Al-Dajani et al., 2016; Barchi-Ferreira Bel & Osório, 2020; Freilich et al., 2023; Zimmerman et al., 2019), with Tucker's congruence coefficient (TCC; ϕ) commonly used as a measure of factor loading similarity, ranging from -1 to $+1$. Lorenzo-Seva and ten Berge (2006) conducted Monte Carlo analyses to determine reasonable thresholds for TCC values, suggesting values of 0.95 or higher indicate the two factors can be considered equal or have "good" similarity and values between 0.85 and 0.94 are indicative of "fair" similarity. For example, in their meta-analysis, Somma et al. (2019) demonstrate that the factor loadings (of the facets onto the five theoretical domains) derived in 25 new samples had at least fair (i.e., $\phi \geq 0.85$) congruence with those of the original Krueger et al. (2012) sample in almost all cases (96%).

These 25 samples spanned 11 different language translations of the PID-5, primarily in Western Europe, as well as diverse samples within the U.S. regarding age, sex, and geography. Similar evidence of the replicability of the five-domain structure has accumulated since, including in Egyptian (Aboul-ata & Qonsua, 2021) and Singaporean (Lim et al., 2019) samples. Thus, the structure of the PID-5 appears to be replicable across U.S. and European populations, with burgeoning evidence in non-Western countries and languages. However, we are unaware of any evidence of replicability (in relation to previously published structures) in a primarily BA sample, with Bagby et al. (2022) demonstrating that the five-factor structure did not emerge with reasonable congruence in either of their two BA samples.

Lack of congruence in a novel sample may result from substantive differences in personality structure or from sampling error. To that end, Somma et al. (2019) examined five potential moderators of factor loading congruence across studies. With limited variability (i.e., most TCC values were relatively high) and only 25 previous studies to consider, no significant

moderators were identified. However, despite not being statistically significant, sample size emerged as potentially relevant. Correlations were estimated between a study's sample size and the congruence values for each domain within that study (e.g., TCC between a given study's Negative Affect loadings and the original, Krueger et al. [2012] Negative Affect loadings). We argue that sample size was potentially relevant because the median correlation coefficient was 0.21 (a larger effect than other tested moderators), indicating that larger sample size may be moderately associated with greater congruence, and, conversely, that obtaining adequate congruence may be less likely without an adequate sample size.

PID-5 Measurement Invariance

Adequate congruence of factor loadings suggests replicability of factor structure, but it does not imply that the substantive test scores can be compared across groups. For that purpose, tests of measurement invariance are needed (Meredith & Teresi, 2006). Replicability of uniquely estimated factor structures is an indicator of configural invariance, and further levels of invariance are typically tested by comparing a sequence of increasingly restricted models. For instance, restricting factor loadings to equality across groups is a test of metric or "weak" invariance, while an additional constraint on indicator intercepts across groups serves as a test of scalar or "strong" invariance. Evidence of weak invariance suggests that the latent factor has the same meaning across groups, as it is defined by the same indicators to same extent. Evidence of strong invariance suggests that group mean differences in indicators are attributable to differences in the latent construct, as item intercepts or thresholds are reasonably equivalent across groups.

The measurement invariance properties of the PID-5 have been tested across several relevant domains. For instance, tests of cross-cultural measurement invariance provide support for the instrument's use in Europe. Thimm et al. (2017) compared Norwegian and U.S. university samples, finding weak and partial strong (i.e., scores were invariant across groups after two facet intercepts were released from constraints) measurement invariance. Sorrel et al. (2021) built on this work, finding evidence of weak and partial strong invariance across samples in Belgium, Switzerland, France, Spain, and Catalonia, suggesting the PID-5 factor structure is stable across countries and languages.

Given that heterogeneity within a given country can be substantial across, for example, culture, race, language, and geographic area, it is also important to consider measurement invariance across diverse within-country populations. Evidence of strong invariance of the five-factor structure across community and clinical samples has been demonstrated in at least three non-U.S. samples (Bach et al., 2018; Riegel et al., 2018; Somma et al., 2019b) as well as in an additional non-US sample using a modified six-factor structure (Zhang et al., 2021). Invariance across biological sex has similarly been demonstrated in American adult (Suzuki et al., 2019), Norwegian adult (South et al., 2017), Australian adult (Gomez et al., 2022), and Italian adolescent (Somma et al., 2017) samples, although only weak invariance was demonstrated in the latter two. In addition, strong invariance has been demonstrated across a heterosexual group and a mixed homosexual and bisexual group in a U.S. sample (Russell et al., 2017). Finally, examinations of differential item functioning across younger and older age groups have raised questions about the age-neutrality of certain items, though

many items have not displayed significant differential functioning (Debast et al., 2018; Van den Broeck et al., 2013). Thus, the limited initial evidence suggests that the PID-5 is largely invariant across age, sex, sexual orientation, sample type, and country.

However, invariance across racial groups within the same country is less studied. Bagby et al. (2022) found evidence suggesting a lack of configural invariance across WA and BA groups in the U.S., interpreting a single-factor solution for the BA group, suggestive of an undifferentiated, broadly based level of demoralization. On the other hand, Becker et al. (2022) examined the PID-5 across a WA and non-white group (26% Black, 44% Hispanic, 11% Asian, 19% other), finding evidence of strong invariance. Clearly, more work is needed to examine the suitability of the instrument across race in the U.S., particularly for BA populations. We attempted to begin to address this gap in the literature by examining the extent to which the PID-5 structure is consistent across WA and BA samples drawn from the same population. We first examined the factor structure replicability in WA and BA groups via TCC values and next examined the instrument's measurement invariance properties.

Method

Participants

Participants included 1,737 undergraduates at a large public Southeastern U.S. university. Participants were recruited through the university's research study pool and were awarded course credit for participation. The university's ethics review board approved all study procedures. Approximately 35.3% of participants identified as African American/Black, 35.2% as Caucasian/White, and 15.2% as Asian/Asian American. Because analyses focused on the BA and WA groups, we removed all other participants from the dataset, yielding a sample size of 1,225 (613 BA and 612 WA). The sample's ensuing mean age was 21.31 years with a standard deviation of 5.44. The sample was 76.3% female.

Measures

Personality Inventory for DSM-5 (PID-5)—The PID-5 (Krueger et al., 2012) is a 220-item self-report scale that measures maladaptive personality traits. In particular, it assesses the 25 specific facets and five, broader domains of personality outlined in the AMPD. The items are scored on a 4-point scale ranging from 0 (Very False or Often False) to 3 (Very True or Often True). Four to fourteen items are averaged to index each facet. A subset of three salient facets can be summed to index each domain, but, in this study, domain scores were latent variables estimated in exploratory structural equation models (ESEMs). The strength of the psychometric properties and construct validity of the PID-5 traits have been supported across many studies (Al-Dajani et al., 2016; Barchi-Ferreira Bel & Osório, 2020). Information about the data collected or code for the analyses run is available from the first author. Hypotheses and design were not preregistered.

Statistical Analyses and Results

Descriptive Statistics

The descriptive statistics for the PID-5 facets are displayed in Supplemental Table S1. Significant mean differences across the BA and WA groups were observed for eight of the 25 facets. The BA group scored higher on Grandiosity and Suspiciousness, whereas the WA group scored higher on Anxiousness, Distractibility, Eccentricity, Perseveration, Separation Anxiety, and Submissiveness. Cohen's d effect sizes were calculated for each of these comparisons. Apart from Submissiveness ($d = 0.52$), each of these effects can be considered small ($d < 0.30$).

Configural Congruence

All models were fit using maximum likelihood estimation with robust standard errors and target rotation in Mplus 8 (Muthén & Muthén, 1998–2022). We first estimated separate five-factor ESEM models for the BA and WA groups. We next estimated a combined model for the whole sample, followed by a two-group model of configural invariance. The fit indices for each of these models are displayed in Supplemental Table S2. Model fit indices were in acceptable ranges and comparable to prior ESEM estimates (e.g., Bach et al., 2018; Bagby et al., 2022; Thimm et al., 2017). Therefore, we next compared the equivalency of the loading matrices (of the facets onto latent domains) across groups as a test of configural invariance using TCCs.

Though sometimes used for congruence analyses, statistical rotations optimize factor solutions to a given sample, so heterogeneity in loadings resulting from sampling variability may be misunderstood as poor factor structure replicability (McCrae et al., 1996; Rolland, 2002). Targeted rotations may represent a viable solution to this problem, and McCrae et al. (1996) provide evidence that targeted rotations are appropriate when using a stringent threshold for factor congruency. Therefore, the loadings were target rotated toward the meta-analytically derived loading matrix reported by Somma et al. (2019; $n = 24,240$).

The resulting loading matrices are summarized in Table 1. We first compared the loadings across the BA and WA groups in the configural model. The TCCs between the BA and WA groups ranged from 0.87 to 0.97 (Negative Affect = 0.97; Detachment = 0.97; Antagonism = 0.96; Psychoticism = 0.92; Disinhibition = 0.87), indicative of fair to good congruence for all domains. Then, we compared loadings from each group (WA and BA) to the Somma et al. (2019) meta-analytically derived loadings¹. The TCCs ranged from 0.86 to 0.98 for the WA group and from 0.88 to 0.99 for the BA group. Each TCC value from these comparisons is also displayed in Table 1.

Measurement Invariance

Because configural invariance was upheld, we executed the remaining levels of measurement invariance testing from the 13-step framework advocated by Marsh et al.

¹The Somma et al. (2019) meta-analytic loadings are nearly identical to the original Krueger et al. (2012) loadings ($\phi = 0.96$ – 0.99 for domains), so comparing to either should not produce meaningfully different results.

(2009) for ESEM analysis of personality data. Results for each step are displayed in Table 2, varying from the least restrictive model of configural invariance to a model of complete factorial invariance. The weak invariance step constrains factor loadings to equality across the WA and BA groups. The strong invariance step additionally constrains facet intercepts across groups, while the strict invariance step additionally constrains facet residual variances (i.e., uniquenesses). There are also supplemental steps that involve the constraint of factor variances and covariances and latent factor means. The most restrictive step (13) represents complete factorial invariance in which each of these values are constrained to equality.

As suggested by Marsh et al. (2009) and Chen (2007), we used differences in CFI and RMSEA as indicators of measurement invariance. To establish invariance across the two samples, neither of those two fit indexes should differ more than 0.010 in comparison to their respective baselines. For additional context, we included Akaike information criteria (AIC) and Bayesian information criteria (BIC), as measures of comparative fit that do not require nesting and are relatively sensitive to scalar noninvariance (Cao & Liang, 2022). We did not set formal criteria for changes in information criteria that would constitute noninvariance.

As indicated in Table 2, measurement invariance was established at each step with the exception of strong invariance (step 5). Relative to the weak invariance models (step 2), there was a decrease in CFI of 0.011, above the 0.010 threshold, but RMSEA increased by only 0.003. No other steps created large changes in RMSEA (< 0.002) or CFI (< 0.010). The strong invariance step also created the largest increases in model information criteria.

As a result of the decreases in model fit at the strong invariance step, we examined modification indices (MI) for the intercepts. Only the intercepts for Submissiveness and Suspiciousness had MIs with values over 20. Therefore, as a test of partial strong invariance, we released these two intercepts from the constraint of equality across groups, while the other 23 remained constrained. This resulted in acceptable levels of invariance (CFI = 0.05). Subsequently, these intercepts were released from the remaining invariance testing procedure (step 7 and onward), and, as before, each of the steps had acceptable differences in CFI, RMSEA, and information criteria. The entire measurement invariance testing sequence with these two intercepts released is summarized in Supplemental Table S3. Taken together, PID-5 scores could be appropriately compared across WA and BA groups in terms of observed factor/domain means and variances. Item/facet residual variance, and, apart from those of Suspiciousness and Submissiveness, item/facet means could also be appropriately compared.

Discussion

We used ESEM to model the latent domains of the PID-5 separately in WA and BA groups and found evidence that the factor structure was replicable with fair to good congruency. Further, we tested for measurement invariance between the two groups. Weak invariance was supported, indicating that the domains hold generally comparable meanings across groups. Strong invariance was partially supported, indicating that mean differences in facet scores are largely comparable, with the possible exception of the Suspiciousness and

Submissiveness facets. While these two facets require further investigation to determine if any items are inapplicable across populations, evidence of partial strong to strong invariance is in line with the degree of invariance observed across countries (Sorrel et al., 2021; Thimm et al., 2017), sex (South et al., 2017; Suzuki et al., 2019), and community versus clinical populations (Bach et al., 2018; Somma et al., 2019b), providing preliminary support for the comparability of PID-5 scores across Black and White Americans.

On the other hand, these results are not consistent with the lack of configural invariance across WA and BA groups observed by Bagby et al. (2022). Bagby and colleagues offer two plausible interpretations for their results: that 1) personality structure as measured by the PID-5 is not similar across racial groups or that 2) the PID-5 is susceptible to considerable sampling error due to the interstitial nature of the facets, that is, that single facets can significantly load onto multiple domains, leading to instability in structure. Our results are not consistent with the first interpretation of a substantive difference in personality structure, lending support to the notion that the complex interstitial structure of the PID-5 may lead to some degree of structural instability across samples.

Sample size emerged as a potentially relevant moderator of factor congruency in the Somma et al. (2019) meta-analysis, suggesting that larger sample size may be moderately associated with greater congruence. In other words, it may be difficult to reliably model the complex PID-5 structure without an adequate number of observations. In the Bagby et al. (2022) BA derivation sample ($n = 255$), congruence with the original Krueger et al. (2012) structure was at least fair for four of five domains², but was inadequate for all five domains in the BA replication sample ($n = 303$). With a larger sample ($n = 613$), we observed adequate to good congruence across each of the five domains. This does not indicate that sample size is necessarily responsible for the conflicting results, though future work could explore associations between TCC values and sample size across studies (especially given that many additional samples have been generated since the Somma et al. [2019] meta-analysis). In addition, researchers could use simulation methods to estimate necessary sample sizes that can consistently achieve adequate congruence given expected population parameters.

Across our and the Bagby et al. (2022) results, adequate congruence has been observed for 9 of 15 domains in primarily BA samples, which is less than what would be expected if personality structure, as measured by the PID-5, was nearly identical across racial groups. To explain the difference in factor structure, Bagby et al. (2022) interpret a one-factor solution in the BA sample, thought to reflect general demoralization. They posit that this factor may reflect a Black American experience of “living in a social milieu in which one experiences and feels a general undifferentiated emotional burden from hidden or “unconscious biases” and frequent microaggressions” (p. 89). Though it is plausible that racially influenced experiences may affect the differentiation within self-reported instruments of personality traits, our data do not suggest a broad lack of differentiation.

²Bagby et al. (2022) report congruence between WA and BA samples but not with original Krueger et al. (2012) loadings. We used their reported loading matrix to examine congruence between the BA sample and the original loadings and switched the loadings labelled Disinhibition and Psychoticism. The resulting TCCs were: Negative Affect $\phi = 0.90$, Detachment $\phi = 0.86$, Antagonism $\phi = 0.93$, Disinhibition $\phi = 0.68$, and Psychoticism $\phi = 0.87$.

Nonetheless, though highly similar across the current WA and BA samples, the five-factor structures were not identical, with notable differences in the Suspiciousness facet. To that end, Wolny et al. (2021) examined invariance of the Schizotypal Personality Questionnaire across BA and WA groups, finding evidence of differential item functioning for items of the suspiciousness and paranoid ideation subscales. BA individuals were more likely to endorse these items at lower latent severity levels, suggesting differences in “normative” or adaptive experiences of suspiciousness. In our data, means were also greater for PID-5 Suspiciousness in the BA group compared to the WA group, perhaps suggesting race-specific experiential processes (e.g., heightened fears regarding how they are viewed by others; heightened mistrust of and guardedness), rather than sampling error, leading to non-invariance. Suspiciousness and Submissiveness were the two facets causing non-invariance of facet intercepts, and thus require more study at the item-level moving forward.

The type of rotation used in estimating the factor loadings may be another potentially relevant component of sampling error. Because factor scores are indeterminate, that is, because an infinite number of loadings can be computed from the same factor solution, factor loadings are routinely “rotated” to increase the interpretability of the estimated model. For instance, a target rotation finds the set of loadings within the solution that is least discrepant to a user-defined expected, or “target”, loading matrix. Rotating towards a shared target can decrease heterogeneity in loadings resulting from sampling variability (McCrae et al., 1996; Rolland, 2002), but other types of rotations are routinely used as well, such as the oblique, geomin rotation used by Bagby et al. (2022).

To examine the impact of rotation choice, we re-ran our configural analyses using a geomin rather than target rotation and observed lower congruence values for each domain, with a three below “fair” levels³. This is not to suggest that a target rotation is necessary for achieving adequate congruence; though it is perhaps beneficial, good congruence has been displayed in several previous studies that did not use target rotations (e.g., Bach et al., 2016; Thomas et al., 2013; Wright & Simms, 2015). Altogether, it is difficult to conclude if rotation type has a large impact on congruency, though it appears to have been important in the current sample. Thus, if the PID-5 is indeed prone to sampling error, especially in smaller samples, then the use of different rotations may further increase the heterogeneity in observed congruencies, as has been observed in BA samples.

One final interpretative note when comparing the replicability of the PID-5 across groups is the proposed hierarchical structure at levels broader than just the five domains. Wright et al. (2012) propose additional one-, two-, three-, and four-factor solutions that resemble existing models of psychopathology and personality. For instance, the two-factor PID-5 solution resembles broad internalizing and externalizing domains, and at the three-factor solution, the internalizing construct splits into Negative Affect and Detachment. With a complex interstitial structure and relatively small sample sizes, the extraction of the fourth Disinhibition domain (splitting from Antagonism/externalizing) and the fifth Psychoticism

³The TCCs between the BA and WA groups ranged from 0.63 to 0.94 (Negative Affect $\phi = 0.94$ [0.97 previously with target rotation]; Detachment $\phi = 0.78$ [0.97]; Antagonism $\phi = 0.92$ [0.96]; Psychoticism $\phi = 0.63$ [0.92]; Disinhibition $\phi = 0.81$ [0.87]), indicative of at least fair congruence for just two domains.

domain can be inconsistent compared to that of Negative Affect, Detachment, and Antagonism (e.g., Quilty et al., 2013; Van den Broeck et al., 2014; Zimmerman et al., 2014). Given a high first-to-second eigenvalue ratio, Bagby et al. (2022) interpret a one-factor solution in the BA sample⁴. Indeed, the PID-5 facets are correlated (Somma et al., 2019; Watters & Bagby, 2018) so an undifferentiated, broadly based level of demoralization would likely be interpretable across many samples, approximating a general factor.

However, it is also plausible that an interpretable three- or four-factor solution would have emerged even in the observed absence of the typical five-factor solution in the BA group. For instance, we fit one-, two-, three-, and four-factor solutions using target rotations toward the loadings derived by Wright et al. (2012). We found at least adequate congruence across the WA and BA groups for three domains in the four-factor solution (Negative Affect $\phi = 0.95$; Detachment $\phi = 0.98$; Antagonism $\phi = 0.92$; Disinhibition $\phi = 0.81$), good congruence for each domain of the three-factor solution (Negative Affect $\phi = 0.98$; Detachment $\phi = 0.96$; Externalizing $\phi = 0.99$), good congruence for both domains of the two-factor solution (Internalizing $\phi = 0.98$; Externalizing $\phi = 0.95$), and loadings to be nearly identical in the one-factor solution (General Factor $\phi = 1.00$)⁵. Given the evidence on the hierarchical structure of the PID-5 and occasional difficulty extracting the Disinhibition and Psychoticism domains, intermediate 2-, 3-, or 4-factor models may be interpretable (and invariant) in the absence of an acceptable 5-factor solution.

Clearly, more work is needed to study the structure of maladaptive personality traits in racial, ethnic, and marginalized groups, though this evidence suggests that the AMPD is promising as an applicable diagnostic framework across BA and WA groups in the U.S. There is growing evidence that the PID-5 is a replicable and invariant measure of traits across White Americans and Western Europeans – with increasing evidence in Eastern Europe, the Middle East and elsewhere – as well as across sex, clinical versus community populations, and age. However, with limited and inconsistent evidence across racial groups, we agree with Bagby et al.'s (2022) conclusion that, with Black Americans, the PID-5 “should be used in combination with the DSM-5 Cultural Formulation Interview and interpreted with informed discretion” (p. 88).

Limitations and Future Directions

The relative homogeneity of the sample limits the generalizability of results. Indeed, the entire sample was recruited from the same university, most were young (mean age = 21.3 years), and the sample was predominantly female (76%), thus making it unlikely to be diverse across relevant dimensions such as economic status, cognitive functioning, emotional functioning, and geographic region. Comparison across a wide range of functioning (e.g., across community and clinical populations) is, indeed, a crucial dimension for the PID-5. Though there is evidence that the PID-5 is invariant across clinical and community populations, this has not been demonstrated in a primarily BA sample. Thus, studying

⁴The Bagby et al. (2022) first-to-second eigenvalue ratio was 6.8:1 in the derivation sample and 4.5:1 in the replication sample (both BA). In our data, the first-to-second eigenvalue ratio was 5.7:1 in the BA sample and 4.4:1 in the WA sample.

⁵Note, that comparative fit indices were smaller for solutions with fewer factors (e.g., CFI for the 5-factor configural model was 0.915; CFI for the 3-factor configural model was 0.835; CFI for the 1-factor configural model was 0.674).

the factor structure of psychometric properties of the PID-5 in a BA clinical sample is a critical future direction for the viability of the AMPD. Given that most prior samples in the literature were primarily WA and Western European, similar work is needed across other racial, ethnic, and marginalized groups. In addition, Bagby et al. (2022) note that racial groupings (e.g., WA and BA) may conflate racial group with culture, so additional work is needed to disentangle race and racialized culture. In addition to race, culture, and sample size, future work should consider other possible moderators of factor structure, such as age, language, and factor rotation used in estimation.

Further, there are multiple dimensions of measurement equivalence; the evidence thus presented is indicative of scalar invariance (i.e., the degree of congruence between measures from the same instrument derived from multiple groups). However, this evidence does not indicate, for instance, conceptual or construct equivalence. In the future, external instruments should be used to validate the “meaning” of PID-5 scores across groups and test if reasonably comparable constructs are being measured.

Methodologically speaking, the reliance on strictly self-reported data is a limitation. The PID-5 has been developed into an informant report form with strong psychometric properties which may overcome limitations of self-report data, such as shared method variance or evaluative consistency bias (Markon et al., 2013). Though CFI is highly sensitive, RMSEA is not sensitive to all noninvariance, especially in samples under 300 participants per group (Cao & Liang, 2022). Finally, these analyses occurred at the level of the facet, rather than the item. With larger samples, multiple-group item response theory analyses can detect invariance at the item-level, which may be important moving forward to understand if items underlying inconsistent facets like Suspiciousness or domains like Disinhibition hold different meanings across cultures, and thus, require modification in future versions of the PID-5.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

R.F. Krueger is a coauthor of the PID-5 and provides consulting services to aid users of the PID-5 in the interpretation of test scores. PID-5 is the intellectual property of the American Psychiatric Association, and Dr. Krueger does not receive royalties or any other compensation from publication or administration of the inventory. We have no other conflicts of interest to disclose. R.F. Krueger is partly supported by the US National Institutes of Health, NIH (R01-AG077742, U19-AG051426). C.D.F. is supported by the National Institute on Drug Abuse (T32DA050560).

References

- Aboul-ata M, & Qonsua F (2021). Validity, reliability and hierarchical structure of the PID-5 among Egyptian college students: Using exploratory structural equation modelling. *Personality and Mental Health*, 15(2), 100–112. 10.1002/pmh.1497 [PubMed: 33205600]
- Al-Dajani N, Gralnick TM, & Bagby RM (2016). A psychometric review of the Personality Inventory for DSM–5 (PID–5): Current status and future directions. *Journal of Personality Assessment*, 98(1), 62–81. 10.1080/00223891.2015.1107572 [PubMed: 26619968]

- American Psychiatric Association. (2013). *Diagnostic and Statistical Manual of Mental Disorders DSM-5 Fifth Edition*. 10.1176/appi.books.9780890425596
- Bach B, Maples-Keller JL, Bo S, & Simonsen E (2016). The alternative DSM–5 personality disorder traits criterion: A comparative examination of three self-report forms in a Danish population. *Personality Disorders: Theory, Research, and Treatment*, 7(2), 124–135. 10.1037/per0000162
- Bach B, Sellbom M, & Simonsen E (2018). Personality Inventory for DSM-5 (PID-5) in clinical versus nonclinical individuals: Generalizability of psychometric features. *Assessment*, 25(7), 815–825. 10.1177/1073191117709070 [PubMed: 28503935]
- Bagby RM, Keeley JW, Williams CC, Mortezaei A, Ryder AG, & Sellbom M (2022). Evaluating the measurement invariance of the Personality Inventory for DSM-5 (PID-5) in Black Americans and White Americans. *Psychological Assessment*, 34(1), 82–90. 10.1037/pas0001085 [PubMed: 34871023]
- Barchi-Ferreira Bel AM, & Osório FL (2020). The Personality Inventory for DSM-5: Psychometric evidence of validity and reliability-updates. *Harvard Review of Psychiatry*, 28(4), 225–237. 10.1097/HRP.0000000000000261 [PubMed: 32692087]
- Becker LG, Asadi S, Zimmerman M, Morgan TA, & Rodriguez-Seijas C (2022). Is there a bias in the diagnosis of borderline personality disorder among racially minoritized patients? *Personality Disorders: Theory, Research, and Treatment*. 10.1037/per0000579
- Cao C, & Liang X (2022). Sensitivity of fit measures to lack of measurement invariance in exploratory structural equation modeling. *Structural Equation Modeling: A Multidisciplinary Journal*, 29(2), 248–258. 10.1080/10705511.2021.1975287
- Chen FF (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling*, 14(3), 464–504. 10.1080/10705510701301834
- Debast I, Rossi G, & van Alphen SPJ (2018). Age-Neutrality of a Brief assessment of the Section III Alternative Model for Personality Disorders in older adults. *Assessment*, 25(3), 310–323. 10.1177/1073191118754706 [PubMed: 29405756]
- Freilich CD, Krueger RF, Hobbs KA, Hopwood CJ, & Zimmerman J (2023). The DSM-5's maladaptive trait model for personality disorders. In Krueger RF, & Blaney PH (Eds.), *Oxford Textbook of Psychopathology Fourth Edition*. Oxford University Press.
- Gomez R, Watson S, Brown T, & Stavropoulos V (2022). Personality inventory for DSM–5-Brief Form (PID-5-BF): Measurement invariance across men and women. *Personality Disorders: Theory, Research, and Treatment*. 10.1037/per0000569
- Krueger RF, Derringer J, Markon KE, Watson D, & Skodol AE (2012). Initial construction of a maladaptive personality trait model and inventory for DSM-5. *Psychological Medicine*, 42(9), 1879–1890. 10.1017/S0033291711002674 [PubMed: 22153017]
- Kupfer DJ, First MB, & Regier DA (Eds.). (2002). *A research agenda for DSM-V (1st ed)*. American Psychiatric Association.
- Lim DSH, Gwee AJ, & Hong RY (2019). Associations between the DSM-5 Section III Trait Model and impairments in functioning in Singaporean college students. *Journal of Personality Disorders*, 33(3), 413–431. 10.1521/pedi_2018_32_353 [PubMed: 29949443]
- Lorenzo-Seva U, & ten Berge JMF (2006). Tucker's Congruence Coefficient as a meaningful index of factor similarity. *Methodology*, 2(2), 57–64. 10.1027/1614-2241.2.2.57
- Markon KE, Quilty LC, Bagby RM, & Krueger RF (2013). The development and psychometric properties of an informant-report form of the personality inventory for DSM-5 (PID-5). *Assessment*, 20(3), 370–383. 10.1177/1073191113486513 [PubMed: 23612961]
- Marsh HW, Muthén B, Asparouhov T, Lüdtke O, Robitzsch A, Morin AJS, & Trautwein U (2009). Exploratory structural equation modeling, integrating CFA and EFA: Application to students' evaluations of university teaching. *Structural Equation Modeling: A Multidisciplinary Journal*, 16(3), 439–476. 10.1080/10705510903008220
- McCrae RR, Zonderman AB, Costa PT Jr., Bond MH, & Paunonen SV (1996). Evaluating replicability of factors in the Revised NEO Personality Inventory: Confirmatory factor analysis versus Procrustes rotation. *Journal of Personality and Social Psychology*, 70(3), 552–566. 10.1037/0022-3514.70.3.552

- Meredith W, & Teresi JA (2006). An essay on measurement and factorial invariance. *Medical Care*, 44(11, Suppl 3), S69–S77. 10.1097/01.mlr.0000245438.73837.89 [PubMed: 17060838]
- Muthén LK, & Muthén BO (1998–2022). *Mplus User’s Guide* (8th ed.). <https://www.statmodel.com/html Ug.shtml>
- Quilty LC, Ayeart L, Chmielewski M, Pollock BG, & Bagby RM (2013). The psychometric properties of the Personality Inventory for DSM-5 in an APA DSM-5 field trial sample. *Assessment*, 20(3), 362–369. 10.1177/1073191113486183 [PubMed: 23588687]
- Riegel KD, Ksinan AJ, Samankova D, Preiss M, Harsa P, & Krueger RF (2018). Unidimensionality of the personality inventory for DSM-5 facets: Evidence from two Czech-speaking samples. *Personality and Mental Health*, 12(4), 281–297. 10.1002/pmh.1423 [PubMed: 29952078]
- Rolland J-P (2002). The cross-cultural generalizability of the Five-Factor Model of personality. In McCrae RR & Allik J (Eds.), *The Five-Factor Model of Personality Across Cultures* (pp. 7–28). Springer US. 10.1007/978-1-4615-0763-5_2
- Russell TD, Pocknell V, & King AR (2017). Lesbians and bisexual women and men have higher scores on the Personality Inventory for the DSM-5 (PID-5) than heterosexual counterparts. *Personality and Individual Differences*, 110, 119–124. 10.1016/j.paid.2017.01.039
- Somma A, Borroni S, Maffei C, Giarolli LE, Markon KE, Krueger RF, & Fossati A (2017). Reliability, factor structure, and associations with measures of problem relationship and behavior of the Personality Inventory for DSM-5 in a sample of Italian community-dwelling adolescents. *Journal of Personality Disorders*, 31(5), 624–646. 10.1521/pedi_2017_31_272 [PubMed: 28072038]
- Somma A, Krueger RF, Markon KE, & Fossati A (2019). The replicability of the personality inventory for DSM-5 domain scale factor structure in U.S. and non-U.S. samples: A quantitative review of the published literature. *Psychological Assessment*, 31(7), 861–877. 10.1037/pas0000711 [PubMed: 30883152]
- Somma A, Krueger RF, Markon KE, Borroni S, & Fossati A (2019b). Item response theory analyses, factor structure, and external correlates of the Italian translation of the Personality Inventory for DSM-5 Short Form in community-dwelling adults and clinical adults. *Assessment*, 26(5), 839–852. 10.1177/1073191118781006 [PubMed: 29902930]
- Sorrel MA, García LF, Aluja A, Rolland JP, Rossier J, Roskam I, & Abad FJ (2021). Cross-Cultural measurement invariance in the Personality Inventory for DSM-5. *Psychiatry Research*, 304, 114134. 10.1016/j.psychres.2021.114134 [PubMed: 34358762]
- South SC, Krueger RF, Knudsen GP, Ystrom E, Czajkowski N, Aggen SH, Neale MC, Gillespie NA, Kendler KS, & Reichborn-Kjennerud T (2017). A population based twin study of DSM-5 maladaptive personality domains. *Personality Disorders*, 8(4), 366–375. 10.1037/per0000220 [PubMed: 27797545]
- Suzuki T, South SC, Samuel DB, Wright AGC, Yalch MM, Hopwood CJ, & Thomas KM (2019). Measurement invariance of the DSM-5 Section III pathological personality trait model across sex. *Personality Disorders: Theory, Research, and Treatment*, 10(2), 114–122. 10.1037/per0000291
- Thimm JC, Jordan S, & Bach B (2017). Hierarchical structure and cross-cultural measurement invariance of the Norwegian version of the personality inventory for DSM-5. *Journal of Personality Assessment*, 99(2), 204–210. 10.1080/00223891.2016.1223682 [PubMed: 27661474]
- Thomas KM, Yalch MM, Krueger RF, Wright AGC, Markon KE, & Hopwood CJ (2013). The convergent structure of DSM-5 personality trait facets and Five-Factor Model trait domains. *Assessment*, 20(3), 308–311. 10.1177/1073191112457589 [PubMed: 22946103]
- Tyrer P, Coombs N, Ibrahimi F, Mathilakath A, Bajaj P, Ranger M, Rao B, & Din R (2007). Critical developments in the assessment of personality disorder. *The British Journal of Psychiatry*, 190(S49), s51–s59. 10.1192/bjp.190.5.s51
- Van den Broeck J, Bastiaansen L, Rossi G, Dierckx E, & De Clercq B (2013). Age-Neutrality of the trait facets proposed for Personality Disorders in DSM-5: A DIFAS analysis of the PID-5. *Journal of Psychopathology and Behavioral Assessment*, 35(4), 487–494. 10.1007/s10862-013-9364-3
- Van den Broeck J, Bastiaansen L, Rossi G, Dierckx E, De Clercq B, & Hofmans J (2014). Hierarchical structure of maladaptive personality traits in older adults: Joint factor analysis of the PID-5 and the DAPP-BQ. *Journal of Personality Disorders*, 28(2), 198–211. 10.1521/pedi_2013_27_114 [PubMed: 23786271]

- Watters CA, & Bagby RM (2018). A meta-analysis of the five-factor internal structure of the Personality Inventory for DSM-5. *Psychological Assessment*, 30(9), 1255–1260. 10.1037/pas0000605 [PubMed: 29952594]
- Wolny J, Moussa-Tooks AB, Bailey AJ, O'Donnell BF, Hetrick WP (2021). Race and self-reported paranoia: Increased item endorsement on subscales of the SPQ. *Schizophrenia Research*. doi.org/10.1016/j.schres.2021.11.034
- Wright AGC, & Simms LJ (2015). A metastructural model of mental disorders and pathological personality traits. *Psychological Medicine*, 45(11), 2309–2319. 10.1017/S0033291715000252 [PubMed: 25903065]
- Wright AGC, Thomas KM, Hopwood CJ, Markon KE, Pincus AL, & Krueger RF (2012). The hierarchical structure of DSM-5 pathological personality traits. *Journal of Abnormal Psychology*, 121(4), 951–957. 10.1037/a0027669 [PubMed: 22448740]
- Zhang P, Ouyang Z, Fang S, He J, Fan L, Luo X, Zhang J, Xiong Y, Luo F, Wang X, Yao S, & Wang X (2021). Personality inventory for DSM-5 brief form (PID-5-BF) in Chinese students and patients: Evaluating the five-factor model and a culturally informed six-factor model. *BMC Psychiatry*, 21(1), 107. 10.1186/s12888-021-03080-x [PubMed: 33596861]
- Zimmermann J, Altenstein D, Krieger T, Holtforth MG, Pretsch J, Alexopoulos J, Spitzer C, Benecke C, Krueger RF, Markon KE, & Leising D (2014). The structure and correlates of self-reported DSM-5 maladaptive personality traits: Findings from two German-speaking samples. *Journal of Personality Disorders*, 28(4), 518–540. 10.1521/pedi_2014_28_130 [PubMed: 24511899]
- Zimmermann J, Kerber A, Rek K, Hopwood CJ, & Krueger RF (2019). A brief but comprehensive review of research on the Alternative DSM-5 Model for Personality Disorders. *Current Psychiatry Reports*, 21(9), 92. 10.1007/s11920-019-1079-z [PubMed: 31410586]

Public Significance Statement

The Personality Inventory for *DSM-5* is increasingly being used to assess personality disorders, though it is unclear if it measures the same constructs in White and Black American samples. Previous evidence suggests some inconsistencies, but the evidence presented here suggests it is an applicable tool across groups. More data from diverse samples is needed to further clarify its measurement properties.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 1.

PID-5 Configural ESEM Model Factor Loadings and Congruence Coefficients

PID-5 Facets	Black Americans (<i>n</i> = 613)					White Americans (<i>n</i> = 612)				
	NA	Det	Ant	Dis	Psyc	NA	Det	Ant	Dis	Psyc
Negative Affectivity										
Anxiousness	0.66	0.24	0.03	-0.09	0.07	0.64	0.29	-0.05	-0.05	0.21
Emotional Lability	0.59	0.09	0.15	0.07	0.13	0.70	0.05	0.11	0.18	0.08
Hostility	0.19	0.29	0.57	0.04	-0.01	0.28	0.18	0.50	0.15	0.06
Perseveration	0.38	0.15	0.06	0.23	0.36	0.43	0.21	0.10	0.16	0.41
Restricted Affectivity (R)	-0.35	0.51	0.21	-0.10	0.38	-0.43	0.48	0.13	-0.03	0.43
Separation Anxiety	0.57	0.00	0.10	0.34	0.04	0.60	0.05	0.18	0.18	0.03
Submissiveness	0.35	0.12	0.06	0.20	0.09	0.34	0.16	0.05	-0.10	0.21
Detachment										
Anhedonia	0.17	0.64	-0.18	0.39	0.01	0.20	0.65	-0.10	0.42	-0.07
Depressivity	0.25	0.43	-0.11	0.51	0.16	0.32	0.50	-0.07	0.41	0.15
Intimacy Avoidance	-0.22	0.57	0.04	0.27	0.20	-0.27	0.50	0.02	0.12	0.26
Suspiciousness	0.30	0.30	0.10	0.05	0.16	0.19	0.37	0.25	0.16	0.06
Withdrawal	0.06	0.81	0.07	-0.05	0.08	-0.02	0.70	-0.03	0.08	0.29
Antagonism										
Attention Seeking	0.22	-0.28	0.55	0.23	0.14	0.22	-0.43	0.60	0.11	0.25
Callousness	-0.22	0.32	0.48	0.37	0.15	-0.20	0.41	0.52	0.30	-0.01
Deceitfulness	0.06	0.08	0.62	0.32	0.07	0.05	0.12	0.60	0.22	0.16
Grandiosity	-0.10	-0.03	0.64	0.06	0.19	-0.05	0.00	0.68	-0.03	0.12
Manipulativeness	0.00	-0.09	0.78	0.07	0.10	-0.06	-0.04	0.73	0.00	0.27
Disinhibition										
Distractibility	0.36	0.17	0.02	0.22	0.31	0.28	0.10	-0.12	0.40	0.46
Impulsivity	0.19	-0.11	0.27	0.38	0.24	0.08	-0.19	0.08	0.63	0.33
Irresponsibility	0.07	0.23	0.16	0.64	0.08	0.05	0.18	0.11	0.60	0.16
Rigid Perfectionism (R)	0.19	0.11	0.20	-0.13	0.35	0.33	0.27	0.36	-0.35	0.21
Risk-Taking	-0.07	-0.34	0.33	0.24	0.28	-0.24	-0.36	0.22	0.49	0.25
Psychoticism										
Eccentricity	0.13	0.11	0.03	0.01	0.65	0.20	0.10	0.01	0.18	0.63
Perceptual Dysregulation	0.08	0.15	0.03	0.23	0.68	0.13	0.25	0.12	0.29	0.48
Unusual Beliefs	-0.04	0.09	0.11	0.09	0.73	-0.01	0.13	0.31	0.16	0.45
Congruence Coefficient	0.96	0.98	0.97	0.86	0.98	0.96	0.99	0.98	0.96	0.88

Note. NA = negative affectivity; Det = detachment; Ant = antagonism; Dis = disinhibition; Psyc = psychoticism. Bold highlights factor loadings [>.35]. TCC calculated in comparison to Somma et al. (2019) meta-analytically derived loading matrix. Facets listed under the domain in which they were initially proposed to load on primarily, though notable cross-loadings of Hostility onto Antagonism, Restricted Affect onto Detachment, Depressivity, Rigid Perfectionism, and Suspiciousness onto Negative Affectivity have been observed elsewhere (Waters & Bagby, 2018). (R) = Item is keyed negatively onto the domain.

Table 2.

ESEM Measurement Invariance Testing across White and Black Americans

Model	χ^2	df	CFI	CFI	RMSEA (90% CI)	RMSEA	AIC	AIC	BIC	BIC	Comparison
1. Configural	1,883.6	370	0.915	0.082 (0.078–0.085)			38,281		39,968		
2. Weak	2,057.0	470	0.911	0.004 (0.071–0.078)	–0.008		38,401	119.4	39,576	–391.6	Model 1
3. Weak + Item	2,179.6	495	0.905	0.006 (0.071–0.078)	0.001		38,506	105.4	39,554	–22.4	2
4. Weak + FVCV	2,126.1	485	0.907	0.004 (0.071–0.078)	0.000		38,444	44.0	39,543	–32.7	2
5. Strong	2,262.7	490	0.900	0.011 (0.074–0.080)	0.003		38,559	158.8	39,633	56.5	2
6. Weak + FVCV + Item	2,255.0	510	0.902	0.009 (0.072–0.078)	0.001		38,570	169.4	39,541	–35.0	2
7. Strict	2,390.6	515	0.894	0.006 (0.077–0.080)	0.000		38,665	105.4	39,610	–22.3	5
8. Strong + FVCV	2,337.2	505	0.897	0.003 (0.074–0.080)	0.000		38,605	45.9	39,602	–30.7	5
9. Strict + FVCV	2,468.3	530	0.891	0.003 (0.074–0.080)	0.000		38,730	64.9	39,599	–11.8	7
10. Strong + LFMn	2,325.5	495	0.897	0.003 (0.075–0.081)	0.001		38,634	74.7	39,682	49.1	5
11. Strict + LFMn	2,450.1	520	0.891	0.003 (0.078–0.081)	0.001		38,739	74.3	39,659	48.8	7
12. Strong + LFMn + FVCV	2,396.1	510	0.894	0.006 (0.075–0.081)	0.001		38,680	120.8	39,651	18.6	5
13. Strict + LFMn + FVCV	2,525.5	535	0.888	0.006 (0.078–0.081)	0.001		38,804	139.3	39,647	37.1	7

Note. CFI = comparative fit index; RMSEA = root mean square error of approximation; CI = confidence interval; AIC = Akaike Information Criteria; BIC = Bayesian Information Criteria; Item = invariance of item/facet uniquenesses (i.e., residual variance); FVCV = invariance of factor variances/covariances; LFMn = invariance of latent factor means. Support for the more parsimonious model requires a change in CFI and RMSEA that does not exceed .010 (Chen, 2007; Marsh et al., 2009).