Personality Disorders: Theory, Research, and Treatment

Measurement Invariance of the Personality Inventory for the DSM-5 Across U.S. East Asian, Southeast Asian, and White Participants

Megan M. Hricovec, Charlie C. Su, Thomas A. Bart, Kaetlin F. Marsh, Clare K. Alsup, and David C. Cicero

Online First Publication, November 2, 2023. https://dx.doi.org/10.1037/per0000647

AMERICAN PSYCHOLOGICAL ASSOCIATION

Personality Disorders: Theory, Research, and Treatment

ISSN: 1949-2715

https://doi.org/10.1037/per0000647

Measurement Invariance of the Personality Inventory for the *DSM-5* Across U.S. East Asian, Southeast Asian, and White Participants

Megan M. Hricovec, Charlie C. Su, Thomas A. Bart, Kaetlin F. Marsh, Clare K. Alsup, and David C. Cicero
Department of Psychology, University of North Texas

The Personality Inventory for the *Diagnostic and Statistical Manual of Mental Disorders (DSM), Fifth Edition* (PID-5) was developed as a measure of the traits included in the alternative model of personality disorders (AMPD) in Section III of the *DSM*. The PID-5 is composed of 25 scales measuring each trait in the AMPD across five domains: negative affectivity, detachment, disinhibition, antagonism, and psychoticism. Previous research suggests that there may be important differences in the expression of personality pathology across race and culture, particularly between people with eastern and western cultural heritages. The goal of the current research was to examine the measurement invariance of the PID-5 across these groups. In the current study, 865 young men and women who identified as White, East Asian, or Southeast Asian completed the PID-5 and international personality item pool (IPIP). On the domain level, a multigroup exploratory structural equation model found that the PID-5 had configural and metric invariance, but lacked complete scalar invariance. On an item level, all scales had configural invariance, one lacked metric invariance, and 11 of the 25 scales lacked scalar invariance across race. For the invariant scales, East and Southeast Asians tended to have higher mean scores than White participants. The PID-5 scales had similar relations with IPIP scales across groups. These results suggest that the PID-5 scales are measuring similar constructs across groups on a global, structural level, but that mean scores may represent different levels of latent personality pathology across groups. The PID-5 may be confidently used in these groups, but mean comparisons should be interpreted with caution.

Keywords: alternative model of personality disorders, measurement equivalence, cross-cultural, self-construal

Supplemental materials: https://doi.org/10.1037/per0000647.supp

The alternative model of personality disorders (AMPD) appears in Section III (emerging models and measures) of the fifth edition of the *Diagnostic and Statistical Manual of Mental Disorders* (*DSM*; American Psychiatric Association, 2013). Section III also includes the Personality Inventory for the *DSM-5* (PID-5; Krueger et al., 2012), a self-report inventory designed to measure Criterion B of the AMPD (i.e., maladaptive personality traits). The PID-5 measures the five proposed personality domains of negative affect, detachment, antagonism, disinhibition, and psychoticism, as well as 25 facet scales that comprise these domains. An emerging body of research has used the PID-5 in many cross-cultural settings (see Watters & Bagby, 2018, for a review), but few studies have examined whether the scale produces equivalent scores across groups (Bagby et al., 2022). Moreover, the studies that have examined measurement equivalency have focused on the scale scores with the proposed five-factor structure, as opposed to the individual items and facet scales. The primary aim of the current study is to examine the measurement invariance of the PID-5 across East Asian, Southeast Asian, and White individuals living in the United States on both the domain and facet level.

David C. Cicero https://orcid.org/0000-0002-5666-9139

This study was not preregistered. These data were presented as a poster at the 2022 annual meeting for the Association for Psychological Science. The authors have no conflicts of interest to disclose.

The materials and analysis code are available at https://osf.io/ybuka/.

Megan M. Hricovec contributed equally to validation and served in a supporting role for conceptualization, data curation, and formal analysis. Charlie C. Su served in a supporting role for conceptualization, data curation, formal analysis, and writing–review and editing. Thomas A. Bart served in a supporting role for conceptualization, writing–original draft, and writing–review and editing. Kaetlin F. Marsh served in a supporting role for writing–original draft and writing–review and editing. Clare K. Alsup served in a supporting role for writing–original draft and writing–review and editing. David C. Cicero served as lead for conceptualization, data curation, formal analysis, methodology, project administration, resources, supervision, and validation and contributed equally to investigation. Megan M. Hricovec and David C. Cicero contributed equally to writing–original draft and writing–review and editing.

Correspondence concerning this article should be addressed to David C. Cicero, Department of Psychology, University of North Texas, 1155 Union Circle Drive, Denton, TX 76203, United States. Email: david.cicero@unt.edu

All of the emerging research using the PID-5 makes the implicit assumption that the psychometric properties of the scales are equivalent across all participants, but few investigations have been conducted to test this assumption. Measurement invariance introduces a statistical approach for assessing the degree to which a scale is measuring the same construct in the same way across groups of participants (Chen, 2008). There are three primary types of measurement invariance: configural, metric, and scalar. Establishing configural invariance demonstrates that there are the same number of factors across groups and that the subscales and/or items load on the same factors across groups. Tests of metric invariance determine whether the factor loadings are equivalent across groups, which can be interpreted as the equivalence of the strength of the relation between the indicator and latent variables (e.g., the items are equally strong indicators of the construct between groups). Finally, scalar invariance establishes that the intercepts are equivalent across groups, which can be interpreted as observed scores reflecting the same latent level of the construct across groups (e.g., a score of six in one group is equivalent to a score of six in the other groups). For scales lacking scalar invariance, mean comparisons across groups are inappropriate because a higher score in one group may be related to the differences in psychometric properties rather than a true difference in the latent level of the construct.

Measurement Invariance of the PID-5

The PID-5 has been used broadly with diverse populations both in the United States and internationally (Watters & Bagby, 2018). It has been translated into many languages, including Portuguese, Dutch, German, Danish, French, Spanish, Norwegian, and Arabic, among others (e.g., Al-Dajani et al., 2016). Despite this broad use, few studies have examined its measurement equivalence, and the studies that did have produced mixed results. The PID-5 has been shown to have configural, metric, and at least partial scalar invariance across sex in American undergraduates, Australian community members, and clinical samples (Choate et al., 2021). It has also been shown to have measurement invariance across age, across European nationalities, between Norwegian and United States samples, and between clinical and undergraduate/community samples (e.g., Bach et al., 2018; Thimm et al., 2017).

In contrast to these findings showing at least partial invariance, a recent study examining the measurement invariance of the PID-5 between Black and White Americans found that it lacked even configural invariance (Bagby et al., 2022). In White participants, the five-factor model fit the data well, but this factor structure could not be replicated for Black participants. Instead, a one-factor model fit the data best in Black participants. Thus, the PID-5 appears to be measuring the proposed five-factor structure of the AMPD in White participants, but to be measuring what Bagby and colleagues described as a "general personality pathology domain and/or demoralization factor" in Black participants (Bagby et al., 2022, p. 88). This finding suggests that the PID-5 conceptualization of personality pathology may not be consistent across race, which underscores the need for more measurement invariance research in the PID-5.

Personality Pathology and Culture

Theorists have suggested that eastern and western cultures emphasize different values (i.e., culturally shared ideas regarding what attitudes and behaviors are considered desirable; Sagiv & Schwartz, 2022). These differences in values may drive real differences in personality and personality pathology but may also contribute to different interpretations of questionnaire items because frame of reference is often impacted by values (Lvina et al., 2012). The most discussed difference between eastern and western cultures is related to self-construal. Eastern cultures tend to value an interdependent self-construal in which individuals view themselves as part of the broader group, while western cultures tend to value an independent self-construal in which individual attributes and achievements are most valued (Markus & Kitayama, 1991). Differences in self-construal have been shown to affect the expression of psychopathology, both between cultures and in individual differences within cultures (Maas et al., 2019).

These cultural differences in values and self-concepts may have specific effects on responses to certain domains on personality pathology measures. Differences in self-construal may lead to differences in the experience and expression of emotion, a part of the detachment domain of the PID-5. Due to cultural differences in emotional display norms, participants from East Asian cultures may report levels of restricted affect that would be considered pathological in western cultures, but are normative in East Asian cultures (e.g., Kim et al., 2005). In addition to detachment, there may be cultural differences in the expression of negative affectivity symptoms. Theorists have long suggested that people of Asian descent report more somatic symptoms, while people of European descent report more mood symptoms (Kleinman, 1977), but empirical results have been mixed (Ryder et al., 2012). The depressivity scale of the PID-5 measures trait-like depression, but does not include any items related to somatic complaints. As a result, items may more strongly represent depressivity in White participants than in East Asian participants.

Researchers have also noted that what is considered "psychoticism" may vary across cultures (Kulhara & Chakrabarti, 2001; Lorr & Klett, 1969). The *DSM* explicitly acknowledges that beliefs should not be considered delusional if they are consistent with the individual's culture or subculture (American Psychiatric Association, 2013). Moreover, hallucinations are common across cultures, and cultural norms often determine whether they are considered pathological, normative, or even desirable (e.g., Larøi et al., 2014). In interviews, the interviewer can ask follow-up questions to determine if the belief is consistent with the interviewee's culture. In a self-report, however, probes are not possible and religious or other culturally appropriate experiences may be wrongly considered to be psychoticism (Cicero, 2016).

The Current Study

Given the theoretical cross-cultural differences in personality functioning, the primary goal of the current study was to examine the measurement invariance of the PID-5 across East Asian, Southeast Asian, and White Americans. We planned to examine measurement invariance at the global scale level (i.e., the factor structure of the 25 scales that comprise the PID-5), as well as the facet level for each of the 25 scales. If any of the scales lacked measurement invariance, we planned to conduct follow-up analyses to determine which specific items contributed to the lack of

invariance. The third goal was to test mean differences in PID-5 scores across race. Finally, the fourth goal was to further test the invariance of the PID-5 scales by examining whether they have equivalent relations with Big-Five personality traits across race.

Method

Participants

Participants were undergraduates at a large public Hawaiian university who completed the study in exchange for partial completion of a course requirement. A total of 1,554 participants enrolled in the study. Embedded within the study was the Chapman Infrequency Scale (Chapman & Chapman, 1983), and 284 participants were removed due to infrequency scores greater than four. We determined that we needed 200 participants per group to conduct the analyses in the current study following the recommendations of Cheung and Rensvold (2002). An additional 93 Pacific Islander, 145 Multiracial, 56 Hispanic, eight Black/African American, and 28 other participants were excluded from the analyses due to low sample size (Meade & Bauer, 2007). This resulted in 291 White, 321 as East Asian, 253 Southeast Asian participants included in the final sample. Participants were 31.6% male and 68.4% female with a mean age of 20.17 ($SD = 3.26$). There were differences among groups in whether they were first, second, or third or longer generation living in the United States, $\chi^2(4) = 227.870$, $p < .001$. White participants were 5.7% first generation, 17.7% second generation, and 76.7% third or longer generation; East Asian participants were 19.6% first generation, 34.4% second generation, and 46.1% third or longer generation; and Southeast Asian participants were 14.7% first generation, 70.5% second generation, and 14.7% third or longer generation. This study was approved by the University of Hawaii Institutional Review Board. The data and Mplus code are available (Hricovec et al., 2023).

Materials

PID-5

Participants completed the full version of the PID-5 (Krueger et al., 2012) as part of a larger study. The PID-5 is a 220-item self-report questionnaire in which participants respond on a scale from 0 (*very false or often false*) to 3 (*very true or often true*). It contains 25 personality facet scales that measure the five personality domains in the AMPD, including negative affectivity, detachment, antagonism, disinhibition, and psychoticism. The PID-5 has been used extensively in diverse populations and translated into at least 12 different languages (Watters & Bagby, 2018). In the current study, Cronbach's α for the subscales ranged from .604 to .948 and Ω ranged from .622 to .948 (see Table S2 in the online supplemental materials for αs and Ωs for every scale).

IPIP

Participants completed the 100-item version of the IPIP (Goldberg, 1999), in which they answered questions on a scale from 1 (*very inaccurate*) to 5 (*very accurate*). In the current study, the scales had high internal consistency for extraversion ($\Omega = .914$, $\alpha = .916$), agreeableness ($\Omega = .872$, $\alpha = .880$), conscientiousness ($\Omega = .878$, $\alpha = .880$), neuroticism ($\Omega = .905$, $\alpha = .905$), and openness to experience ($\Omega = .861$, $\alpha = .869$).

Procedure

Participants completed the PID-5 and IPIP as part of a larger online study that also included the Wisconsin Schizotypy Scales, Schizotypal Personality Questionnaire, Rosenberg Self-Esteem Scale, Self-Concept Clarity Scale, Multiethnic Identity Measure, Collective-Self-Esteem Scale, and Self-Concept Identity Measure. None of these other measures were analyzed as part of the current research, which focused narrowly on the psychometric properties of the PID-5 and its relation to Big-Five personality. The order of presentation of questionnaires was randomized across participants.

Data Analysis

All analyses were conducted with Mplus Version 8.7 (Muthen & Muthen, 1998). We first tested the measurement invariance of the global model of the scale scores. Most previous research examining the factor structure of the PID-5 has failed to find an adequately fitting model including all 25 scale scores using confirmatory factor analyses. However, some studies have successfully used exploratory structural equation modeling (ESEM), which combines features of both exploratory and confirmatory factor analysis, to examine this factor structure (Somma et al., 2019). Thus, in the current research, we chose to utilize ESEM with all scales loading on all factors. Five factors were specified using maximum likelihood estimation with robust standard errors (MLR). We first tested a single-group model to ensure that the five-factor ESEM model fit the data well. If the single-group model did not fit the data well, we planned to explore an alternative ESEM model, beginning with a parallel analysis to determine the number of factors to extract.

To evaluate the measurement invariance of the scales, we tested a series of models using the "configural metric scalar" model command with Mplus defaults. Mplus defaults set the metric of a factor by setting a factor loading to one. Factor variances, item intercepts, and residual variances are freely estimated across groups. We began with the configural model in which the factor loadings and scale intercepts are free to vary across races. We then planned to test a metric invariance model in which the factor loadings are constrained to be equal across groups but the scale intercepts are allowed to vary across groups. Finally, we planned to test a scalar invariance model in which the factor loadings and scale intercepts are constrained to be equal across groups. We did not test the invariance of residuals because this level of invariance is not necessary for mean comparisons (Putnick & Bornstein, 2016).

Following convention, the single-group model and configural models were considered good fit if the root-mean-squared error of approximation (RMSEA) $< .10$ and Confirmatory Fit Index (CFI) $> .90$ and excellent if RMSEA $< .05$ and CFI $> .95$ (Bentler, 1990; Hu & Bentler, 1999). To establish measurement invariance, four indices were used including (a) ΔCFI $< .010$, (b) change in McDonald's Noncentrality Index (ΔM_c; Cheung & Rensvold, 2002; Mcdonald, 1989; Meade & Bauer, 2007) $< .020$, (c) whether the Bayesian information criterion (BIC) decreases in each successive model (Cao & Liang, 2022), and (d) $RMSEA_D < .010$ (Savalei et al., 2023). The $RMSEA_D$ differs from the ΔRMSEA in that $RMSEA_D$ is calculated based on the difference in χ^2 rather than subtracting the difference in RMSEA calculated independently for the

two models. $RMSEA_D$ can be interpreted like a typical RMSEA in which lower values represent a smaller difference in fit between the models. Due to the well-known problems with chi-square difference testing for measurement invariance, we report but do not interpret the χ^2 difference test (Cheung & Rensvold, 2002).

After testing the measurement invariance of the global scales, we examined each scale individually on the item level following the same format. The PID-5 items have four rank-ordered response options, which are ordinal data. However, categorical data presents some problems for measurement invariance analyses. First, the oftencited simulation studies establishing cut points for ΔCFI and ΔM_c were established with continuous variables using MLR, and are problematic when applied to weighted least squares mean and varaince adjusted estimation (WLSMV) (Sass et al., 2014). Second, multigroup analyses with categorical variables require that at least one participant from each group selected each of the possible options, which was not the case for several scales. Thus, we first tested the fit of a single-group unidimensional model for all 25 scales using MLR estimation. If these models fit the data well, we planned to use MLR estimation to take advantage of its benefits over WLSMV.

Following the same template as the scale-level analyses, we first tested a single-group unidimensional model for all 25 scales. For each of the models that fit the data well, we tested the configural, metric, and scalar invariance of the scale individually using a multigroup CFA. If the model did not fit the data well, we planned to test the measurement invariance of the scales with ESEM with all items loading on all factors. If a scale lacked metric or scalar invariance, we planned to consult the modification indices to determine which item intercepts need to be freed to improve model fit, as is commonly done in measurement invariance research (Skriner & Chu, 2014; Spaapen et al., 2014; Torres et al., 2013). Item intercepts were freed iteratively and the fit of the model was retested. This continued until the scalar model fit as well as the configural model. Next, we conducted mean comparisons for the scales that displayed scalar invariance with a one-way analysis of variance (ANOVA) and Bonferroni-corrected planned comparisons among the three groups. A sensitivity analysis revealed we had .80 power to detect a minimum effect size of $\eta^2 = .011$ for the ANOVAs. As we had uneven group sizes, the minimum effect we could observe with .80 power was slightly different for each of the comparisons ($d = 0.26$ for East Asian/White, $d = .28$ for Southeast Asian/White, and $d = 0.27$ for East Asian/Southeast Asian).

Finally, we regressed each of the five IPIP scales on all 25 of the PID facets. For each analysis, we specified a model in which the relation between the IPIP scale (i.e., separately for each of the five IPIP scales) and the facet (i.e., separately for each of the 25 facets) were allowed to vary among the three groups in the study. This resulted in 125 models. Then, we tested the same models with this parameter to be constrained to be equal across groups. We evaluated the fit between these models with a Satorra–Bentler chi-squared difference test. If the constrained model fit as well as the unconstrained model, then the relations between IPIP scales and PID-5 facets can be interpreted to be equal across groups.

Results

We first tested the measurement invariance of the PID-5 scale scores across race using ESEM. As can be seen in Table 1, the configural model with all scales freely loading on all factors fit the data well, suggesting that the PID-5 has configural invariance across racial groups. The metric invariance model also fit the data well and fit just as well as the configural model according to the ΔCFI, ΔM_c, BIC, and $RMSEA_D$. This suggests that the scales have equivalent factor loadings across groups (i.e., the scales are similar indicators of the latent constructs across groups). Finally, the scalar invariance model also fit the data as well as the configural model according to ΔCFI, BIC, and $RMSEA_D$ but not ΔM_c. Since the scalar invariance model did not fit as well as the configural model according to ΔM_c, we consulted the modification indices to determine which intercepts were responsible for the lack of clear scalar invariance. Freeing the intercepts for submissiveness in the East Asian group and anhedonia in the White group resulted in a modified scalar model that fit as well as the configural model. Taken together, these results provide equivocal support for the scalar invariance of the global scale scores. Factor loadings by group for the scale level ESEM are shown in Table S1 in the online supplemental materials.

Following measurement invariance analyses for the full scale, we conducted item-level measurement invariance analyses for each individual scale. As can be seen in Table S2 in the online supplemental materials, the single-group unidimensional models fit the data reasonably well, with the exception of emotional lability, separation insecurity, risk taking, hostility, and perceptual dysregulation. Thus, we used ESEM with two factors to test the fit of these models. Manipulativeness also had a RMSEA $> .100$, but the configural model for a two-factor ESEM did not converge. Thus, we ran this model as a single factor. Table 2 shows that 24 of the 25 scales displayed configural and metric invariance and 14 of the 25 scales displayed scalar invariance. Five of the six detachment scales (withdrawal, intimacy avoidance, anhedonia, depressivity, and restricted affectivity), all five antagonism scales (manipulativeness, deceitfulness, grandiosity, attention seeking, and callousness), and one disinhibition scale (irresponsibility) lacked complete scalar invariance. All negative affectivity and psychoticism scales displayed scalar invariance. Similar to the scale-level analyses, we iteratively freed one item intercept based on the modification indices and tested the fit until the model fit as well as the configural model according to the ΔCFI, ΔM_c, $RMSEA_D$, and BIC. Factor loadings for the configural models can be found in Tables S3–S27 in the online supplemental materials. Table S28 in the online supplemental materials shows the items that were responsible for the lack of scalar invariance.

We next examined mean differences between groups for the scales that displayed scalar invariance with a one-way ANOVA and a Bonferroni correction for multiple comparisons. As can be seen in Table S29 in the online supplemental materials, White participants tended to have lower scores than East and Southeast Asian participants. East and Southeast Asian participants did not differ on any scale. In the negative affectivity domain, White participants had lower scores than East and Southeast Asian participants on all six facets. These effects were small to moderate. On the detachment domain, White participants had lower scores than East Asian participants on the suspiciousness facet, which was the only facet in this domain with scalar invariance. On the disinhibition domain, White participants had lower scores than East and Southeast Asian participants on impulsivity, distractibility, and risk taking, and lower scores than Southeast Asians on rigid perfectionism. There were no differences between groups on the psychoticism domain. None of the antagonism scales displayed scalar invariance. We

Table 1
Scale-Level Measurement Invariance of the Personality Inventory for the DSM-5 Across Race

Model	χ^2	*df*	RMSEA	CFI	BIC	SB χ^2 diff.	*df*	ΔCFI	ΔM_c	$RMSEA_D$
Single group analysis										
Single group	1,335.865	185	.080 (.076–.084)	.919	122,656.79					
Multiple group analyses										
Configural	1,434.445	555	.082 (.076–.087)	.923	91,476.070					
Metric	1,649.961	755	.071 (.066–.075)	.921	90,560.173	276.047	200	.002	.005	.031
Scalar	1,747.862	795	.071 (.066–.075)	.916	90,396.581	363.509	240	.007	.027[b]	.062
Modified scalar[a]	1,722.559	793	.070 (.066–.075)	.918	90,383.756	342.612	238	.005	.018	.052

Note. All participants are included in the single group analysis. The configural, metric, scalar, and modified scalar model are multiple-group exploratory structural equation models. In the configural model, the factor loadings and intercepts are free to vary across groups. In the metric invariance model, the factor loadings are constrained to be equal across groups. In the scalar invariance model, the factor loadings and item intercepts are constrained to be equal across groups. Lower BIC, $\Delta CFI < .010$, $\Delta M_c < .020$, and $\Delta RMSEA > -.015$ indicate invariance. RMSEA = root-mean-squared error of approximation; CFI = Confirmatory Fit Index; BIC = Bayesian information criterion; SB χ^2 diff. = Satorra–Bentler χ^2 difference test; ΔCFI = change in CFI; ΔM_c = change in McDonald's Noncentrality Index. $RMSEA_D$ = RMSEA based on χ^2 difference test.
[a] Intercepts allowed to vary for submissiveness in the East Asian group and anhedonia in the White group. [b] Metric indicates a lack of measurement invariance.

also examined differences in latent means with the White group as the reference group. As shown in Table S30 in the online supplemental materials, the latent mean comparisons yielded the same results as the raw score comparisons with the exceptions of the difference between White and Southeast Asian being nonsignificant for separation insecurity and submissiveness and the difference between White, East Asian, and Southeast Asian being significant for perceptual dysregulation.

Finally, we examined the relations among all 25 PID-5 facet scales and the Big-Five personality traits as measured with the IPIP. As can be seen in Tables S31–S33 in the online supplemental materials, most PID-5 facets were negatively associated with Big-Five personality traits in all three groups (294 of 375 regression weights are statistically significant). As expected, the relations tended to be strongest between negative affectivity facets and neuroticism, detachment facets and extraversion, antagonism facets and agreeableness, and disinhibition facets and conscientiousness. Of the 125 relations between Big-Five traits and PID-5 facets, 108 were not significantly statistically different among the groups (see Table S34 in the online supplemental materials). Of the 17 relations that were different among groups, 11 were with conscientiousness including anxiousness on the negative affectivity domain, withdrawal and restricted affectivity on the detachment domain, manipulativeness, deceitfulness, and callousness on the antagonism domain, risk taking and rigidity on the disinhibition domain, and eccentricity and unusual beliefs and experiences on the psychoticism domain. All five of the antagonism facets (i.e., manipulativeness, deceitfulness, grandiosity, attention seeking, and callousness) had unequal relations with neuroticism. Finally, anhedonia had different associations with extraversion.

Discussion

The PID-5 is the most commonly used measure that is explicitly keyed to the AMPD in the *DSM-5*. The current study found that the factor structure of the global scale had configural, and metric invariance, but lacked complete scalar invariance across race. Follow-up analyses found that the Anhedonia intercept needed to be freed in the White group and the Submissiveness intercept needed to be freed in the East Asian group. These results suggest that the scales are likely measuring the same constructs across groups but mean scores may represent different latent levels of personality pathology across groups. On an individual scale level, 11 of the 25 scales lacked scalar invariance across race. Mean comparisons with these 11 scales should be interpreted with caution or not done at all. Taken together, these results suggest that many scale scores may represent different latent levels of personality pathology across groups, and global factor-structure measurement invariance analyses may obscure some of the biases that are evident at the individual scale/item level.

To our knowledge, the current study is the first study to examine the measurement equivalence of the PID-5 in White, East Asian, and Southeast Asian American participants. The findings of the current research suggest that the PID-5 is less biased in these minority groups than in other racial and ethnic minority groups in the United States. For example, Bagby et al. (2022) found that the scale lacked even the most basic form of invariance between White and Black Americans, which suggests that research may need to reevaluate its use for Black Americans. The finding that the domain scales have configural, metric, and partial scalar invariance suggests that it can be used in these populations for most purposes. The finding that many facet-level scales lacked scalar invariance suggests that mean comparisons of these scores may result in biased conclusions. At the same time, psychopathology researchers have argued that mean comparisons of psychopathology between racial groups are of limited utility and may reinforce outdated cultural deficit models that portray differences as deficits in racial and ethnic minorities (Coll et al., 2000; Medin et al., 2010). Thus, the lack of scalar invariance may reinforce the idea that mean comparisons of the PID-5 have little practical value.

The specific scales that lacked complete invariance may contribute to our understanding of cultural differences in personality pathology. Five of the six scales on the detachment domain lacked invariance. Items responsible for lack of invariance were related to interpersonal distance, lack of interest in sex, low energy, and emotional expression, and the majority of the problematic intercepts were in the White group. For example, the items "I don't get emotional," "I keep my distance from people," and "I talk about suicide a lot" all had unequal intercepts across groups. This finding may be related to cultural differences in the experience and expression of emotion,

Table 2
Item-Level Measurement Invariance of the PID-5 Scales by Race

Model	χ^2	*df*	RMSEA	CFI	BIC	SB χ^2 diff.	*df*	ΔCFI	ΔM_c	$RMSEA_D$
Negative affectivity										
Emotional lability[a] (seven items)										
Configural	34.956	24	.044 (.000–.073)	.993	11,017.46					
Metric	54.112	44	.031 (.000–.056)	.994	10,909.53	19.269	20	−.001	.004	.000
Scalar	59.248	54	.020 (.000–.047)	.997	10,847.60	23.900	30	−.004	.012	.000
Anxiousness (nine items)										
Configural	225.640	81	.086 (.073–.100)	.936	14,676.07					
Metric	250.972	97	.081 (.069–.094)	.932	14,590.50	20.562	16	.004	.006	.085
Scalar	276.974	113	.078 (.066–.090)	.927	14,508.96	44.180	32	.009	.012	.087
Separation insecurity[a] (seven items)										
Configural	58.773	24	.078 (.053–.103)	.972	11,690.52					
Metric	73.855	44	.053 (.031–.074)	.976	11,583.46	19.089	20	−.004	−.003	.000
Scalar	97.356	54	.058 (.039–.076)	.965	11,540.63	40.703	30	.007	.006	.044
Submissiveness (four items)										
Configural	3.899	6	.000 (.000–.064)	1.000	6,367.02					
Metric	8.695	12	.000 (.000–.049)	1.000	6,333.02	4.935	6	.000	−.001	.000
Scalar	22.023	18	.030 (.000–.068)	.993	6,308.57	19.447	12	.007	.004	.044
Hostility[a] (10 items)										
Configural	155.705	78	.068 (.054–.083)	.955	15,866.76					
Metric	201.542	110	.059 (.046–.071)	.953	15,707.51	38.880	32	.002	.007	.074
Scalar	212.339	126	.053 (.041–.066)	.956	15,610.07	48.671	48	−.001	.005	.047
Perseveration (nine items)										
Configural	136.447	81	.054 (.037–.069)	.969	13,485.53					
Metric	148.493	97	.047 (.031–.062)	.971	13,387.63	7.302	16	−.002	−.002	.000
Scalar	160.978	113	.042 (.026–.056)	.973	13,291.72	16.561	32	−.004	−.004	.000
Detachment										
Withdrawal (10 items)										
Configural	262.722	105	.079 (.067–.091)	.937	14,411.56					
Metric	290.850	123	.075 (.064–.086)	.933	14,317.01	23.822	18	.004	.006	.084
Scalar	346.049	141	.078 (.067–.088)	.918	14,255.84	81.752	36	.019[b]	.029[b]	.128[b]
Modified	317.509	138	.073 (.063–.084)	.928	14,242.19	48.166	33	.009	.013	.091
Intimacy avoidance (six items)										
Configural	36.174	27	.038 (.000–.067)	.990	8,929.69					
Metric	46.915	37	.033 (.000–.060)	.989	8,876.18	10.523	10	.001	.001	.031
Scalar	74.139	47	.049 (.026–.070)	.970	8,840.83	39.176	20	.020[b]	.012	.106[b]
Modified	55.531	45	.031 (.000–.056)	.989	8,831.87	18.523	18	.001	.001	.031
Anhedonia (eight items)										
Configural	155.844	60	.081 (.066–.097)	.935	12,023.20					
Metric	175.412	74	.075 (.061–.090)	.931	11,954.97	19.369	14	.005	.007	.070
Scalar	221.225	88	.079 (.066–.092)	.909	11,911.66	64.793	48	.028[b]	.016	.129[b]
Modified	190.993	85	.072 (.058–.085)	.928	11,895.57	32.792	25	.007	.013	.071
Depressivity (14 items)										
Configural	627.252	231	.084 (.077–.092)	.910	18,476.09					
Metric	662.453	257	.081 (.073–.089)	.908	18,335.21	25.307	26	.002	.005	.067
Scalar	725.650	283	.081 (.073–.088)	.899	18,224.46	83.125	52	.011[b]	.024[b]	.096
Modified	718.192	282	.080 (.073–.087)	.901	18,221.28	74.021	51	.009	.020	.086
Restricted affectivity (seven items)										
Configural	109.136	42	.081 (.063–.100)	.936	11,233.18					
Metric	123.460	54	.073 (.056–.090)	.934	11,165.92	11.273	12	.002	.002	.049
Scalar	151.380	66	.073 (.058–.089)	.918	11,114.85	38.898	24	.016[b]	.012	.097
Modified	136.771	65	.068 (.052–.083)	.931	11,104.63	45.959	35	.005	.003	.050
Suspiciousness (seven items)										
Configural	110.481	42	.082 (.064–.101)	.882	12,005.23					
Metric	120.279	54	.071 (.054–.088)	.886	11,940.59	11.854	12	−.004	−.001	.000
Scalar	137.441	66	.067 (.051–.083)	.877	11,878.08	27.847	24	.005	−.002	.039
Antagonism										
Manipulativeness (five items)										
Configural	48.493	15	.097 (.067–.128)	.958	7,777.01					
Metric	67.833	23	.090 (.066–.115)	.944	7,742.11	17.421	8	.014	.008	.133
Modified	60.489	22	.083 (.061–.106)	.952	7,740.67	9.153	7	.006	.003	.095
Scalar	79.863	30	.087 (.066–.109)	.937	7,706.65	35.936	16	.021[b]	.010	.117[b]
Modified	68.516	28	.078 (.055–.101)	.949	7,707.02[b]	26.008	21	.009	.004	.082

(*table continues*)

Table 2 (*continued*)

Model	χ^2	*df*	RMSEA	CFI	BIC	SB χ^2 diff.	*df*	ΔCFI	ΔM_c	$RMSEA_D$
Deceitfulness (10 items)										
Configural	296.649	105	.087 (.076–.099)	.906	15,262.76					
Metric	325.273	123	.083 (.072–.094)	.901	15,165.44	21.700	18	.005	.006	.086
Scalar	359.089	141	.080 (.070–.091)	.894	15,077.32	52.230	36	.012[b]	.016	.096
Modified	352.576	140	.079 (.069–.090)	.896	15,076.52	44.711	35	.010	.013	.087
Grandiosity (six items)										
Configural	72.180	27	.083 (.060–.107)	.943	8,925.61					
Metric	81.700	37	.071 (.050–.092)	.944	8,868.12	7.429	10	−.001	−.001	.000
Scalar	104.676	47	.061 (.045–.076)	.927	8,824.97	29.545	20	.016[b]	.008	.089
Modified	93.777	46	.066 (.047–.085)	.940	8,819.36	17.679	19	.003	.002	.041
Attention seeking (eight items)										
Configural	212.736	60	.103 (.088–.118)	.914	12,182.30					
Metric	240.675	74	.097 (.083–.111)	.906	12,109.79	20.892	14	.008	.007	.111
Scalar	305.414	88	.101 (.089–.114)	.877	12,086.05	91.053	28	.037[b]	.039[b]	.170[b]
Modified	250.142	82	.092 (.079–.105)	.905	12,062.99	26.483	39	.009	.010	.093
Callousness (14 items)										
Configural	378.950	231	.051 (.042–.061)	.957	17,049.44					
Metric	423.935	257	.052 (.043–.061)	.952	16,933.23	45.440	26	.004	.007	.095
Scalar	468.789	283	.052 (.044–.060)	.947	16,807.66	91.071	52	.010[b]	.023[b]	.095
Modified	462.130	281	.052 (.043–.060)	.948	16,812.73	83.493	50	.009	.020	.091
Disinhibition										
Irresponsibility (seven items)										
Configural	74.355	42	.057 (.035–.077)	.964	10,500.84					
Metric	85.915	54	.050 (.029–.069)	.964	10,434.33	10.962	12	.000	−.003	.000
Scalar	138.999	66	.068 (.052–.084)	.918	10,413.71	66.854	24	.048[b]	.027[b]	.146[b]
Modified	105.134	64	.052 (.033–.069)	.954	10,388.16	38.851	34	.010	.006	.070
Impulsivity (six items)										
Configural	89.448	27	.098 (.076–.120)	.954	9,179.080					
Metric	101.064	37	.085 (.065–.104)	.953	9,121.567	8.798	10	.001	.001	.044
Scalar	117.380	47	.079 (.061–.097)	.948	9,071.701	24.809	20	.006	.005	.070
Distractibility (nine items)										
Configural	121.855	81	.046 (.028–.062)	.980	13,696.27					
Metric	138.954	97	.042 (.025–.057)	.979	13,604.18	14.541	16	.001	.001	.029
Scalar	163.688	113	.043 (.027–.057)	.975	12,197.30	40.208	32	.005	.007	.062
Risk taking[a] (14 items)										
Configural	411.921	192	.069 (.060–.078)	.918	21,889.62					
Metric	457.703	240	.061 (.053–.070)	.919	21,626.23	45.896	48	−.001	−.001	.000
Scalar	495.017	264	.060 (.052–.068)	.914	21,504.77	81.024	72	.004	.007	.044
Rigid perfectionism (10 items)										
Configural	233.875	105	.071 (.059–.084)	.938	15,454.89					
Metric	252.602	123	.066 (.055–.078)	.938	15,351.98	14.721	18	.000	.001	.023
Scalar	281.097	141	.064 (.053–.075)	.933	15,260.27	41.066	36	.005	.007	.063
Psychoticism										
Unusual beliefs and experiences (eight items)										
Configural	120.477	60	.065 (.048–.082)	.956	11,699.06					
Metric	140.122	74	.061 (.045–.076)	.952	11,622.94	22.895	21	.004	.003	.071
Scalar	165.556	88	.061 (.046–.075)	.944	11,556.39	46.602	42	.012	.011	.088
Perceptual dysregulation[a] (12 items)										
Configural	248.677	129	.062 (.050–.074)	.955	16,677.64					
Metric	280.318	169	.052 (.041–.063)	.958	16,458.67	32.088	40	−.003	−.004	.000
Scalar	307.284	189	.051 (.040–.061)	.955	14,508.96	55.202	60	.000	−.001	.000
Eccentricity (13 items)										
Configural	494.780	195	0.091 (0.082–0.100)	.934	18,115.93					
Metric	534.363	219	0.077 (0.069–0.085)	.931	17,979.39	5.986	22	.003	.002	.090
Scalar	570.328	243	0.075 (0.067–0.083)	.928	17,849.68	32.786	44	.007	.007	.084

Note. The formula for $RMSEA_D$ results in a square root of a negative number if the difference in χ^2 is less than the difference in degrees of freedom. These values are set to .000. RMSEA = root-mean-squared error of approximation; CFI = Confirmatory Fit Index; BIC = Bayesian information criterion; SB χ^2 diff. = Satorra–Bentler χ^2 difference test; ΔCFI = change in CFI; ΔM_c = change in McDonald's Noncentrality Index; $RMSEA_D$ = RMSEA based on χ^2 difference test.

[a] Model tested with exploratory structural equation modeling. All other models tested with a one-factor model. Modified = a scalar model with unequal intercepts allowed to vary between groups. [b] A lack of measurement invariance for the given metric.

where restricted emotional expression may be considered pathological in western but not East Asian cultures (Lim, 2016; Schouten et al., 2020). These findings are consistent with previous research indicating a lack of scalar invariance in scales measuring detachment constructs, which found a lack of complete scalar invariance between White and Asian (East and Southeast Asians together) groups on scales such as the suspiciousness subscale of the Schizotypal Personality Questionnaire (Cicero, 2016) and the Revised Social Anhedonia Scale (Cicero et al., 2019).

In the disinhibition domain, the irresponsibility scale lacked scalar invariance. Example items with unequal intercepts include "others see me as irresponsible" and "I'm often pretty careless with my own and other's things." Differences in these intercepts may represent cultural differences in values, in which eastern cultures tend to value restraint more than western cultures (Levinson et al., 2011; Su et al., 2013; Tams, 2008). Thus, irresponsibility may need a higher cutoff in the White group to be considered pathological than in the East and Southeast Asian groups. Likewise, all five of the scales on the antagonism domain lacked scalar invariance. Example items on this domain include "I do things to make sure people notice me," "I'm good at making people do what I want them to do," and "I use people to get what I want." This may also represent differences in cultural values, such that manipulating or influencing others may be more normative in independent cultures where assertiveness is valued, but more pathological in interdependent cultures in which group harmony is valued more than the individual (Lui et al., 2018; Thalmayer & Rossier, 2019).

On the negative affectivity domain, the separation insecurity scale lacked complete scalar invariance. Similar to the differences in detachment and antagonism, the lack of complete scalar invariance in this domain may be related to cultural differences in the self-construal (Kim et al., 2005; Markus & Kitayama, 1991; Su et al., 2013). Items measuring separation insecurity (e.g., I fear being alone in my life more than anything) may tap an especially negative experience in cultures that value interdependence more than independence.

The interpretation that the lack of invariance across race is due to differences in the cultural heritage of participants should be interpreted with the caveat that the acculturation of the groups is not homogenous. A limitation of the current research is that race was measured, but not culture. This is especially true in the East Asian group in which just over half were first or second generation and just under half were third-or-higher generation. Participants with both parents born in the United States (i.e., third-or-higher generation) may be more acculturated to mainstream American culture than participants whose parents or themselves were born in East or Southeast Asian. At the same time, participants were living in Hawaii, where East and Southeast Asians are a plurality of residents (U.S. Census Bureau, 2022). It is likely that participants in the current research are more acculturated to East and Southeast Asian culture than people with East and Southeast Asian heritage living in the continental United States. The current study did not have a large enough sample to examine measurement invariance across generation status. Thus, the results of the current research may not generalize to people living in East and Southeast Asia or people in the continental United States.

For the scales that displayed complete scalar invariance, East and Southeast Asians tended to have higher scores than White participants, and East and Southeast Asians did not differ from each other on any scores. This is consistent with a long line of research that suggests racial and ethnic minorities tend to report higher levels of psychopathology than members of majority populations (Breslau et al., 2005). These differences may be related to a host of factors under the umbrella of minority stress, such as interpersonal prejudice, discrimination, and systematic racism (Vaid & Lansing, 2020).

One limitation of the current research is that the participants were undergraduates. Thus, results may not generalize to young people who are not attending college. At the same time, over 40% of Americans between the ages of 18 and 24 are currently enrolled in college, and over 60% of Americans 25 years and older have at least some college education (U.S. Census Bureau, 2021). Moreover, recent research suggests that rates of mental illness are increasing in college students throughout the world, making this an appropriate participant group for these analyses (Auerbach et al., 2018; Oswalt et al., 2020). Future research may examine the measurement invariance of the PID-5 in individuals not attending college. Another potential limitation of the data analysis plan in the current research is that we ran many different tests of measurement invariance, which may lead to a higher Type 1 error rate. In the context of the current research, this suggests that some of the findings of noninvariance (similar to finding a significant effect in a null hypothesis significance testing framework) may be false positives and as a result our conclusions may be overestimates of the lack of invariance. In other words, the scales may have less bias than the current study estimates. Monte Carlo studies have established cutoffs for measurement invariance analyses, but to our knowledge they have not established empirically how to adjust these cutoffs for multiple tests. One solution would be to increase the cutoffs for ΔCFI and ΔM_c to conclude a scale lacks invariance. However, it is unclear how far these adjustments should go and what effect they would have on Type II errors. Future simulation studies may address this point directly by testing Type I and Type II error rates in multiple measurement invariance analyses in a single dataset. In addition, the current research used MLR instead of WLSMV because MLR has several advantages over WLSMV in measurement invariance analyses. However, MLR has limitations with ordinal data with fewer than five response options which may have affected the results of the study (Rhemtulla et al., 2012).

Overall, the results of the current research provide mixed evidence for the measurement invariance of the PID-5 scales. Both the global and facet/item-level analyses found that the scales displayed configural and metric invariance. These results suggest that the PID-5 is likely measuring the same constructs across these groups. However, nearly half of the individual scales lacked complete scalar invariance, suggesting that mean comparisons by race should be interpreted with caution. Future research may continue to examine the measurement invariance of the PID-5 across additional demographic groups to determine whether the scales produce reliable and valid scores across diverse groups.

References

Al-Dajani, N., Gralnick, T. M., & Bagby, R. M. (2016). A psychometric review of the Personality Inventory for DSM-5 (PID-5): Current status and future directions. *Journal of Personality Assessment*, *98*(1), 62–81. https://doi.org/10.1080/00223891.2015.1107572

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders, DSM-5* (5th ed.).

Auerbach, R. P., Mortier, P., Bruffaerts, R., Alonso, J., Benjet, C., Cuijpers, P., Demyttenaere, K., Ebert, D. D., Green, J. G., Hasking, P., Murray, E., Nock, M. K., Pinder-Amaker, S., Sampson, N. A., Stein, D. J., Vilagut, G., Zaslavsky, A. M., & Kessler, R. C. (2018). WHO World Mental Health Surveys International College Student Project: Prevalence and distribution of mental disorders. *Journal of Abnormal Psychology*, *127*(7), 623–638. https://doi.org/10.1037/abn0000362

Bach, B., Sellbom, M., & Simonsen, E. (2018). Personality Inventory for DSM-5 (PID-5) in clinical versus nonclinical individuals: Generalizability of psychometric features. *Assessment*, *25*(7), 815–825. https://doi.org/10.1177/1073191117709070

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

Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin*, *107*(2), 238–246. https://doi.org/10.1037/0033-2909.107.2.238

Breslau, J., Kendler, K. S., Su, M., Gaxiola-Aguilar, S., & Kessler, R. C. (2005). Lifetime risk and persistence of psychiatric disorders across ethnic groups in the United States. *Psychological Medicine*, *35*(3), 317–327. https://doi.org/10.1017/S0033291704003514

Cao, C., & Liang, X. (2022). Sensitivity of fit measures to lack of measurement invariance in exploratory structural equation modeling. *Structural Equation Modeling*, *29*(2), 248–258. https://doi.org/10.1080/10705511.2021.1975287

Chapman, L. J., & Chapman, J. P. (1983). *Infrequency scale for personality measures* [Unpublished scale available from T.R. Kwapil]. UIUC Department of Psychology.

Chen, F. F. (2008). What happens if we compare chopsticks with forks? The impact of making inappropriate comparisons in cross-cultural research. *Journal of Personality and Social Psychology*, *95*(5), 1005–1018. https://doi.org/10.1037/a0013193

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling*, *9*(2), 233–255. https://doi.org/10.1207/S15328007SEM0902_5

Choate, A. M., Gorey, C., Rappaport, L. M., Wiernik, B., & Bornovalova, M. (2021). Alternative model of personality disorders traits predict residential addictions treatment completion. *Drug and Alcohol Dependence*, *228*, Article 109011. https://doi.org/10.1016/j.drugalcdep.2021.109011

Cicero, D. C. (2016). Measurement invariance of the Schizotypal Personality Questionnaire in Asian, Pacific Islander, White, and multiethnic populations. *Psychological Assessment*, *28*(4), 351–361. https://doi.org/10.1037/pas0000180

Cicero, D. C., Martin, E. A., & Krieg, A. (2019). Differential item functioning of the full and brief Wisconsin schizotypy scales in Asian, White, Hispanic, and multiethnic samples and between sexes. *Assessment*, *26*(6), 1001–1013. https://doi.org/10.1177/1073191117719509

Coll, C. G., Akerman, A., & Cicchetti, D. (2000). Cultural influences on developmental processes and outcomes: Implications for the study of development and psychopathology. *Development and Psychopathology*, *12*(3), 333–356. https://doi.org/10.1017/S0954579400003059

Goldberg, L. R. (1999). A broad-bandwidth, public health domain, personality inventory measuring the lower-level facets of several five-factor models. In I. Mervielde, I. Deary, F. De Fruyt, & F. Ostendorf (Eds.), *Personality psychology in Europe* (Vol. 77, pp. 7–28). Tilburg University Press.

Hricovec, M., Su, C. C.-Y., Bart, T., Marsh, K., Alsup, C., & Cicero, D. C. (2023). *Measurement invariance of the Personality Inventory for the DSM-5 across U.S. East Asian, Southeast Asian, and White Participants*. PsyArXiv. https://doi.org/10.31234/osf.io/23nhf

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, *6*(1), 1–55. https://doi.org/10.1080/10705519909540118

Kim, B. K., Li, L. C., & Ng, G. F. (2005). The Asian American values scale—multidimensional: Development, reliability, and validity. *Cultural Diversity and Ethnic Minority Psychology*, *11*(3), 187–201. https://doi.org/10.1037/1099-9809.11.3.187

Kleinman, A. M. (1977). Depression, somatization and the "new cross-cultural psychiatry". *Social Science & Medicine (1967)*, *11*(1), 3–9. https://doi.org/10.1016/0037-7856(77)90138-X

Krueger, R. F., Derringer, J., Markon, K. E., Watson, D., & Skodol, A. E. (2012). Initial construction of a maladaptive personality trait model and inventory for DSM-5. *Psychological Medicine*, *42*(9), 1879–1890. https://doi.org/10.1017/S0033291711002674

Kulhara, P., & Chakrabarti, S. (2001). Culture and schizophrenia and other psychotic disorders. *Psychiatric Clinics of North America*, *24*(3), 449–464. https://doi.org/10.1016/S0193-953X(05)70240-9

Larøi, F., Luhrmann, T. M., Bell, V., Christian, W. A., Deshpande, S., Fernyhough, C., Jenkins, J., & Woods, A. (2014). Culture and hallucinations: Overview and future directions. *Schizophrenia Bulletin*, *40*(Suppl_4), S213–S220. https://doi.org/10.1093/schbul/sbu012

Levinson, C. A., Langer, J. K., & Rodebaugh, T. L. (2011). Self-construal and social anxiety: Considering personality. *Personality and Individual Differences*, *51*(3), 355–359. https://doi.org/10.1016/j.paid.2011.04.006

Lim, N. (2016). Cultural differences in emotion: Differences in emotional arousal level between the east and the west. *Integrative Medicine Research*, *5*(2), 105–109. https://doi.org/10.1016/j.imr.2016.03.004

Lorr, M., & Klett, C. J. (1969). Cross-cultural comparison of psychotic syndromes. *Journal of Abnormal Psychology*, *74*(4), 531–543. https://doi.org/10.1037/h0027783

Lui, P. P., Vidales, C. A., & Rollock, D. (2018). Personality and the social environment: Contributions to psychological adjustment among Asian and Euro American students. *Journal of Social and Clinical Psychology*, *37*(9), 659–696. https://doi.org/10.1521/jscp.2018.37.9.659

Lvina, E., Johns, G., Treadway, D. C., Blickle, G., Liu, Y., Liu, J., Atay, S., Zettler, I., Solga, J., Noethen, D., & Ferris, G. R. (2012). Measure invariance of the Political Skill Inventory (PSI) across five cultures. *International Journal of Cross Cultural Management*, *12*(2), 171–191. https://doi.org/10.1177/1470595812439870

Maas, J., Van Assen, M., Van Balkom, A., Rutten, E. A. P., & Bekker, M. H. J. (2019). Autonomy–connectedness, self-construal, and acculturation: Associations with mental health in a multicultural society. *Journal of Cross-Cultural Psychology*, *50*(1), 80–99. https://doi.org/10.1177/0022022118808924

Markus, H. R., & Kitayama, S. (1991). Culture and self: Implications for cognition, emotion, and motivation. *Psychological Review*, *98*(2), 224–253. https://doi.org/10.1037/0033-295X.98.2.224

Mcdonald, R. P. (1989). An index of goodness-of-fit based on noncentrality. *Journal of Classification*, *6*(1), 97–103. https://doi.org/10.1007/BF01908590

Meade, A. W., & Bauer, D. J. (2007). Power and precision in confirmatory factor analytic tests of measurement invariance. *Structural Equation Modeling*, *14*(4), 611–635. https://doi.org/10.1080/10705510701575461

Medin, D., Bennis, W., & Chandler, M. (2010). Culture and the home-field disadvantage. *Perspectives on Psychological Science*, *5*(6), 708–713. https://doi.org/10.1177/1745691610388772

Muthen, L. K., & Muthen, B. O. (1998). *Mplus user's guide* (8th ed.).

Oswalt, S. B., Lederer, A. M., Chestnut-Steich, K., Day, C., Halbritter, A., & Ortiz, D. (2020). Trends in college students' mental health diagnoses and utilization of services, 2009–2015. *Journal of American College Health*, *68*(1), 41–51. https://doi.org/10.1080/07448481.2018.1515748

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

Rhemtulla, M., Brosseau-Liard, P. E., & Savalei, V. (2012). When can categorical variables be treated as continuous? A comparison of robust continuous and categorical SEM estimation methods under suboptimal conditions. *Psychological Methods*, *17*(3), 354–373. https://doi.org/10.1037/a0029315

Ryder, A. G., Sun, J., Zhu, X., Yao, S., & Chentsova-Dutton, Y. E. (2012). Depression in China: Integrating developmental psychopathology and cultural-clinical psychology. *Journal of Clinical Child & Adolescent Psychology*, *41*(5), 682–694. https://doi.org/10.1080/15374416.2012.710163

Sagiv, L., & Schwartz, S. H. (2022). Personal values across cultures. *Annual Review of Psychology*, *73*(1), 517–546. https://doi.org/10.1146/annurev-psych-020821-125100

Sass, D. A., Schmitt, T. A., & Marsh, H. W. (2014). Evaluating model fit with ordered categorical data within a measurement invariance framework: A comparison of estimators. *Structural Equation Modeling*, *21*(2), 167–180. https://doi.org/10.1080/10705511.2014.882658

Savalei, V., Brace, J. C., & Fouladi, R. T. (2023). We need to change how we compute RMSEA for nested model comparisons in structural equation modeling. *Psychological Methods.* Advance online publication. https://doi.org/10.1037/met0000537

Schouten, A., Boiger, M., Kirchner-Häusler, A., Uchida, Y., & Mesquita, B. (2020). Cultural differences in emotion suppression in Belgian and Japanese couples: A social functional model. *Frontiers in Psychology*, *11*, Article 1048. https://doi.org/10.3389/fpsyg.2020.01048

Skriner, L. C., & Chu, B. C. (2014). Cross-ethnic measurement invariance of the SCARED and CES-D in a youth sample. *Psychological Assessment*, *26*(1), 332–337. https://doi.org/10.1037/a0035092

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

Spaapen, D. L., Waters, F., Brummer, L., Stopa, L., & Bucks, R. S. (2014). The emotion regulation questionnaire: Validation of the ERQ-9 in two community samples. *Psychological Assessment*, *26*(1), 46–54. https://doi.org/10.1037/a0034474

Su, J. C., Lee, R. M., & Oishi, S. (2013). The role of culture and self-construal in the link between expressive suppression and depressive symptoms. *Journal of Cross-Cultural Psychology*, *44*(2), 316–331. https://doi.org/10.1177/0022022112443413

Tams, S. (2008). Self-directed social learning: The role of individual differences. *Journal of Management Development*, *27*(2), 196–213. https://doi.org/10.1108/02621710810849335

Thalmayer, A. G., & Rossier, J. (2019). 7—Agreeableness, antagonism, and mental health across cultures. In J. D. Miller & D. R. Lynam (Eds.), *The handbook of antagonism* (pp. 97–111). Academic Press.

Thimm, J. C., Jordan, S., & Bach, B. (2017). Hierarchical structure and cross-cultural measurement invariance of the Norwegian version of the Personality Inventory for DSM-5. *Journal of Personality Assessment*, *99*(2), 204–210. https://doi.org/10.1080/00223891.2016.1223682

Torres, L., Miller, M. J., & Moore, K. M. (2013). Factorial invariance of the Brief Symptom Inventory-18 (BSI-18) for adults of Mexican descent across nativity status, language format, and gender. *Psychological Assessment*, *25*(1), 300–305. https://doi.org/10.1037/a0030436

U.S. Census Bureau. (2021). *Educational attainment in the United States: 2021*. Retrieved October 25, 2022, from https://www.census.gov/data/tables/2021/demo/educational-attainment/cps-detailed-tables.html

U.S. Census Bureau. (2022). *QuickFacts: Hawaii*. https://www.census.gov/quickfacts/HI

Vaid, E., & Lansing, A. H. (2020). Discrimination, prejudice, and oppression and the development of psychopathology. In L. T. Benuto, M. P. Duckworth, A. Masuda, & W. O'Donohue (Eds.), *Prejudice, stigma, privilege, and oppression: A behavioral health handbook* (pp. 235–248). Springer. https://doi.org/10.1007/978-3-030-35517-3_14

Watters, C. A., & Bagby, R. M. (2018). A meta-analysis of the five-factor internal structure of the Personality Inventory for DSM-5. *Psychological Assessment*, *30*(9), 1255–1260. https://doi.org/10.1037/pas0000605