AMERICAN PSYCHOLOGICAL ASSOCIATION

Psychological Assessment

© 2025 American Psychological Association
ISSN: 1040-3590

2026, Vol. 38, No. 2, 71–84
https://doi.org/10.1037/pas0001421

Longitudinal Measurement Invariance of the Personality Inventory for *ICD-11* Across Black and White American Older Adults

Pooja Heragu[1], Nathalie Dieujuste[2], Yara Mekawi[3], and Joshua R. Oltmanns[4]
[1] School of Psychology, Xavier University
[2] Department of Psychology, University of Denver
[3] Department of Psychological and Brain Sciences, University of Louisville
[4] Department of Psychological and Brain Sciences, Washington University, St. Louis

The Personality Inventory for *ICD-11* (PiCD) assesses five maladaptive trait domains from the *International Classification of Diseases*–11th edition's dimensional model of personality disorder. Validity evidence of PiCD scores has relied primarily on White samples and there have been no evaluations of measurement invariance (MI). Research examining use of PiCD scores with diverse populations is needed. The present study investigated MI of PiCD scores across race and time in sample of White and Black American older adults (n = 843, ∼20% Black). Cross-sectionally, Marsh et al.'s (2009) 13-step exploratory structural equation modeling was used to determine MI of the five domains across Black and White participants at two waves of data collection about 2 years apart. Findings revealed partial strong invariance across race at both waves. At Wave 1, intercepts for two Anankastia items and two negative affectivity items (only one negative affectivity item at Wave 2) were noninvariant across race. Longitudinal exploratory structural equation modeling suggested strict invariance across time for the entire sample. Domain-level longitudinal confirmatory factor analysis indicated strict invariance across time for Black participants in each PiCD domain. Findings suggest four item means demonstrated noninvariance and require further examination, but the PiCD scores showed a high level of invariance (factor structure, factor loadings, 56 of 60 item intercepts). Reasons for the four noninvariant item intercepts are probed by examining scale score differences with and without the items and external correlates. Findings indicate partial strong invariance for PiCD scores, but the four item mean scores need further exploration across race, and potential revision.

Public Significance Statement
This 2-year study on the Personality Inventory for *ICD-11* (PiCD) examines the structure and, by extension, the interpretability of the PiCD across a community sample of Black and White older adults. Importantly, the study considers differences in four items that could impact how the measure may function across Black and White participants. Further research using the PiCD with diverse groups is needed in the future.

Keywords: measurement invariance, *ICD-11*, PiCD, older adults, community

Supplemental materials: https://doi.org/10.1037/pas0001421.supp

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.
All rights, including for text and data mining, AI training, and similar technologies, are reserved.

This article was published Online First November 3, 2025.
Javier Suárez-Álvarez served as action editor.
Pooja Heragu https://orcid.org/0009-0007-5983-2433
Data and code using for the present study are listed on the Open Science Framework (https://osf.io/s9b4x/?view_only=ad6ea60012b54ae4b11f3196e3a20bf8).

This research was supported by the National Institutes of Health, National Institute on Aging, National Institute of Mental Health under Award Numbers R01-AG061162, R01-AG045231, and R01-MH077840. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

The authors thank the participants and researchers of the St. Louis Personality and Aging Network (SPAN) for making this study possible. Also thanks to Michael R. Bagby, Martin Sellbom, and colleagues for sharing MPlus measurement invariance code.

Pooja Heragu played a lead role in conceptualization, formal analysis, methodology, software, and writing–original draft. Nathalie Dieujuste played a supporting role in conceptualization and writing–original draft. Yara Mekawi played a supporting role in conceptualization, writing–original draft, and writing–review and editing. Joshua R. Oltmanns played a lead role in conceptualization, data curation, and writing–review and editing and an equal role in formal analysis, methodology, and software.

Correspondence concerning this article should be addressed to Pooja Heragu, School of Psychology, Xavier University, 3800 Victory Parkway, Cincinnati, OH 45207, United States. Email: poojheragu@gmail.com

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.
All rights, including for text and data mining, AI training, and similar technologies, are reserved.

Categorical diagnoses have limitations warranting a shift toward dimensional diagnostic models (e.g., Clark, 2007; Tyrer et al., 2015; Widiger & Trull, 2007). In a section for emerging measures and models, the 5th edition of the *Diagnostic and Statistical Manual of Mental Disorders*, fifth edition (*DSM-5*) included a dimensional model of personality disorder diagnosis called the alternate model of personality disorders (AMPD). This model consisted of one general severity dimension as well as five maladaptive personality trait domains: negative affectivity, detachment, antagonism, disinhibition, and psychoticism. Soon after, the 11th edition of the *International Classification of Diseases* (*ICD-11*) officially adopted a dimensional model of personality diagnosis (World Health Organization, 2025). The *ICD-11* model includes three broad components: a required rating of general severity of personality dysfunction, participant functioning regarding five personality trait domains, and a borderline pattern specifier. The five personality trait domains are negative affectivity, detachment, dissociality, disinhibition, and anakastia. These five domains can be conceptualized as maladaptive variants of five-factor model personality trait domains. Specifically, negative affectivity aligns with neuroticism, detachment aligns with extroversion, dissociality aligns with agreeableness, and disinhibition and anakastia align with opposing ends of conscientiousness (Mulder et al., 2016).

The Personality Inventory for the *ICD-11* (PiCD), a 60-question self-report measure, was developed to assess the five maladaptive train domains include in the *ICD-11* model (J. R. Oltmanns & Widiger, 2018). Three rounds of item performance evaluation were completed to construct a working structure and select final items. Subsequently, two initial validation studies comparing convergent and discriminant validity between both general and maladaptive personality trait measures were conducted, with results supporting the construct validity of the PiCD scale scores (J. R. Oltmanns & Widiger, 2018). Research to date has repeatedly replicated a four-factor structure, with anakastia and disinhibition forming opposite ends of a single bipolar factor (J. R. Oltmanns, 2021).

After the initial construction and validation of the PiCD, subsequent studies conducted in American and European settings have yielded findings that strengthen the convergent, discriminant, and criterion validity of the PiCD scores, and supported the four-factor structure of the PiCD (J. R. Oltmanns & Widiger, 2021; Somma et al., 2020; Stricker et al., 2022). Most of this evidence, however, was based on predominantly White or European samples. Given that the PiCD will be used to study personality in people of varying racial and ethnic backgrounds across the world, it is imperative to evaluate whether the PiCD is across different races and ethnicities. Fair assessment requires validity evidence across populations and a core tenant of such validity includes the assurance that the measure is evaluating the same or similar constructs across groups, allowing the scores to be interpreted in a similar manner for different groups of people (Ramírez et al., 2005).

The *ICD-11* and *DSM-5* AMPD trait domains can be conceptualized as maladaptive variations of four of the five-factor model trait domains. Of note is that psychoticism, which is included in the *DSM-5* AMPD, is not included in the *ICD-11* model, while the *ICD-11* model includes an anankastia domain that is not included in the *DSM-5* AMPD. This is because schizotypal personality disorder (which aligns with psychoticism) has traditionally been included in the schizophrenia-related disorders section of the *ICD*. *DSM-5* AMPD included a compulsivity domain in an earlier edition of that model (analogous to anankastia), but ultimately eliminated it for parsimony (Krueger et al., 2012). Regardless, measurement equivalence of personality trait domain assessments is important to demonstrate across ethnoracial groups.

Regarding measures of maladaptive personality traits, researchers have investigated measurement invariance of the Personality Inventory for the *DSM-5* (PID-5) in a variety of samples. The PID-5 is a 220-item self-report measure that produces a score for the five maladaptive trait domains included in the *DSM-5* AMPD (Krueger et al., 2012). At the domain level, Somma et al. (2019) assessed the factor structure (configural invariance) that emerged from samples of participants who took the PID-5 from the United States, European, and Middle Eastern countries. Results indicated that the expected five-factor structure emerged in all samples. Zhang et al. (2021) assessed measurement invariance for the Personality Inventory for the *DSM-5* brief form (PID-5-BF) in Chinese undergraduate students and clinical patients. Findings indicate that, although the expected five-factor model emerged with acceptable fit, an exploratory six-factor structure, in which negative affectivity was divided into two separate factors, produced better model fit comparative fit index (CFI = .905 undergraduates, CFI = .904 clinical sample). Encouragingly, partial strict invariance was found across both the clinical and undergraduate students in the samples using the six-factor model.

Within the United States, Asadi et al. (2024) investigated measurement invariance of the PID-5-BF across sexual and gender minority (SGM) individuals. In both clinical (n = 1,174 total, n = 254 SGM) and nonclinical (n = 1,456 total undergraduates, n = 151 SGM) samples, the predicted five-factor structure of the PID-5-BF was found along with invariance of factor loadings and item thresholds. Invariance was not found at the item intercept level. In general, SGM individuals in both samples endorsed a higher average score in the antagonism, negative affectivity, disinhibition, and psychoticism than the non-SGM individuals.

Focusing on racial measurement invariance in the United States, Becker et al. (2023), who were broadly investigating the potential bias in borderline personality diagnosis of non-White individuals, conducted measurement invariance of the PID-5-BF among White and non-White (all participants who identified with a racial category other than White) participants to ensure that the measure of maladaptive trait domains in their study was unbiased. The authors found that, in a sample of 2,657 partial hospitalization patients, measurement invariance of the PID-5-BF across race was supported in their sample. These findings suggest that at least partial measurement invariance can be found across individuals of different races in the United States.

To investigate PID-5 applicability across Black and White Americans, Bagby et al. (2022) tested measurement invariance of the PID-5 between Black and White American college students. Participants were broken up into a derivation (n = 590 White, n = 255 Black) and replication (n = 1,317, n = 456 Black) samples. In the derivation sample, five-factor exploratory structural equation modeling (ESEM) models were analyzed separately for White and Black participants to test whether it was plausible that a five-factor structure would emerge for both White and Black participants. However, upon testing configural invariance (i.e., factor structure invariance) in the derivation sample, the authors found that the expected five-factor structure only emerged for White participants. To investigate the best possible factor structure for the Black participants, exploratory factor analysis was conducted. It was found that the ratio between the first and second factor extracted was very

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.
All rights, including for text and data mining, AI training, and similar technologies, are reserved.

large, indicating that a single factor structure was the most suitable PID-5 structure for Black participants. A similar pattern of results for the ESEM and exploratory factor analysis analyses were found in the replication sample—that is configural invariance was not found and a one-factor solution fit Black participants best. Based on these findings, Bagby et al. (2022) concluded that the one-factor solution found for Black Americans may reflect that the PID-5 is measuring overall demoralization or negative emotions associated with the discrimination Black Americans face in a racialized society like the United States. Specifically, overall demoralization can be understood as an overarching factor that may be contributing to increased intercorrelations between the expected five factors for only the Black participants in the sample. Thus, for Black Americans, the PID-5 may measure negative emotions, experiences, and demoralization associated with Black individuals' racialization in American society.

A follow-up study investigated the merit of the aforementioned racialization theory for noninvariance across Black and White Americans by assessing measurement invariance of the PID-5 across White American (n = 212) and Nigerian (n = 250) undergraduate students (Orjiakor et al., 2023). A Nigerian sample was chosen, because Black individuals in Nigeria may not experience the same racialization as Black Americans, allowing the authors to test whether similar factor level differences would be found with Black participants across racialized and nonracialized societies. In this case, the configural invariance model demonstrated acceptable fit, suggesting that, in general the factor structure is similar across Nigerian and White American students. Ultimately, significant differences among facet loadings, mean scores, and residuals were found. These findings provide tentative support for the idea that the race-based discrimination faced by Black individuals in the United States may contribute to the differences in factor-structure. However, these findings should be interpreted with caution as racialization was not isolated from other country-related differences.

In contrast, in a replication of study conducted by Bagby et al. (2022), Freilich et al. (2023) found that, for the PID-5, a five-factor structure emerged for both Black (n = 613) and White (n = 612) American students. Subsequent testing suggested partial strong invariance of the PID-5 across the Black and White undergraduate students in the sample. These findings provide some support for the notion that partial invariance can be found across Black and White Americans. Taken together, however, the summation of evidence above suggests mixed findings regarding the applicability of the PID-5 across Black and White Americans, highlighting the need for further testing.

The disparities in the aforementioned findings may be due in part to methodological differences. First, it should be noted that Bagby et al. (2022) used a derivation sample that included American students from three separate universities from different regions in the United States and a replication sample, while Freilich et al. (2023) used a sample from one university located in the southeastern United States. Additionally, Orjiakor et al. (2023) used a sample of Nigerian students to directly assess whether there was support for the racialization hypothesis. Sample geographic differences may contribute to discrepancies. Second, Bagby et al. used geomin rotation whereas Freilich et al. used target rotation in their analyses. Target rotation involves defining a target matrix based on the expected structure found in previous literature, whereas geomin rotation aims to mitigate variable complexity to extract a simple structure that best fits the data (Myers et al., 2015). In Freilich et al., factor loadings were rotated to be in line with a preexisting loading matrix based on available literature to avoid conflating loading differences as being related to the replicability of factor structure as opposed to sample differences. Previous research has suggested that, in factor analyses, target rotation can demonstrate better accuracy than geomin rotation, whereas geomin rotation can demonstrate better stability than target rotation (Myers et al., 2015). It is important to note that the reviewed findings concern the PID-5, which measures the *DSM-5* AMPD. However, as discussed above, the *DSM-5* AMPD and *ICD-11* personality disorder trait models include four conceptually analogous domains (negative affectivity, detachment, disinhibition, and antagonism/dissociality). The above evidence informs the expected results for the PiCD—measurement invariance of the PiCD scores among diverse populations is a strong possibility, but it is also possible that the PiCD structure will dovetail that of the PID-5 found by Bagby et al., whereby only a one-factor structure is best for Black Americans. However, targeted measurement invariance research with the PiCD is necessary to shed important light on this subject. To date, we are unaware of any measurement invariance studies of the PiCD.

Method

Transparency and Openness

This study's design and its analysis were not preregistered. Analysis follows guidelines that have been widely and methodically used in previous studies reviewed below. Data were analyzed using Mplus, Version 8 (Muthén & Muthén, 2011). All data, analysis code, and research materials are available on the Open Science Framework (https://osf.io/s9b4x/?view_only=ad6ea60012b54ae4b11f3196e3a20bf8; Heragu & Oltmanns, 2025). The present study was approved by the local university Institutional Review Board.

Participants

Participants selected for this analysis were part of a larger, longitudinal study that began in 2007 to investigate personality-related aging outcomes in a group of older adults from St. Louis, Missouri, USA (T. F. Oltmanns, Rodrigues, et al., 2014). Data collection procedures for the SPAN study were approved by the Washington University institutional review board. Participants were St. Louis community members who were 55 to 64 years old at baseline, lacked significant preexisting illness (e.g., psychosis) and did not have significant language/reading difficulties. Recruitment efforts included an overrecruitment of Black men participants after initially lower participation rates, and the sample is representative of race and ethnicity in the St. Louis metropolitan area (Spence & Oltmanns, 2011). Participants completed the PiCD as a part of a larger assessment battery at follow-up 13 (Wave 1 for the present analysis) between October of 2017 and April of 2021 and at follow-up 14 (Wave 2 for the present analysis) between December 2019 and January 2022. At Wave 1, $N = 711$ target participants ($M_{age} = 69.5$ years, SD = 5.4 years) completed the PiCD. Wave 1 participants identified as 55.1% men, 76.4% White/Caucasian, and 21.5% Black/African-American. At Wave 2, N = 748 target participants (M_{age} = 69.6 years, SD = 4.5 years) completed the PiCD. Wave 2 participants identified as 56.2% men, 78.4% White/Caucasian, and 19.8% Black/African-American. This sample provides a strong

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.
All rights, including for text and data mining, AI training, and similar technologies, are reserved.

opportunity to assess measurement invariance across Black and White participants in a large, representative sample of older adults from a large metropolitan city in the United States. Previous research has suggested that PiCD scores from this community sample include individuals who report high levels of PiCD maladaptive traits (J. R. Oltmanns & Widiger, 2021). Full details about recruitment and the participants can be found in other sources (T. F. Oltmanns, Rodrigues, et al., 2014).

Measures

Personality Inventory for the ICD-11

The PiCD (J. R. Oltmanns & Widiger, 2018) is a 60-item measure with five subscales of 12 items each that correspond the five *ICD-11* maladaptive trait domains: anankastia, disinhibition, detachment, dissociality, and negative affectivity. Each item is rated on a Likert type scale ranging from 1 (*strong disagree*) to 5 (*strong agree*). The PiCD is freely available as part of the online Supplemental Materials in its developmental article. Previous validation evidence for the PiCD scale scores is briefly discussed in the introduction and is reviewed in greater detail by J. R. Oltmanns (2021). Internal consistency estimates were acceptable for all five PiCD subscale scores at Wave 1, ranging from $\alpha = .72$ (anankastia) to $\alpha = .86$ (negative affectivity), with a median of .76. At Wave 2, internal consistency estimates for all five PiCD subscale scores were also acceptable, ranging from $\alpha = .71$ (anankastia) to $\alpha = .84$ (negative affectivity), with a median of .73.

Race

Race was measured with one multiple choice question that had eight possible responses including "White, Caucasian," "Black, African-American," "East Asian-Pacific Islander," "South Asian (e.g., India, Pakistan)," "Middle Eastern," "Native American," "Biracial," and "Other." Only participants who identified themselves as "White, Caucasian" or "Black, African-American" were included within the current analysis.

External Correlate Measures

Scores from four external outcome measures were utilized as correlates to better assess the implications of excluding noninvariant measure elements on scale-level predictive validity.

Self-Report Criteria Variables. The Beck Depression Inventory–II (BDI-II; Beck et al., 1996) was used to assess depressive symptoms and includes 21 items rated on a 4-point scale and has extensive validation evidence (Erford et al., 2016). The Satisfaction With Life Scale (Diener et al., 1985) assesses life satisfaction among target participants and includes five items rated on a 7-point Likert scale from 1 (*strongly disagree*) to 7 (*strongly agree*). The RAND-36 Health Status Inventory (P-HSI; Hays & Morales, 2001) was used to assess subjective physical functioning. The scale consists of 10 items assessing limitations of physical functioning especially relevant to older adults (e.g., "Does your health now limit you in these activities? If so, how much?"), with an example item being "Lifting or carrying groceries." Higher scores indicate better physical functioning. Coefficient alphas for the HSI Physical Functioning scale and the BDI-II scale in the SPAN study are about .90 (Cruitt & Oltmanns, 2019). Participants completed an adapted version of the Social Network Questionnaire (SNQ; Kahn & Antonucci, 1980) in which they were asked to imagine three concentric circles that illustrated their social network and complete the seven-items from the Quality of Relationship Inventory–General Support Subscale (Pierce et al., 1991). Regarding the concentric circles, the closest circle was meant to represent individual the person could not imagine their life without, the next largest circle represented important individuals who were not as close as the innermost circle, and the largest circle represented individuals who were less close than the first two circles, but still important in their life. Participants were asked to list up to 10 adult individuals for the first two circles to determine the breadth of an individuals close social network. The Quality of Relationship Inventory items assessed the extent of perceived social support and included items rated on a scale from 1 (*not at all*) to 4 (*very much*), with higher scores indicating greater perceived social support. Last, the seven-item Insomnia Severity Index (ISI; Bastien et al., 2001) was completed by participants to assess insomnia symptoms, with higher scores indicating greater insomnia severity. Coefficient α for ISI scores in the SPAN study are .89 (J. R. Oltmanns, Weinstein, & Oltmanns, 2014).

Informant-Report Criteria Variable. Informants completed a short form of the Informant Health Status Inventory (I-HSI), which includes 10 adapted items about general emotional and physical health functioning from the RAND-36 Health Status Inventory (Hays & Morales, 2001). An example item is "During the past 4 weeks, to what extent has his/her physical health [or emotional problems] interfered with his/her normal social activities with family, neighbors, or groups?" Items were rated on 5- or 6-point scales, for example, from *poor* to *excellent*. Higher scores indicated better health. Coefficient α was .87. Informants also completed the informant version of the Health Behavior Checklist (HBC; Vickers et al., 1990), which included 10 adapted items to assess engagement in health-related behavior (e.g., exercise, lack of alcohol consumption, regular doctor visits). These 10 items were rated on a Likert type rating system ranging from (*diagree strongly*) to (*agree strongly*), coefficient α in previous research using data from the SPAN study was .84 (Wright et al., 2022). Higher scores indicated greater levels of engagement in health behaviors. In addition, informants completed the Ascertain Dementia eight-item Questionnaire (AD8; Galvin et al., 2005), an eight-item measure meant to assess issues associated with change in cognition over the past few years. Response options included "yes, a change," "no, no change," or "N/A, don't know" and an example item is "Repeats the same things over and over (questions, stories, or statement)." The cutoff for possible cognitive decline is indicated by an affirmative answer on two or more of the questions.

Statistical Analyses

Measurement invariance testing is a quantitative method to determine whether the structure, and by extension interpretability, of a measure is similar across multiple groups of individuals or time points (Putnick & Bornstein, 2016). Invariance testing can include multiple methods under various frameworks, such as structural equation modeling and item-response theory. Under the structural equation modeling framework, measurement invariance testing is usually based on confirmatory factor analysis (CFA) and examines whether the factor structure, scores, and residual error of a measure have the same meaning across groups of individuals or time. Although measurement invariance testing can be conducted in different ways, there is a broadly agreed upon method for evaluating it (Kline, 2016). Four nested models are fit, each with increasing

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.
All rights, including for text and data mining, AI training, and similar technologies, are reserved.

structural constraints. These models progressively test different levels of invariance: configural (factor structure), metric/weak (factor structure and loadings), scalar/strong (factor structure, loadings, and intercepts), and strict (factor structure, loadings, intercepts, and residuals). Invariance testing includes evaluating whether the change in goodness-of-fit indices and standard error estimates are significantly different from one level to the next. A significant change between models suggests noninvariance, or that the structure, items, or error of the measure—and therefore the interpretability of the measure—is dissimilar across groups. Full, or strict measurement invariance is achieved when there are no significant changes during sequential testing of the increasingly stricter stages (Kline, 2016)

For the present study, ESEM was conducted to test measurement invariance of the PiCD across race and time. ESEM was chosen over traditional multigroup CFA because ESEM better accounts for the fact that intercorrelations between personality traits are an expected and natural part of their structure. These natural intercorrelations reduce model fit in traditional CFA, where indicators are often assumed to be orthogonal (Marsh et al., 2014). ESEM accommodates for these intercorrelations by allowing each indicator (e.g., score, item) to load on all factors.

The present study follows Guo et al. (2017) in which (a) measurement invariance across race was conducted cross-sectionally at each of two time points and (b) longitudinal measurement invariance testing was completed for the overall sample. For the cross-sectional measurement invariance testing, the 13-step ESEM procedure described by Marsh et al. (2009) was used. This 13-step procedure includes the testing varying combinations of configural, metric, scalar, and strict invariance with varying combinations for additional constraints of latent means, variances, and covariances. Longitudinal ESEM solely included testing of configural, metric, scalar, and strict invariance without latent structural constraints.

Longitudinal measurement invariance analyses were also completed with only the Black participants to examine longitudinal PiCD invariance in Black participants specifically. However, the sample size was not big enough to conduct full ESEM for the 60 items of the PiCD with only the Black participants (n = 192). To address this issue in a pragmatic way, we conducted five item-level longitudinal measurement invariance analyses *per domain* using CFA. This significantly reduced the number of indicators (24 instead of 120), enabling the tests. These domain-level CFAs provided insight into whether the PiCD structures found for Black individuals also held over time.

The models were each evaluated via a relative fit index: the CFI and two absolute fit indices: the root-mean-squared error of approximation (RMSEA), and standardized root-mean-square residual (SRMR). Although previous literature suggests that a CFI of 0.90 or above is considered the acceptable, it was expected that the present study would produce lower CFI statistics because the analysis was conducted at the item-level versus the facet-level, which is a tougher test and introduces more error into the model. A larger number of variables is generally associated with decreases in fit indices like CFI (Kenny & McCoach, 2003). Along these lines, the PiCD has demonstrated relatively lower CFI (e.g., ∼.75) in prior studies (J. R. Oltmanns & Widiger, 2018, 2021). Regarding absolute fit indices, acceptable error would be demonstrated by RMSEA and SRMR values which fall below 0.08 (Kline, 2016).

Following the guidance of Chen (2007), a change in CFI of greater than 0.010 was used an indicator of noninvariance across models. Changes in RMSEA and SRMR were also evaluated, with a change of greater 0.015 between models indicating noninvariance. However, CFI remained the primary indicator of significant change between levels of invariance testing.

In the event of noninvariance across subsequent models, partial invariance was tested. Modification indices were used to determine noninvariant items. The constraints on the items with the largest modification indices were freed sequentially, from the most to least noninvariant items, until the CFI change across models was no longer above .01. At that point, the remaining invariance testing was conducted with the relevant parameters freed. Mplus Version 8 was used to conduct the analysis. Independent samples t tests were conducted to better estimate the true effect size of the differences between the mean scores of invariant items. This was done, as there exists no known method to examine effect size of noninvariant parameters in ESEM.

Once noninvariant parameters were identified, subsequent analyses were tested to better assess whether these item-level parameters contributed to scale-level differences and to determine whether the presence of these noninvariant parameters may be related to changes in the strength of scale-level associations with external criteria variables. To determine whether noninvariant item means were associated with domain level differences, independent samples t tests were conducted to determine whether there were significant differences between the total anankastia and negative affectivity domain-level scores for Black and White participants with and without the noninvariant items in each scale.

Correlation matrices were created for the Negative Affectivity and Anankastia scales including combined sample total scale score, combined total scale score without the noninvariant items, and scores from eight external correlates Wave 1 to determine whether the noninvariant items compromised the strength of the correlation between each scale and relevant external correlates. They were also computed for Black and White participant total scale scores with and without noninvariant item scores to determine whether the exclusion of noninvariant item scores would impact scale-level associations differently for Black and White participants. Differences in Pearson correlation coefficients under 0.10 were regarded as similar, given that difference would fall below small effect size as deemed by Cohen's guidelines (Cohen, 1992).

Results

Descriptive Statistics

The mean domain scores and comparisons are summarized in Table 1. At Wave 1, there were no significant differences in mean total domain score across race, except within the anankastia domain ($p < .001$), with Black participants scoring significantly higher than their White counterparts. The magnitude of this difference is considered small (Cohen's $d = -0.33$). Mirroring Wave 1 total domain scores at Wave 2 were not significantly different across race, except for the anankastia domain ($p < .001$), with Black participants scoring significantly higher than their White counterparts. The magnitude of this difference is considered small (Cohen's $d = -0.45$).

Measurement Invariance

Cross-Sectional Wave 1

Results from the Wave 1 13-step ESEM conducted across race are listed in Table 2. As demonstrated in Table 2, RMSEA and SRMR estimates for all 13 steps were below the .08 threshold, suggesting

Table 1
Mean Domain Score Comparisons Across Race

Scale	*M* White	*M* Black	*t*	*p*	*d*
Wave 1					
NA	25.99 (*SD* = 6.51)	26.12 (*SD* = 6.64)	−0.23	.82	−0.02
Nano26or36	21.43 (*SD* = 5.96)	21.49 (*SD* = 6.16)	−0.11	.91	−0.01
DN	23.24 (*SD* = 5.23)	22.90 (*SD* = 5.89)	0.69	.49	0.06
DT	26.25 (*SD* = 6.88)	26.75 (*SD* = 5.98)	−0.81	.42	−0.07
DL	22.92 (*SD* = 5.26)	22.70 (*SD* = 5.21)	0.45	.65	0.04
AK	38.57 (*SD* = 5.34)	40.31 (*SD* = 5.23)	−3.59	<.001	−0.33
AKno40or55	32.16 (*SD* = 4.41)	32.76 (*SD* = 4.47)	−1.49	.014	−0.14
Wave 2					
NA	25.32 (*SD* = 6.58)	25.48 (*SD* = 6.30)	−0.25	.81	−0.03
NAno26	23.08 (*SD* = 6.53)	22.83 (*SD* = 6.19)	0.38	.71	0.04
DN	23.12 (*SD* = 5.55)	21.95 (*SD* = 4.81)	2.12	.03	0.22
DT	25.70 (*SD* = 6.68)	27.64 (*SD* = 6.58)	−2.86	.004	−0.29
DL	22.61 (*SD* = 5.28)	22.08 (*SD* = 5.25)	0.98	.33	0.10
AK	38.40 (*SD* = 5.25)	40.87 (*SD* = 6.18)	−4.44	<.001	−0.45
AKno40or55	31.95 (*SD* = 4.42)	33.12 (*SD* = 5.43)	−2.45	.014	−0.25

Note. This table includes mean (*M*) domain scores and standard deviations (*SD*) for Black and White participants at both time points. The significance (*p*) and effect size estimates (Cohen's *d*) generated from independent samples *t* tests comparing mean domain scores (with and without noninvariant items) for Black and White participants at both time points is provided. Domains include negative affectivity (NA), disinhibition (DN), dissociality (DL), detachment (DT), and Anankastia (AK). A Bonferroni correction was used to determine a significant cutoff of .002, instead of .05, given that there were 23 independent samples conducted for this study.

acceptable fit. Additionally, CFI estimates ranged from .748 to .763 which, though below the desired .90 cutoff, may still be considered acceptable, as they are consistent with prior item-level analysis of the PiCD (CFI ~ .75; J. R. Oltmanns & Widiger, 2018, 2021). Configural and metric/weak invariance were established, but there was not support for full strong invariance. The change in RMSEA and SRMR between the weak and strong invariance models both fell below the cutoff of .015, but the change in CFI between weak invariance and strong invariance was .017, which is above the cutoff of .010, indicating significant change.

Given that there was significant change in CFI between the weak and strong invariance models, partial strong invariance testing was conducted. In this procedure, modification indices were used to set item intercepts free sequentially, starting with the most noninvariant item and ending with the least noninvariant item. This process was conducted until the model demonstrated a change in CFI of less than .010. Partial strong invariance was established after releasing the following items, listed from most to least noninvariant: PiCD40 "My top priority is being safe and secure" (anankastia), PiCD55 "I tend to be very cautious and careful" (anankastia), PiCD26 "Changes in my mood are

Table 2
Wave 1 Cross-Sectional ESEM Analysis Across Black and White Participants

Model	Free parameter	χ^2	*df*	CFI	ΔCFI	RMSEA (CI)	ΔRMSEA	SRMR	ΔSRMR	BIC	ΔBIC
1. Configural	708	6126.91	3,072	0.758		.053 (.052, .055)		0.049		99433.75	
2. Weak	484	6286.21	3,296	0.763	0.005	.051 (.049, .053)	0.002	0.058	0.009	98309.83	1123.91
3. Weak + Item	424	6409.89	3,356	0.758	0.005	.051 (.049, .053)	0	0.062	0.004	98116.61	193.22
4. Weak + FVCV	474	6302.19	3,306	0.763	0	.051 (.049, .053)	0	0.061	0.003	98266.68	43.15
5. Strong	428	6564.85	3,352	0.746	0.017	.052 (.051, .054)	0.001	0.061	0.003	98227.27	82.57
5.5 Partial strong	432	6455.59	3,348	0.754	−0.006	.052 (.050, .054)	0.001	0.059	−0.006	98141.05	−221.40
6. Weak + FVCV + Item	414	6423.80	3,366	0.758	0	.051 (.049, .053)	0	0.065	−0.003	98071.57	45.049
7. Strict	372	6590.41	3,408	0.748	0.006	.052 (.050, .054)	0	0.062	0.003	97964.64	176.41
8. Strong + FVCV	422	6471.68	3,358	0.754	0	.052 (.050,.054)	0	0.062	0.003	98098.19	42.86
9. Strict + FVCV	362	6604.37	3,418	0.748	0	.052 (.050,.054)	0	0.065	0.003	97919.65	221.40
10. Strong + LFMn	428	6460.60	3,352	0.754	0	.052 (.050,.054)	0	0.059	0	98121.64	19.40
11. Strict + LFMn	368	6595.09	3,412	0.748	0	.052 (.050,.054)	0	0.062	0	97944.99	19.65
12. Strong + LFMn + FVCV	418	6476.44	3,362	0.754	0	.052 (.050,.053)	0	0.063	0.004	98077.86	63.18
13. Strict + LFMn + FVCV	358	6608.90	3,422	0.748	0	.052 (.050,.054)	0	0.065	0.003	97899.17	65.47

Note. This table summarizes the model information and change in fit indices across the thirteen invariance models for the cross-sectional ESEM conducted across race at Wave 1. The absolute values of the change in fit indices are provided in the change columns. The FCVC acronym alludes to a constraint of factor variances and covariances across groups. The LFMn acronym means that latent factor means are constrained across groups. Last, item refers to the residuals being constrained across groups. ESEM = exploratory structural equation modeling; CFI = comparative fit index; RMSEA = root-mean-square error of approximation; CI = confidence interval values; SRMR = standardized root-mean-square residual; BIC = Bayesian information criterion.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.
All rights, including for text and data mining, AI training, and similar technologies, are reserved.

unrelated to what is happening in my life" (negative affectivity), and PiCD36 "am thin-skinned" (negative affectivity).

Given that there does not currently exist any method to determine the effect size of these noninvariant items via ESEM, independent samples *t* tests were conducted to determine the effect size of the differences between Black and White participant scores on each of the invariant items. The independent samples *t* tests for each of the four freed items were significant ($p < .001$). The effect sizes for PiCD55 (Cohen's $d = -.520$) and PiCD40 (Cohen's $d = -.742$) were both medium, suggesting that the difference in average scores between Black and White participants on these items, were moderate. For both items, Black participants scores, were on average, higher than White participants. The effect sizes for PiCD36 (Cohen's $d = .321$) and PiCD26 (Cohen's $d = -.434$) were both small, suggesting that the magnitude of the differences between the average Black and White participant scores on these items was relatively small in nature. Of note, Black participants, on average, scored lower than White participants on PiCD36 and Black participants scored higher than White participants on PiCD26.

Once partial strong invariance was established, the subsequent eight steps were conducted with the intercepts for PiCD40, PiCD55, PiCD26, and PiCD36 freed. As shown in Table 2, throughout the remaining eight steps of cross-sectional invariance testing, the changes in between each subsequent step of testing for RMSEA and SRMR all fell below the cutoff .015, and the changes in CFI all fell below the cutoff of .010. In sum, at Wave 1, partial strong invariance of the PiCD was found across Black and White participants.

Cross-Sectional Wave 2

Results from the Wave 2 13-step ESEM conducted across race are listed in Table 3. Once again, the RMSEA and SRMR estimates for all 13 steps were below the .08 threshold, suggesting acceptable fit. CFI estimates ranged from .734 to .752, once again falling in line with previous estimates (CFI ~ .75; J. R. Oltmanns & Widiger, 2018, 2021) with the PiCD. Configural and metric/weak invariance were established, but, again full strong invariance was not found: The change in RMSEA and SRMR between the weak and strong invariance models both fell below the cutoff of .015, but the change in CFI between weak invariance and strong invariance was, again, .017, which falls above the cutoff of .010, indicating significant change.

Partial strong invariance testing, which included setting item intercepts free from most to least invariant, until the change in CFI from weak to strong invariance fell below .010, was conducted. Partial strong invariance was established after releasing the following items, listed from most to least noninvariant: PiCD40 "My top priority is being safe and secure" (anankastia), PiCD55 "I tend to be very cautious and careful" (anankastia), and PiCD26 "Changes in my mood are unrelated to what is happening in my life" (negative affectivity). Subsequently, independent samples *t* tests were conducted to determine the effect size of the differences between Black and White participant scores on each of the invariant items. The independent samples *t* tests for all of the three freed items were significant ($p < .001$). The effect sizes for PiCD26 (Cohen's $d = -.514$), PiCD55 (Cohen's $d = -.673$), and PiCD40 (Cohen's $d = -.764$) were all medium, suggesting that the difference in average scores between Black and White participants were moderate. For all three items, Black participants scores were, on average, higher than White participants.

After establishing partial strong invariance, the subsequent eight steps were conducted with the intercepts for PiCD40, PiCD55, and PiCD26 freed. As shown in Table 3, throughout the remaining eight steps of cross-sectional invariance testing, the changes in between each subsequent step of testing for RMSEA and SRMR all fell below the cutoff .015, and the changes in CFI all fell below the cutoff of .010.

Taken together, these results suggest that there is partial strong invariance across Black and White participants at Wave 2. The pattern of results found at Wave 2 provide a near perfect replication of those found at Wave 1, suggesting that the significant similarities and differences in PiCD structure found across Black and White participants may generally hold across time. However, these analyses were cross-sectional.

Table 3
Wave 2 Cross-Sectional ESEM Analysis Across Black and White Participants

Model	Free parameter	χ^2	*df*	CFI	ΔCFI	RMSEA (CI)	ΔRMSEA	SRMR	ΔSRMR	BIC	ΔBIC
1. Configural	708	6092.85	3,072	0.752		.052 (.050, .054)		0.049		108752.31	
2. Weak	484	6340.21	3,296	0.75	0.002	.050 (.048, .052)	0.002	0.058	0.009	107678.21	1074.11
3. Weak + Item	424	6455.10	3,356	0.745	0.005	.050 (.048, .052)	0	0.063	0.005	107487.95	190.26
4. Weak + FVCV	474	6350.43	3,306	0.75	0	.050 (.048, .052)	0	0.06	0.002	107629.36	141.41
5. Strong	428	6594.14	3,352	0.733	0.017	.051 (.050, .053)	0.001	0.06	0.002	107564.11	114.10
5.5 Partial strong	431	6502.61	3,349	0.741	0	.051 (.049, .053)	0.001	0.059	−0.002	107490.10	48.58
6. Weak + FVCV + Item	414	6466.642	3,366	0.745	0.005	.050 (.048, .052)	0	0.064	0.006	107439.256	238.95
7. Strict	371	6624.50	3,409	0.736	0.005	.051 (.049, .053)	0	0.063	0.004	107311.69	178.41
8. Strong + FVCV	421	6512.90	3,359	0.741	0	.051 (.049, .052)	0	0.061	0.002	107441.53	48.58
9. Strict + FVCV	361	6635.02	3,419	0.736	0	.051 (.049, .053)	0	0.064	0.001	107262.96	48.73
10. Strong + LFMn	427	6528.64	3,353	0.739	0.002	.051 (.049, .053)	0	0.06	0.001	107492.21	2.11
11. Strict + LFMn	367	6649.43	3,413	0.734	0.002	.051 (.049, .053)	0	0.063	0	107312.50	0.81
12. Strong + LFMn + FVCV	417	6535.50	3,363	0.739	0.002	.051 (.049, .053)	0	0.061	0.002	107440.16	49.95
13. Strict + LFMn + FVCV	357	6657.014	3,423	0.734	0.002	.051 (.049, .053)	0	0.064	0.001	107260.994	50.70

Note. This table summarizes the model information and change in fit indices across the thirteen invariance models for the cross-sectional ESEM conducted across race at Wave 2. The absolute values of the change in fit indices are provided in the change columns. The FCVC acronym alludes to a constraint of factor variances and covariances across groups. The LFMn acronym means that latent factor means are constrained across groups. Last, item refers to the residuals being constrained across groups. ESEM = exploratory structural equation modeling; CFI = comparative fit index; RMSEA = root-mean-square error of approximation; CI = confidence interval values; SRMR = standardized root-mean-square residual; BIC = Bayesian information criterion.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.
All rights, including for text and data mining, AI training, and similar technologies, are reserved.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.
All rights, including for text and data mining, AI training, and similar technologies, are reserved.

Longitudinal: Whole Sample ESEM

Findings from the four-step longitudinal ESEM testing involving the full sample are listed in Table 4. For all four steps, the RMSEA and SRMR estimates fell below the threshold of .08, suggesting acceptable fit. CFI estimates ranged from .826 to .832, falling below the general cutoff of .090. However, these CFI estimates were higher than previous PiCD CFI estimates (CFI ~ .75; J. R. Oltmanns & Widiger, 2018, 2021). As demonstrated in Table 4, the change in RMSEA and SRMR fell below .015 and change in CFI fell below .010 for all four steps in the longitudinal invariance testing. Thus, using the full sample of participants, strict measurement invariance of the PiCD was found longitudinally across the two time points about 2 years apart.

Longitudinal: Black Participants CFA

Given that strict measurement invariance was not found cross-sectionally, longitudinal invariance was also assessed for the Black participants specifically. Although the sample included close to 200 Black participants at both time points, full longitudinal ESEM tested could not be completed, given that there were 120 indicators (i.e., PiCD items). As such, longitudinal CFA was completed for each of the five domains with the Black participants. Findings are listed in detail in the Supplemental Table S1–S5. CFI ranged from 0.708 (anankastia) to .908 (detachment), suggesting acceptable-to-good model fit across domains, considering past studies presenting CFI at the item level for the PiCD (J. R. Oltmanns & Widiger, 2018). RMSEA estimates generally fell below the cutoff of .08, however, and some SRMR estimates fell above .08, especially on the models generated for the anankastia (highest SRMR .107) and disinhibition (highest SRMR .093) domains. This may stem from the potentially problematic use of CFA instead of ESEM with personality constructs that are often interrelated (cf. Marsh et al., 2014). Additionally, results suggest that full strict measurement invariance was found for all five domains across time for Black participants. These finding suggest that the structures found for Black participants holds over time.

Noninvariant Item Mean Impact on Scale Differences

It is important to investigate whether differences between Black and White participants in the four noninvariant item intercepts extend to differences at the scale level. To probe this issue, anankastia and negative affectivity domain scores were computed excluding the noninvariant items for the Anankastia (PiCD item 40 and 55) and Negative Affectivity (PiCD 26 and 36) scales. Subsequently, independent samples *t* tests were completed to determine whether there were significant differences in domain score across Black and White participants (presented in Table 1).

At both Waves 1 and 2, Black participants scored significantly higher on anankastia ($p < .001$) when all 12 items were included in the scale score. In contrast, at both waves, there were not significant differences when PiCD 40 and PiCD 55 were excluded from the Anankastia scale score. For negative affectivity, the difference between Black and White average total scale score, including the noninvariant items, was not significantly different at both Waves 1 and 2. Independent samples *t* tests computed at both waves comparing total scale score without PiCD 26 (and PiCD 36 at Wave 1) indicated that the Negative Affectivity scale score was (still) not significantly different across Black and White participants when items 26 and 36 were excluded.

Noninvariant Item Mean Impact on Correlations With External Criteria

It is also important to investigate whether differences between Black and White participants in the four noninvariant item intercepts extend to differences at the scale level in terms of construct validity. To probe this issue, scale scores for Negative Affectivity and Anankastia both with and without the noninvariant items at Wave 1 were correlated with several external correlates to assess differences in the extent of predictive validity (Table 5). The associations between Negative Affectivity combined total scale scores and scores from measures of external outcomes like health status, satisfaction with life, and depression were very similar (difference in *r* less than 0.10; less than small effect size), regardless of whether the noninvariant items were included or excluded from the total scale score. Comparisons between total Negative Affectivity scores for White and Black participants, separately, also demonstrated similar associations with external correlate scores with and without the two noninvariant item scores.

Of note, the strength of the correlations that the Negative Affectivity total scores (with and without the noninvariant items) for Black participants demonstrated with scores from the AD8 and ISI were lower (greater than 0.10 difference in *r*) than that of the total scores from the combined sample and White participants. Additionally, although the correlations between combined sample, Black, and White total scores and the P-HSI and I-HSI scores were similar, correlations between Black participant Negative Affectivity scores and these P- and I-HSI scores were not significant ($p > .001$). This difference in

Table 4
Longitudinal ESEM for the Whole Sample

Model	Free parameter	χ^2	*df*	CFI	ΔCFI	RMSEA (CI)	ΔRMSEA	SRMR	ΔSRMR	BIC	ΔBIC
1. Configural	784	11896.41	6,596	0.832		0.031 (.030, .032)		0.043		202564.07	
2. Weak	560	12103.81	6,820	0.832	0	0.030 (.029, .031)	0.001	0.045	0.002	201317.90	1246.12
3. Strong	500	12250.37	6,880	0.829	0.003	0.030 (.029, .031)	0	0.046	0.001	201063.07	254.83
4. Strict	440	12410.73	6,940	0.826	0.003	0.030 (.030, .031)	0	0.047	0.001	200881.43	181.63

Note. This table summarizes the model information and change in fit indices across the four invariance models for the whole sample ESEM conducted across Waves 1 and 2. The absolute values of the change in fit indices are provided in the change columns. ESEM = exploratory structural equation modeling; CFI = comparative fit index; RMSEA = root-mean-square error of approximation; CI = confidence interval values; SRMR = standardized root-mean-square residual; BIC = Bayesian information criterion.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.
All rights, including for text and data mining, AI training, and similar technologies, are reserved.

Table 5
Noninvariant Item Mean Impact on Correlations With External Criteria

Scale	AD8	HBC	I-HSI	BDI-II	ISI	SNQ	SWLS	P-HSI
AKtotal	−0.04	0.12	0.01	0.06	0.08	−0.06	0.00	−0.03
AKtotalno40or55	−0.06	0.12	−0.02	0.00	0.03	−0.03	0.05	0.03
AKwhite	−0.04	0.12	0.04	0.04	0.07	−0.08	0.01	−0.02
AKwhiteno40or55	−0.06	0.11	−0.01	−0.02	0.02	−0.05	0.06	0.03
AKblack	−0.11	0.15	−0.13	0.07	0.08	0.08	0.06	0.06
AKblackno40or55	−0.12	0.13	−0.12	0.05	0.04	0.07	0.08	0.08
NAtotal	0.08	−0.04	**0.27**	**0.52**	**0.35**	**−0.29**	**−0.40**	**−0.21**
NAtotalno26or36	0.08	−0.05	**0.27**	**0.54**	**0.37**	**−0.30**	**−0.41**	**−0.21**
NAwhite	0.10	−0.01	**0.25**	**0.51**	**0.32**	**−0.29**	**−0.41**	**−0.22**
NAwhiteno40or55	0.10	−0.03	**0.24**	**0.52**	**0.33**	**−0.29**	**−0.42**	**−0.23**
NAblack	−0.02	−0.07	0.26	**0.54**	**0.41**	**−0.27**	**−0.33**	−0.14
NAblackno40or55	−0.02	−0.08	0.27	**0.56**	**0.42**	**−0.28**	**−0.35**	−0.14

Note. This table summarizes findings from the correlation matrices including correlations between the Negative Affectivity scale score, Anankastia scale score, Anankastia scale score without item 40 or 55 scores, and Negative Affectivity scale score without PiCD item 26 or 36 scores with eight external self-report and informant-report external outcomes at Wave 1. Correlation coefficients are in bold for those associations for which the $p < .001$. AD8 = Ascertain Dementia eight-item questionnaire; HBC = Health Behavior Checklist; I-HSI = Health Status Inventory Informant; BDI-II = Beck Depression Inventory; ISI = Insomnia Severity Index; SNQ = social network questionnaire; SWLS = Satisfaction With Life Scale; P-HIS = Health Status Inventory Participant; AK = Anankastia; NA = negative affectivity; PiCD = Personality Inventory for *ICD-11*.

significance may be attributable to the fact that Black participants had a much lower sample size than the combined group and White participant. Our analyses demonstrate that these correlational differences are likely not due to measurement noninvariance of the two item means in the scale.

There were not statistically significant correlations between the Anankastia scale and the external correlates (mirroring prior research; Mays et al., 2024). However, Anankastia combined, White, Black total scale scores showed Pearson correlation coefficients that were similar in value regardless of whether the noninvariant item scores for items 26 and 36 were excluded. Interestingly, there were notable differences between the r values between Anankastia combined or White participant total domain scores versus that of Black participant total scores and the following external correlates: I-HSI and SNQ. Again, our analyses demonstrate that these correlational differences are not due to measurement noninvariance of two item means in the scale.

Discussion

The *ICD-11*'s shift to a dimensional model of personality disorder diagnosis necessitates the creation and validation of assessment tools that can provide fair assessment information for a wide variety of people. The present study provides the first invariance test of the PiCD, a measure of the maladaptive personality trait domains from the *ICD-11*, across a community sample of Black and White older adults. Previous research suggests that strong invariance is required to assume similar interpretability across groups (Luong & Flake, 2023).

The present study provides evidence that there are three items in the PiCD (two from anankastia, and one from negative affectivity) that were noninvariant across race at two time points (one negative affectivity item was noninvariant at the first time point only). Black older adults tended to score higher on two anankastia items, scored higher on one negative affectivity item across both time points, and scored lower on the one negative affectivity item that was noninvariant only at the first time point. Our follow-up analyses indicated that the two noninvariant anankastia means were responsible for the observed scale-level difference in Anankastia. There were not observed scale-level differences in negative affectivity. Our follow-up analyses also indicated that correlations with several external correlates were similar for both Black and White participants with and without the items with noninvariant mean levels. This provides a first test of the significance of the differences in mean levels of the three/four items, but future research should continue to evaluate the correlations of the scales with and without these items with further external correlates, and in clinical samples. Researchers should be aware of these mean differences, especially in the Anankastia scale. Future research will need to investigate the significance of these differences in younger populations and other racial groups. Although future research is needed, these findings provide initial evidence that these three/four items should perhaps be revised in future editions of the PiCD.

Findings from the present study indicated that partial strong invariance was found across Black and White participant PiCD scores. From a bird's eye view, the PiCD scores demonstrated an impressive amount of invariance—configural and structural invariance across Black and White older adults, indicating the same factor structure and factor loadings. The PiCD scores demonstrated intercept (mean level) invariance for the other 56 out of the 60 items. When the four intercepts that were noninvariant were held constant, the PiCD scores then demonstrated invariance of the residuals at the item-level. These results are impressive given that testing was completed at the item-level, and demonstrate at least equal measurement invariance than the PID-5, in which previous research has indicated factor differences across Black and White Americans when testing was conducted at the facet-level (Bagby et al., 2022) and scale-level mean noninvariance across race for two PID-5 scales when testing at the facet level (Freilich et al., 2023). However, the present study highlighted four important mean item-level differences for the PiCD across Black and White old adult participants that require further investigation.

PiCD Noninvariance Findings

There were notable differences in the mean scores on four of the 60 PiCD items found across Black and White Americans. Two of the items with noninvariant means included PiCD40 "My top priority is

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.
All rights, including for text and data mining, AI training, and similar technologies, are reserved.

being safe and secure" (anankastia) and PiCD55 "I tend to be very cautious and careful" (anankastia). Black participants, on average, scored higher on both items. Given the ongoing legacies of racism in the United States and resulting threats to physical and psychological safety, it is enlightening that PiCD item 40, "My top priority is being safe and secure" was higher among Black versus White Americans. A prioritization of safety and security is consistent with literature finding that Black Americans feel less safe than White Americans (Jordan & Gabbidon, 2010). A disproportionate exposure to violence could influence perception of safety (Thomas et al., 2016). Furthermore, the systems in place assumed to increase safety for all of society may function differently for Black Americans. For example, there is evidence that placing officers in schools is associated with lower perception of safety among Black versus White students (Theriot & Orme, 2016) and that in general, police presence may signal more threat (Najdowski et al., 2015) and may be associated with less safety (Wheelock et al., 2019) for Black versus White Americans. It is worthwhile to note that these reactions are grounded in historical (Petersen & Ward, 2015) and ongoing (F. Edwards et al., 2019) disparities in police use of violence toward Black Americans. Furthermore, legacies of racism extend to other contexts that can reduce Black Americans sense of safety, including in medical settings (Boakye & Prendergast, 2024), school settings (E. C. E. C. Edwards, 2021), and other professional settings (Purdie-Vaughns et al., 2008). Taken together, there is evidence that these items may reflect an understandable and adaptive coping with a disproportionally threatening environment for Black Americans. Conversely, White Americans may endorse this item less due, at least in part, to their historical privilege in terms of safety and security in the U.S.

Our results also indicated that on average, Black versus White Americans scored higher on Item 55, "I tend to be very cautious and careful." Related to the threats to safety discussed earlier, one possibility is that engaging in cautious and careful behavior is a way to cope with lack of safety or threat in the environment. Although hypervigilance is different from being cautious and careful in that it is more about being alert and oriented to threat in the environment, it is possible that hypervigilance may be driving the disproportionate tendency to be cautious and careful. Indeed, hypervigilance is a consequence of trauma (Smith et al., 2019) and racial trauma in particular (Carter & Kirkinis, 2021) and is thought to be a way to protect oneself from threat. Indeed, there is evidence that Black Americans attend to and seek identity safety and threat cues in the environment, such as in a medical and other settings (Derricks et al., 2023). This can be conceptualized as a way to cautiously avoid risk. This is consistent with Brownlow (2023) model of "culturally compelled" coping which contextualizes Black Americans' coping in the context of racism and highlights distinct aspects, including hypersensitivity and avoiding unplanned risks. Brownlow argues it is critical to understand sensitivity to threat as being the result of adaptive coping to threatening environments as opposed to a hardwired biological reality for Black Americans. Further, and consistent with our findings, Brownlow argues that avoiding unplanned risks is a way to cope with threat and is a strategy often encouraged in the racial socialization process (e.g., parents transmitting racially relevant information about what it means to be a part of a racial group). Thus, Black Americans scoring higher on Item 55 compared with their White counterparts may reflect an adaptive coping response in the context of real racialized and other threats in the environment.

Additionally, mean scores for negative affectivity items PiCD26 "Changes in my mood are unrelated to what is happening in my life" and PiCD36 "am thin-skinned" were also noninvariant across Black and White participants. On average, Black participants scored higher than White participants on item 26. This difference may be influenced by systemic stressors experienced by many Black individuals in society, such as racial discrimination and socioeconomic challenges (Jones, 2023), which may lead to a form of perceived emotional resilience where Black individuals mentally disconnect one's internal mood states from external events. This aligns with the racial healing perspective (French et al., 2020), which notes that one principle of coping with racism, labeled strength and resistance, is maintaining a commitment to living a joyful life despite the challenges of racism and systemic oppression. This strategy, therefore, could function as a protective mechanism to reduce the psychological burden of these external pressures, resulting in a perception that their emotional state operates independently of external circumstances. Furthermore, this mindset may be influenced by cultural attitudes toward stoicism and self-reliance within the Black community (Dennis & Zolnikov, 2024; Scott, 2003), emphasizing an internal locus of control (Brown et al., 2017; James, 2017) that may reinforce the perception that one's mood is not directly tied to external events.

Relatedly, on item 36, "I am thin-skinned," which uses an idiom, Black participants scored lower on average compared with White participants. Given the universality of the word in English for hundreds of years, we do not believe differences have to do with familiarity with the idiom. The difference may be attributable to the cultural and historical factors that have shaped how Black individuals respond to criticism and adversity, which is often passed down through racial socialization (Hughes et al., 2006). Over generations, confronting and navigating systemic racism and oppression could have fostered a "thick-skinned" attitude as a form of emotional resilience. Additionally, cultural norms within the Black community may prioritize collective strength and support (Selvanathan et al., 2023), which could further buffer individuals against negative feedback and external criticism, leading to a lower self-perception of being thin skinned. Thus, for this particular item, there may be genuine differences in how older adults in this sample perceive and endorse their own vulnerability or sensitivity. There is some evidence that White Americans perceive Black Americans as having "thicker skin" than White individuals (Hoffman et al., 2016), which may provide some insight on the way White Americans view themselves. Furthermore, a growing body of work has documented the ways in which racial privilege may influence the degree to which White individuals tolerate emotions (i.e., White fragility; Ford et al., 2022) which may provide some context for the relatively higher endorsement of thin skin among White individuals.

Comparison With the PID-5

Research with the PID-5 raised concerns regarding the PID-5's applicability, even at the configural level, across Black and White American college students (Bagby et al., 2022; Orjiakor et al., 2023). The present study suggested that PiCD scores demonstrate partial strong invariance across Black and White older adult participants. It should be noted that the invariance testing for the PID-5

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.
All rights, including for text and data mining, AI training, and similar technologies, are reserved.

was conducted at the facet level, whereas, for the PiCD, the test was done at the item-level, which is a more rigorous test. Also of importance, different sample demographics—older adults versus college students—may play a part in differences across these measures. The present study suggested that the PiCD and PID-5 scores may demonstrate different extents of racial invariance across Black and White Americans. Further research with similar populations may provide further insight on these differences.

Additionally, the content of these measures may align or differ in important ways that impact the differential patterns of racial invariance found for the PID-5 and PiCD. In fact, Freilich et al. (2023) found facet scale intercept (i.e., mean) differences across Black and White Americans for the submissiveness and suspiciousness PID-5 scale scores. Suspiciousness items assess concern with safety with a focus on paranoia regarding other's intentions. It is possible that the similar content of these items and the two noninvariant items in the PiCD Anankastia scale account for shared findings of racial noninvariance across both measures. Items from the PID-5 submissiveness scale assesses dependence or deference to others. This facet scale is part of the detachment domain. In contrast, the PiCD detachment domain focuses primarily on lack of interpersonal warmth or engagement, which may contribute to racial invariance for this domain. Further testing is required to better understand how sample and content level differences may impact racial invariance across Black and White Americans.

Implications and Future Directions

It is important to consider these differences and the potential context in which they may arise to avoid making biased conclusions about Black participants. This is especially important given the differential mean levels of the four items found for Black older adults was replicated across the two cross-sectional analyses, and the longitudinal analysis indicated that the differential mean levels for Black and White participants holds across time.

Specifically, higher average scores for Black participants on anankastia items 40 and 55 may reflect normative responses to cultural circumstances in our sample as opposed to pathological fears regarding safety or control. Our current analyses show that significant differences in the observed Anankastia scale scores are explained by differences in higher mean scores of PiCD items 40 and 55 in Black participants. This pattern of results show that the mean difference between Black and White participants on the Anankastia scale is explained by these two items. Correlations of the Anankastia scale score with and without items 40 and 55 for the combined sample, White participants, and Black participants, separately, showed similar relations with relevant external criteria scores. Correlation analyses comparing the scale with and without the items indicates that the construct validity of the Anankastia scale may not compromised by the mean level differences. Importantly, however, notable differences were found in the strength of correlations between Black participant total scores (with and without the two noninvariant item scores) and that of the White participant total scale scores and the SNQ and I-HSI scores. Further testing of substantive racial differences will need to be conducted to replicate these findings and better understand whether and why any differences in predictive validity at the domain score-level may occur across Black and White older adults. However, our analyses indicate that these differences are not due to the item-level mean measurement noninvariance found for these items.

The observed negative affectivity domain scores were not significantly different across Black and White participants—with or without the two noninvariant items. Correlational tests between the Negative Affectivity scale scores (combined, White, and Black participant scores) and several external outcome scores were also similar for this scale with and without the two noninvariant items, once again indicating that the mean level differences of the two noninvariant negative affectivity item scores may not compromise criterion validity at the scale level. Of note, however, differences were found between the extent of scale predictive validity of the Black participant total scores (with and without the two noninvariant item scores) and that of the White participant total scale scores and the AD8 and ISI scores. Again, further replications and extensions on these findings are necessary with a larger sample of Black participants. Given that this study briefly investigated associations between scale scores and eight external outcomes, there is also need for tests with other criteria scores. Again, however, our findings indicate that the differences in the correlations across race are not due to the mean level item measurement noninvariance of the items on the Negative Affectivity scale.

Although we found two items with noninvariant mean levels each in the anankastia and negative affectivity domains (only one in negative affectivity at Wave 2), there are items with similar content included in each scale that had invariant mean levels. For example, the noninvariant anankastia item "I tend to be very cautious and careful" is complemented with the items "I always choose the safest option" and "I do not take risks." This indicates that the content validity of the scales may not be completely sacrificed. This face validity, combined with the lack of correlational differences in the scales with and without the four items with noninvariant means, indicates that—at least with the external criteria tested herein—construct validity may not be sacrificed with the exclusion of these items.

Future research is needed to evaluate the construct validity of the Anankastia and Negative Affectivity scales including the four items with noninvariant mean levels across Black and White older adults. This line of research should focus on assessing the impact of removing the noninvariant items on overall measure structure and interpretability across Black and White Americans. Item response theory analyses may be helpful in examining differential item functioning. Although the PiCD did demonstrate an impressive amount of configural, structural, intercept, and residual invariance at the item-level, the meaningful differences found in the four items' mean levels across racial groups indicate that it may not be appropriate to interpret these Anankastia and Negative Affectivity scores on the PiCD across Black and White American older adults. It is possible that revisions to these items or replacement would be helpful for a revised PiCD. However, future research and replication will be needed to determine this.

Strengths

This study offered the first test of the measurement invariance of the PiCD across Black and White American participants. This study possessed several strengths: First, the use of a large, representative sample size allowed for useful analyses from which meaningful conclusions could be drawn regarding structural invariance in a community sample and important areas for further testing to develop the measure for more accurate assessment with future participants. Additionally, this study conducted measurement invariance testing at the item level, allowing for more fine-grained analysis and understanding of differences that could impact the usage of these measures. Last, the study included both cross-sectional and longitudinal analyses at two time points across 2 years, providing the opportunity to not only investigate the presence of differences across Black and White American older adults, but the consistency of these differences across time.

Constraints on Generality

This is the first step into the study of the PiCD with diverse populations. Conclusions should be extended with caution, as the current sample, though representative of the older adult population in St. Louis, Missouri, do not represent other community or racial demographics. There is a need for further measurement invariance research with PiCD scores across Black and White Americans from other areas of the United States to determine whether and how findings differ based on region, social context and opportunity, and so on. Additionally, invariance testing will need to be conducted with other racial/ethnic groups in America and across the world to formally assess the interpretability of the PiCD across diverse groups of individuals. Research must also expand to include the use of clinical samples to determine how the measure functions clinically prior to use in those settings.

Last, explicit testing to better understand the causes or correlates that may be contributing to differences between Black and White Americans on the PiCD is needed. The present study does not empirically investigate what factors may be contributing to these differences, so there is a need for explicit investigation determining what mechanisms are empirically associated with differences in mean item scores across Black and White Americans. That is, differences are likely explained by broader environmental, social, and/or cultural issues and it would be important to empirically identify them.

Conclusions

The present study offered important insight into the use of PiCD scores across Black and White American older adults. In the present community sample of Black and White older adults, it was found that the PiCD demonstrated similar factor structure, loadings, and item means for 56 out of 60 items across our Black and White participants. Four items (two from anankastia, two from negative affectivity) were found to be noninvariant in mean levels at Time 1, and three of those four items at Time 2 (the same two from anankastia, and one of the two from negative affectivity at Time 1). It is important to conduct further research on the anankastia and negative affectivity items and domains in future research, prior to use of the PiCD scores with Black American participants. Our study indicates the value of continuing research in this area with samples of different demographic backgrounds to add valuable insight into the use of the PiCD.

In the investigation of the measurement invariance of the PiCD scores, it is not only important to establish whether the expected underlying structure of the PiCD applies differentially to White and Black American participants, but to investigate *why* these differences may occur. Because measurement noninvariance was found for four item means, possible explanatory factors such as traumatic event exposure, differential neighborhood conditions, physical or mental health disparities, and the experiences of discrimination between the groups of Black and White participants offer important areas for follow up.

References

Asadi, S., Cunningham, T. J., Morgan, T. A., Zimmerman, M., & Rodriguez-Seijas, C. (2024). Examining measurement invariance in the Personality Inventory for *DSM*-5 Brief Form across sexual and gender minority status. *Assessment*, *31*(3), 678–697. https://doi.org/10.1177/10731911231176449

Bagby, R. M., Keeley, J. W., Williams, C. C., Mortezaei, A., Ryder, A. G., & 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. https://doi.org/10.1037/pas0001085

Bastien, C. H., Vallières, A., & Morin, C. M. (2001). Validation of the insomnia severity index as an outcome measure for insomnia research. *Sleep Medicine*, *2*(4), 297–307. https://doi.org/10.1016/s1389-9457(00)00065-4

Beck, A. T., Steer, R. A., Ball, R., & Ranieri, W. (1996). Comparison of Beck Depression Inventories-IA and -II in psychiatric outpatients. *Journal of Personality Assessment*, *67*(3), 588–597. https://doi.org/10.1207/s15327752jpa6703_13

Becker, L. G., Asadi, S., Zimmerman, M., Morgan, T. A., & Rodriguez-Seijas, C. (2023). Is there a bias in the diagnosis of borderline personality disorder among racially minoritized patients? *Personality Disorders*, *14*(3), 339–346. https://doi.org/10.1037/per0000579

Boakye, P. N., & Prendergast, N. (2024). "There is nothing to protect us from dying": Black women's perceived sense of safety accessing pregnancy and intrapartum care. *Nursing Inquiry*, *31*(3), Article e12638. https://doi.org/10.1111/nin.12638

Brown, D. L., Rosnick, C. B., & Segrist, D. J. (2017). Internalized racial oppression and higher education values: The mediational role of academic locus of control among college African American men and women. *Journal of Black Psychology*, *43*(4), 358–380. https://doi.org/10.1177/0095798416641865

Brownlow, B. N. (2023). How racism "gets under the skin": An examination of the physical- and mental-health costs of culturally compelled coping. *Perspectives on Psychological Science*, *18*(3), 576–596. https://doi.org/10.1177/17456916221113762

Carter, R. T., & Kirkinis, K. (2021). Differences in emotional responses to race-based trauma among black and white Americans. *Journal of Aggression, Maltreatment & Trauma*, *30*(7), 889–906. https://doi.org/10.1080/10926771.2020.1759745

Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling*, *14*(3), 464–504. https://doi.org/10.1080/10705510701301834

Clark, L. A. (2007). Assessment and diagnosis of personality disorder: Perennial issues and an emerging reconceptualization. *Annual Review of Psychology*, *58*, 227–257. https://doi.org/10.1146/annurev.psych.57.102904.190200

Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*(1), 155–159. https://doi.org/10.1037/0033-2909.112.1.155

Cruitt, P. J., & Oltmanns, T. F. (2019). Unemployment and the relationship between borderline personality pathology and health. *Journal of Research in Personality*, *82*, Article 103863. https://doi.org/10.1016/j.jrp.2019.103863

Dennis, J., & Zolnikov, T. R. (2024). Demystifying male emotional development: Exploring the implications of forced gender roles on black American males. *Journal of Black Psychology*, *50*(1), 68–90. https://doi.org/10.1177/00957984231191868

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.
All rights, including for text and data mining, AI training, and similar technologies, are reserved.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.
All rights, including for text and data mining, AI training, and similar technologies, are reserved.

Derricks, V., Johnson, I. R., & Pietri, E. S. (2023). Black (patients') lives matter: Exploring the role of identity-safety cues in healthcare settings among black Americans. *Journal of Health Psychology*, *28*(1), 30–47. https://doi.org/10.1177/13591053221090850

Diener, E., Emmons, R. A., Larsen, R. J., & Griffin, S. (1985). The satisfaction with life scale. *Journal of Personality Assessment*, *49*(1), 71–75. https://doi.org/10.1207/s15327752jpa4901_13

Edwards, E. C. (2021). Centering race to move towards an intersectional ecological framework for defining school safety for black students. *School Psychology Review*, *50*(2–3), 254–273. https://doi.org/10.1080/2372966X.2021.1930580

Edwards, F., Lee, H., & Esposito, M. (2019). Risk of being killed by police use of force in the United States by age, race-ethnicity, and sex. *Proceedings of the National Academy of Sciences of the United States of America*, *116*(34), 16793–16798. https://doi.org/10.1073/pnas.1821204116

Erford, B. T., Johnson, E., & Bardhoshi, G. (2016). Meta-analysis of the English version of the Beck Depression Inventory–Second Edition. *Measurement and Evaluation in Counseling and Development*, *49*(1), 3–33. https://doi.org/10.1177/0748175615596783

Ford, B. Q., Green, D. J., & Gross, J. J. (2022). White fragility: An emotion regulation perspective. *American Psychologist*, *77*(4), 510–524. https://doi.org/10.1037/amp0000968

Freilich, C. D., Palumbo, I. M., Latzman, R. D., & Krueger, R. F. (2023). Assessing the measurement invariance of the Personality Inventory for *DSM-5* across black and white Americans. *Psychological Assessment*, *35*(9), 721–728. https://doi.org/10.1037/pas0001255

French, B. H., Lewis, J. A., Mosley, D. V., Adames, H. Y., Chavez-Dueñas, N. Y., Chen, G. A., & Neville, H. A. (2020). Toward a psychological framework of radical healing in communities of color. *The Counseling Psychologist*, *48*(1), 14–46. https://doi.org/10.1177/0011000019843506

Galvin, J. E., Roe, C. M., Powlishta, K. K., Coats, M. A., Muich, S. J., Grant, E., Miller, J. P., Storandt, M., & Morris, J. C. (2005). The AD8: A brief informant interview to detect dementia. *Neurology*, *65*(4), 559–564. https://doi.org/10.1212/01.wnl.0000172958.95282.2a

Guo, B., Kaylor-Hughes, C., Garland, A., Nixon, N., Sweeney, T., Simpson, S., Dalgleish, T., Ramana, R., Yang, M., & Morriss, R. (2017). Factor structure and longitudinal measurement invariance of PHQ-9 for specialist mental health care patients with persistent major depressive disorder: Exploratory structural equation modelling. *Journal of Affective Disorders*, *219*, 1–8. https://doi.org/10.1016/j.jad.2017.05.020

Hays, R. D., & Morales, L. S. (2001). The RAND-36 measure of health-related quality of life. *Annals of Medicine*, *33*(5), 350–357. https://doi.org/10.3109/07853890109002089

Heragu, P., & Oltmanns, J. R. (2025, April 17). *Longitudinal Measurement Invariance of the Personality Inventory for the ICD-11 across black and white older adults*. https://osf.io/s9b4x

Hoffman, K. M., Trawalter, S., Axt, J. R., & Oliver, M. N. (2016). Racial bias in pain assessment and treatment recommendations, and false beliefs about biological differences between blacks and whites. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(16), 4296–4301. https://doi.org/10.1073/pnas.1516047113

Hughes, D., Rodriguez, J., Smith, E. P., Johnson, D. J., Stevenson, H. C., & Spicer, P. (2006). Parents' ethnic-racial socialization practices: A review of research and directions for future study. *Developmental Psychology*, *42*(5), 747–770. https://doi.org/10.1037/0012-1649.42.5.747

James, D. (2017). Internalized racism and past-year major depressive disorder among African-Americans: The role of ethnic identity and self-esteem. *Journal of Racial and Ethnic Health Disparities*, *4*(4), 659–670. https://doi.org/10.1007/s40615-016-0269-1

Jones, J. M. (2023). Surviving while black: Systemic racism and psychological resilience. *Annual Review of Psychology*, *74*(1), 1–25. https://doi.org/10.1146/annurev-psych-020822-052232

Jordan, K. L., & Gabbidon, S. L. (2010). Race/ethnicity and perceptions of safety among a national sample of Americans. *Criminal Justice Review*, *35*(3), 281–294. https://doi.org/10.1177/0734016810366453

Kahn, R. L., & Antonucci, T. C. (1980). Convoys over the life course: Attachment, roles, and social support. In P. B. Baltes & O. G. Brim (Eds.), *Life-span development and behavior* (pp. 254–283). Academic Press.

Kenny, D. A., & McCoach, D. B. (2003). Effect of the number of variables on measures of fit in structural equation modeling. *Structural Equation Modeling*, *10*(3), 333–351. https://doi.org/10.1207/S15328007SEM1003_1

Kline, R. B. (2016). *Principles and practice of structural equation modeling* (4th ed.). Guilford Press.

Krueger, R. F., Derringer, J., Markon, K. E., Watson, D., & Skodol, A. E. (2012). *Personality Inventory for DSM-5 (PID-5)* [Database record]. APA PsycTests. https://doi.org/10.1037/t30042-000

Luong, R., & Flake, J. K. (2023). Measurement invariance testing using confirmatory factor analysis and alignment optimization: A tutorial for transparent analysis planning and reporting. *Psychological Methods*, *28*(4), 905–924. https://doi.org/10.1037/met0000441

Marsh, H. W., Morin, A. J., Parker, P. D., & Kaur, G. (2014). Exploratory structural equation modeling: An integration of the best features of exploratory and confirmatory factor analysis. *Annual Review of Clinical Psychology*, *10*(1), 85–110. https://doi.org/10.1146/annurev-clinpsy-032813-153700

Marsh, H. W., Muthén, B., Asparouhov, T., Lüdtke, O., Robitzsch, A., Morin, A. J. S., & Trautwein, U. (2009). Exploratory structural equation modeling, integrating CFA and EFA: Application to students' evaluations of university teaching. *Structural Equation Modeling*, *16*(3), 439–476. https://doi.org/10.1080/10705510903008220

Mays, A. A., Mills, C. J., & Oltmanns, J. R. (2024). Two-year retest reliability and predictive validity of the Self- and Informant-Personality Inventory for ICD-11 in older adults. *Psychological Assessment*, *36*(6–7), 433–439. https://doi.org/10.1037/pas0001316

Mulder, R. T., Horwood, J., Tyrer, P., Carter, J., & Joyce, P. R. (2016). Validating the proposed *ICD-11* domains. *Personality and Mental Health*, *10*(2), 84–95. https://doi.org/10.1002/pmh.1336

Muthén, L. K., & Muthén, B. O. (2011). *Mplus user's guide* (6th ed.).

Myers, N. D., Jin, Y., Ahn, S., Celimli, S., & Zopluoglu, C. (2015). Rotation to a partially specified target matrix in exploratory factor analysis in practice. *Behavior Research Methods*, *47*(2), 494–505. https://doi.org/10.3758/s13428-014-0486-7

Najdowski, C. J., Bottoms, B. L., & Goff, P. A. (2015). Stereotype threat and racial differences in citizens' experiences of police encounters. *Law and Human Behavior*, *39*(5), 463–477. https://doi.org/10.1037/lhb0000140

Oltmanns, J. R. (2021). Personality traits in the International Classification of Diseases 11th Revision (*ICD-11*). *Current Opinion in Psychiatry*, *34*(1), 48–53. https://doi.org/10.1097/YCO.0000000000000656

Oltmanns, J. R., Weinstein, Y., & Oltmanns, T. F. (2014). Borderline personality pathology and insomnia symptoms in community-dwelling older adults. *Personality and Mental Health*, *8*(3), 178–187. https://doi.org/10.1002/pmh.1259

Oltmanns, J. R., & Widiger, T. A. (2018). A self-report measure for the *ICD-11* dimensional trait model proposal: The personality inventory for *ICD-11*. *Psychological Assessment*, *30*(2), 154–169. https://doi.org/10.1037/pas0000459

Oltmanns, J. R., & Widiger, T. A. (2021). The self- and informant-personality inventories for *ICD-11*: Agreement, structure, and relations with health, social, and satisfaction variables in older adults. *Psychological Assessment*, *33*(4), 300–310. https://doi.org/10.1037/pas0000982

Oltmanns, T. F., Rodrigues, M. M., Weinstein, Y., & Gleason, M. E. (2014). Prevalence of personality disorders at midlife in a community sample: Disorders and symptoms reflected in interview, self, and informant reports. *Journal of Psychopathology and Behavioral Assessment*, *36*(2), 177–188. https://doi.org/10.1007/s10862-013-9389-7

Orjiakor, C. T., Sellbom, M., Keeley, J. W., & Bagby, R. M. (2023). Measurement invariance of the Personality Inventory for the *DSM-5* (PID-5) for Nigerian and white American university students. *Psychological Assessment*, *35*(8), 715–720. https://doi.org/10.1037/pas0001251

Petersen, N., & Ward, G. (2015). The transmission of historical racial violence: Lynching, civil rights–era terror, and contemporary interracial homicide. *Race and Justice*, *5*(2), 114–143. https://doi.org/10.1177/2153368714567577

Pierce, G. R., Sarason, I. G., & Sarason, B. R. (1991). General and relationship-based perceptions of social support: Are two constructs better than one? *Journal of Personality and Social Psychology*, *61*(6), 1028–1039. https://doi.org/10.1037/0022-3514.61.6.1028

Purdie-Vaughns, V., Steele, C. M., Davies, P. G., Ditlmann, R., & Crosby, J. R. (2008). Social identity contingencies: How diversity cues signal threat or safety for African Americans in mainstream institutions. *Journal of Personality and Social Psychology*, *94*(4), 615–630. https://doi.org/10.1037/0022-3514.94.4.615

Putnick, D. L., & Bornstein, M. H. (2016). Measurement invariance conventions and reporting: The state of the art and future directions for psychological research. *Developmental Review*, *41*, 71–90. https://doi.org/10.1016/j.dr.2016.06.004

Ramírez, M., Ford, M. E., Stewart, A. L., & Teresi, J. A. (2005). Measurement issues in health disparities research. *Health Services Research*, *40*(5), 1640–1657. https://doi.org/10.1111/j.1475-6773.2005.00450.x

Scott, L. D., Jr. (2003). Cultural orientation and coping with perceived discrimination among African American youth. *Journal of Black Psychology*, *29*(3), 235–256. https://doi.org/10.1177/0095798403254213

Selvanathan, H. P., Jetten, J., & Umeh, A. (2023). A history of collective resilience and collective victimhood: Two sides of the same coin that explain black Americans' present-day responses to oppression. *British Journal of Social Psychology*, *62*(1), 136–160. https://doi.org/10.1111/bjso.12562

Smith, N. A., Voisin, D. R., Yang, J. P., & Tung, E. L. (2019). Keeping your guard up: Hypervigilance among urban residents affected by community and police violence. *Health Affairs*, *38*(10), 1662–1669. https://doi.org/10.1377/hlthaff.2019.00560

Somma, A., Gialdi, G., & Fossati, A. (2020). Reliability and construct validity of the Personality Inventory for *ICD-11* (PiCD) in Italian adult participants. *Psychological Assessment*, *32*(1), 29–39. https://doi.org/10.1037/pas0000766

Somma, A., Krueger, R. F., Markon, K. E., & 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. https://doi.org/10.1037/pas0000711

Spence, C. T., & Oltmanns, T. F. (2011). Recruitment of African American men: Overcoming challenges for an epidemiological study of personality and health. *Cultural Diversity & Ethnic Minority Psychology*, *17*(4), 377–380. https://doi.org/10.1037/a0024732

Stricker, J., Jakob, L., Köhler, D., & Pietrowsky, R. (2022). Six-month stability and predictive validity of the Personality Inventory for *ICD-11*. *BMC Psychology*, *10*(1), Article 270. https://doi.org/10.1186/s40359-022-00979-2

Theriot, M. T., & Orme, J. G. (2016). School resource officers and students' feelings of safety at school. *Youth Violence and Juvenile Justice*, *14*(2), 130–146. https://doi.org/10.1177/1541204014564472

Thomas, A., Caldwell, C. H., Jagers, R. J., & Flay, B. R. (2016). It's in my hood: Understanding African American boys' perception of safety in their neighborhoods. *Journal of Community Psychology*, *44*(3), 311–326. https://doi.org/10.1002/jcop.21768

Tyrer, P., Reed, G. M., & Crawford, M. J. (2015). Classification, assessment, prevalence, and effect of personality disorder. *The Lancet*, *385*(9969), 717–726. https://doi.org/10.1016/S0140-6736(14)61995-4

Vickers, R. R., Jr., Conway, T. L., & Hervig, L. K. (1990). Demonstration of replicable dimensions of health behaviors. *Preventive Medicine*, *19*(4), 377–401. https://doi.org/10.1016/0091-7435(90)90037-K

Wheelock, D., Stroshine, M. S., & O'Hear, M. (2019). Disentangling the relationship between ace and attitudes toward the police: Police contact, perceptions of safety, and procedural justice. *Crime and Delinquency*, *65*(7), 941–968. https://doi.org/10.1177/0011128718811928

Widiger, T. A., & Trull, T. J. (2007). Plate tectonics in the classification of personality disorder: Shifting to a dimensional model. *The American Psychologist*, *62*(2), 71–83. https://doi.org/10.1037/0003-066X.62.2.71

World Health Organization. (2025). *International statistical classification of diseases and related health problems* (11th ed.). https://icd.who.int/

Wright, A. J., Weston, S. J., Norton, S., Voss, M., Bogdan, R., Oltmanns, T. F., & Jackson, J. J. (2022). Prospective self- and informant-personality associations with inflammation, health behaviors, and health indicators. *Health Psychology*, *41*(2), 121–133. https://doi.org/10.1037/hea0001162

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), Article 107. https://doi.org/10.1186/s12888-021-03080-x

Received October 3, 2024
Revision received August 5, 2025
Accepted August 12, 2025 ■

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.
All rights, including for text and data mining, AI training, and similar technologies, are reserved.