American Psychological Association

Psychological Assessment

1040-3590/19/$12.00

2019, Vol. 31, No. 6, 833–838
http://dx.doi.org/10.1037/pas0000708

BRIEF REPORT

Measurement Invariance of Depression Symptom Ratings Across African American, Hispanic/Latino, and Caucasian Adolescent Psychiatric Inpatients

William Mellick
Medical University of South Carolina

Claire Hatkevich
University of Houston

Amanda Venta
Sam Houston State University

Ryan M. Hill
Baylor College of Medicine

Iram Kazimi
University of Texas Health Science Center at Houston

Jon D. Elhai
University of Toledo

Carla Sharp
University of Houston

The Beck Depression Inventory–II (BDI-II) is widely used to assess adolescent depressive symptom severity. Psychometric investigations, including factor-analytic studies, with adolescents support the reliability and validity of the BDI-II. However, a major limitation of this research is that samples have been predominantly Caucasian/White. This is critical because depressive illness is highly prevalent across race and ethnicity, and the extent to which reliability and findings generalize to non-Caucasian populations is in question. The present study recruited African American/Black (n = 96), Hispanic/Latino(a) (n = 151), and Caucasian/White (n = 97) adolescent psychiatric inpatients (M_{age} = 14.73) to test the measurement invariance of the BDI-II, using Osman and colleagues' two-factor solution while also assessing within-group reliability and concurrent validity by examining associations with other symptom measures. Across groups, the two-factor solution, factor loadings, and indicator thresholds were invariant. Within-group reliability estimates were adequate, and the concurrent validity was supported. This suggests BDI-II symptom comparisons between African American/Black, Hispanic/Latino(a), and Caucasian/White adolescent inpatients are valid. Critical extensions of this work may include the examination of potential invariance across depressive symptom clusters via network analysis and invariance testing of depression symptom ratings over time in ethnoracially diverse children and adolescents.

Public Significance Statement
Study findings suggest that Beck Depression Inventory–II (BDI-II) depressive symptom ratings are similarly reliable, valid, and invariant among African American/Black, Caucasian/White, and Hispanic/Latino(a) adolescent psychiatric inpatients. Therefore, these racial-ethnic groups can be meaningfully compared on BDI-II scores.

Keywords: Beck Depression Inventory, race, ethnicity, measurement invariance

Developed for use with adults and adolescents, the Beck Depression Inventory–II (BDI-II; Beck, Steer, & Brown, 1996) is one of the most widely used depressive symptom severity self-report rating scales in clinical and nonclinical settings worldwide (Wang & Gorenstein, 2013). Studies with psychiatric inpatient, outpatient, and nonclinical adolescent samples have shown BDI-II scores to be a reliable and valid measure of depressive symptoms (i.e., Osman, Barrios, Gutierrez, Williams, & Bailey, 2008; Osman, Kopper, Barrios, Gutierrez, & Bagge, 2004; Steer, Ball, Ranieri, & Beck, 1999; Steer, Kumar, Ranieri, & Beck, 1998). However, a major limitation of this research, which includes factor-analytic studies (i.e., Osman et al., 2004, 2008), is that participant samples

This article was published Online First February 25, 2019.

William Mellick, Department of Psychiatry and Behavioral Sciences, Medical University of South Carolina; Claire Hatkevich, Department of Psychiatry, University of Houston; Amanda Venta, Department of Psychology, Sam Houston State University; Ryan M. Hill, Section of Psychology, Department of Pediatrics, Baylor College of Medicine; Iram Kazimi, Department of Psychiatry and Behavioral Sciences, University of Texas Health Science Center at Houston; Jon D. Elhai, Department of Psychology, University of Toledo; Carla Sharp, Department of Psychology, University of Houston.

We thank the teens and their families for their participation in the study.

Correspondence concerning this article should be addressed to Carla Sharp, Department of Psychology, University of Houston, 126 Heyne Building, Houston, TX 77204. E-mail: csharp2@uh.edu

have been racially and ethnically homogeneous, composed predominantly of Caucasian adolescents. Therefore, it is unknown whether extant findings generalize to non-Caucasian patient populations, which is critical because depressive illness is highly prevalent across race and ethnicity (González, Tarraf, Whitfield, & Vega, 2010).

Arguably, the most replicable BDI-II factor structure consists of cognitive-affective and somatic factors similar to the original proposed by Beck and colleagues (Beck et al., 1996; Dozois, Dobson, & Ahnberg, 1998; Wang & Gorenstein, 2013). Particularly relevant to the present adolescent study, Osman and colleagues (2004) conducted several confirmatory factor analyses (CFAs) of extant factorial solutions with a sample of adolescent psychiatric inpatients to determine which solution provided the best model fit. None of the tested models fit the data well, including that of Beck et al. (1996), and instead a revised two-factor cognitive-affective and somatic solution received support. This factor structure was subsequently supported in a sample of nonclinical adolescents in another study by Osman and colleagues (2008). Both studies additionally examined the reliability and concurrent validity of the BDI-II with other psychopathology and suicide risk measures. Reliability for factor scores was adequate (i.e., $\alpha = .78–0.92$), and concurrent validity was supported as BDI-II scores significantly correlated with internalizing symptoms, hopelessness, and suicide-related behaviors. These findings suggest the BDI-II is a reliable and valid measure of adolescent depressive symptom ratings, and this two-factor solution (Osman et al., 2004, 2008) provides good fit to data from adolescent samples.

Whether BDI-II scores are measurement invariant across racially and ethnically diverse adolescent samples importantly remains to be determined. Approximately 46% of American youth belong to racial/ethnic minority groups (Mather, Pollard, & Jacobsen, 2009; Vaughn-Coaxum, Mair, & Weisz, 2015), and depression among these groups is a growing public health concern (i.e., Eaton et al., 2011). For instance, as compared to Caucasian/White adolescents, Hispanic/Latino(a) adolescents reported elevated depressive symptoms and suicide risk while African American/Black adolescents, in turn, are indicated to have greater depressive symptom-related impairment and poorer functional treatment outcomes (i.e., Brown, Schulberg, Sacco, Perel, & Houck, 1999; Lowry, Crosby, Brener, & Kann, 2014).

For these reasons, among others, it is important to know whether BDI-II scores measure the same constructs in the same manner across adolescents of differing racial/ethnic backgrounds (Chen & West, 2008). Different experience of depression-related constructs (i.e., internal vs. external locus of control, self-efficacy, etc.), different interpretation of item content, and expressive differences in depressive symptomology may contribute to BDI-II measurement invariance (Ayalon & Young, 2003; Milfont & Fischer, 2010; Mosotho, Louw, Calitz, & Esterhuyse, 2008). A fairly recent investigation by Vaughn-Coaxum and colleagues (2015) found racial/ethnic noninvariance for youth depression symptom ratings with another common measure indicating a potential need for revised scoring and/or interpretation. Similar recommendations may need to be made for the BDI-II if it is indeed found to be noninvariant.

Against this background, the present study recruited self-identified African American/Black, Hispanic/Latino(a), and Caucasian/White adolescent psychiatric inpatients to test the measurement invariance of the BDI-II using Osman and colleagues' (2004, 2008) two-factor solution. Demonstration of nonequivalence for other youth depression symptom ratings (Vaughn-Coaxum et al., 2015) tempered expectations that BDI-II racial-ethnic invariance would be found in the present study as it has in young-adult samples (Hambrick et al., 2010; Whisman, Judd, Whiteford, & Gelhorn, 2013). Within-group evaluation of BDI-II scores' reliability and concurrent validity with other symptom measure scores was an exploratory second aim.

Method

In total, 405 adolescent inpatients were recruited; however, 35 belonged to racial/ethnic groups of insufficient sample size for invariance testing (i.e., multiracial, $n = 24$; Asian, $n = 6$; Native American, $n = 1$) and 26 were excluded for not completing any portion of the BDI-II. Item-level data were imputed (mean item response values) for cases with $< 20\%$ missing data ($n = 27$). Twenty-five of these participants were missing data for two or fewer items. Imputation was necessary to maximize subsample sizes, which is consistent with established methods of practice and does not degrade the statistical models tested (Graham, 2009; Schafer & Graham, 2002). Failure to complete supplemental study measures (to examine BDI-II concurrent validity) did not result in study exclusion. A final sample of $N = 344$ adolescents was included: African American/Black (non-Hispanic; $n = 96$), Caucasian/White (non-Hispanic; $n = 97$), and Hispanic/Latino(a) ($n = 151$). Among Hispanic/Latino(a) adolescents, 79.5% originated from the United States with 16.6% from Mexico, 1.3% from Guatemala, and 0.7% from Bolivia, Columbia, El Salvador, and Honduras, respectively. Thirty percent of Hispanic/Latino(a) adolescents self-identified as first-generation Americans, with 19.2% and 11.9% identifying as second- and third-generation, respectively (36.4% did not identify their generation).

The present study received institutional review board approval. Participants were recruited from the adolescent acute inpatient unit of a county psychiatric hospital serving the greater Houston, Texas, metropolitan area. Parents provided consent in English or Spanish, and adolescents were then approached for assent. Assessments were typically completed within 2–3 days of admission in a private setting by a doctoral-level clinical psychology graduate student who received supervised training by the principal investigator. Inclusion criteria required participants to be between 12 and 17 years old, possess English fluency, and, at minimum, have a fifth-grade reading level as determined by the Wide Range Achievement Test 4 (Wilkinson & Robertson, 2006). Inpatients with intellectual disability or any psychotic spectrum disorder were not recruited. Participation was voluntary and withdrawal was permitted at any time.

Participants completed a sociodemographic questionnaire and the BDI-II (Beck et al., 1996), among other symptom measures, including the Youth Self-Report (YSR; Achenbach & Rescorla, 2001) and the Interpersonal Needs Questionnaire (Van Orden, Witte, Gordon, Bender, & Joiner, 2008), a measure of thwarted belongingness and perceived burdensomeness. Clinicians administered the Modified Scale for Suicidal Ideation (Miller, Norman, Bishop, & Dow, 1986). Scores from these measures were used to examine BDI-II concurrent validity.

Group comparisons on sociodemographic and symptom variables were performed using chi-square tests of independence, latent factor mean comparisons, and one-way analyses of variance with post hoc Tukey's tests. Effect sizes were reported as Cramer's *v* and partial-eta squared, respectively. A structural equation modeling (SEM) approach was used to calculate within-group reliability estimates for Cronbach's alpha (α) with 95% confidence intervals (CIs; Raykov & Marcoulides, 2015). Bivariate Pearson correlations were used to examine associations between BDI-II factor scores and concurrent measure scores.

Osman and colleagues' (2004, 2008) two-factor solution included cognitive-affective (Items 1–9, 12–14) and somatic (Items 11, 15–21) factors. Since Item 10 ("Crying") did not adequately load onto either factor (≤0.32; Osman et al., 2004), it was excluded. An SEM approach with ordinal items, polychoric covariance matrices, and probit factor loadings was taken with CFAs (Hambrick et al., 2010). A weighted least squares estimator with a mean and variance (WLSMV)–adjusted chi-square was employed (Bowen & Masa, 2015). Chi-square goodness-of-fit test (χ^2), root mean square error of approximation (RMSEA), comparative fit index (CFI), and Tucker–Lewis index (TLI) were assessed for model fit using traditionally accepted standards. Acceptable model fit required CFI ≥ 0.90, TLI ≥ 0.90, and RMSEA ≤ 0.08, whereas excellent model fit required CFI ≥ 0.95, TLI ≥ 0.95, and RMSEA ≤ 0.06 (Hu & Bentler, 1999). The Mplus DIFFTEST option was used to obtain a corrected chi-square difference test since WLSMV chi-square values do not fit a chi-square distribution (Muthén & Muthén, 1998–2012). This option simply exports adjusted derivatives file for model comparisons. A significant chi-square difference test did not unequivocally support the less restrictive model due to known limitations of this test (Meade, Johnson, & Braddy, 2008; Milfont & Fischer, 2010). CFI was also considered in evaluating model comparisons since it conveys the magnitude of the effect of parameter constraints and provides a better fit index for smaller sample sizes (Hu & Bentler, 1999). Configural, metric (weak factorial), and scalar (strong factorial) measurement invariances were tested. Strict invariance (equivalence of factor and error variances) went untested as it is commonly untenable in social science research (Bowen & Masa, 2015). SPSS v. 19 (IBM Corp, 2010) and MPlus v. 7.2 (Muthén & Muthén, 1998–2012) were used for analyses.

Results and Discussion

Attrition Analyses

Sociodemographic differences between study completers and noncompleters were tested given the significant number of participants without BDI-II data ($n = 26$). Groups did not differ on age, $t(368) = 0.66$, $p = .511$, Cohen's $d = 0.132$; sex, $\chi^2(1, N = 370) = 1.22$, $p = .268$, Cramer's $v = 0.058$; or group assignment, $\chi^2(2, 370) = 3.42$, $p = .181$, Cramer's $v = 0.096$. Attrition effects were not further explored given these nonsignificant findings.

Sample Characteristics

Table 1 lists participant sociodemographic characteristics, co-occurring psychiatric symptoms, and BDI-II descriptive statistics. BDI-II total scores indicated moderate depressive symptoms among participants. The marginally significant group difference on BDI-II total scores was to be followed up after invariance testing. Post hoc Tukey's tests on YSR Anxiety Problem scores showed only trend-level group differences despite a significant omnibus test.

Factor Reliability Estimates

Within-group reliability estimates as measured by Cronbach's alpha (α) were calculated for each BDI-II factor. Reliability was similarly adequate across groups for the cognitive-affective factor: African American/Black, α = .92, 95% CI [0.89, 0.94]; Caucasian/White, α = .93, 95% CI [0.90, 0.94]; and Hispanic/Latino(a), α = .92, 95% CI [0.89, 0.94]. Within-group somatic factor reliability was as follows: African American/Black, α = .77, 95% CI [0.69, 0.83]; Caucasian/White, α = .80, 95% CI [0.73, 0.84]; and Hispanic/Latino(a), α = .72, 95% CI [0.64, 0.77]. Thus, there was greater variability among groups' somatic factor reliability estimates. However, these reliabilities are all adequate for this scale length (Slobodskaya, 2007).

CFAs

The distributional properties of individual items were examined to justify the modeling approach prior to performing CFAs. No

Table 1
Descriptive Characteristics and Between-Group Comparison Results

Variable	Combined sample ($n = 344$)	African American/ Black ($n = 96$)	Caucasian/ White ($n = 97$)	Hispanic/ Latino(a) ($n = 151$)	F/χ^2	p	η^2/v
Age (years)	14.73 (1.47)	14.73 (1.59)	14.57 (1.52)	14.83 (1.36)	.93	.396	.005
Sex (% female)	62.20	63.50	63.90	60.30	.44	.485	.003
BDI-II	22.60 (13.16)	20.26 (13.36)	24.89 (14.17)	22.61 (12.16)	3.02	.050	.017
Anxiety SXs	60.12 (8.42)	58.08 (8.06)	61.07 (8.13)	60.75 (8.67)	3.46	.033	.022
ADHD SXs	61.17 (8.05)	60.94 (8.07)	62.00 (8.21)	60.76 (7.96)	.68	.509	.004
ODD SXs	61.06 (8.58)	60.77 (8.58)	61.89 (8.56)	60.71 (8.62)	.58	.563	.004
Conduct SXs	65.20 (10.09)	65.50 (10.40)	65.09 (9.08)	65.08 (10.59)	.05	.950	.000

Note. Data are means (standard deviations) unless otherwise specified. Effect sizes reported in partial-eta squared (η^2) and Cramer's *v*. Combined sample data for reference only. BDI-II = Beck Depression Inventory–II total raw score; Anxiety SXs = YSR Anxiety Problems *t* score; ADHD SXs = YSR Attention-Deficit-Hyperactivity-Disorder Problems *t* score; ODD SXs = YSR Oppositional Defiant Disorder Problems *t* score; Conduct SXs = YSR Conduct Problems *t* score.

item within any group had skewness or kurtosis ≥ ± 2 indicating normality of indicator variables. Independent CFAs were then performed to provide assurance that the two-factor model adequately fit to each subsample's data (Milfont & Fischer, 2010). Factor means fixed to 0 and factor and item residual variances fixed to 1. The model provided excellent fit to each group's data: African American/Black, $\chi^2(169, N = 96) = 222.69$, $p = .003$, CFI = 0.970, TLI = 0.966, RMSEA = 0.058; Hispanic/Latino(a), $\chi^2(169, N = 151) = 207.28$, $p = .024$, CFI = 0.984, TLI = 0.982, RMSEA = 0.039; and Caucasian/White, $\chi^2(169, N = 97) = 234.20$, $p < .001$, CFI = 0.970, TLI = 0.967, RMSEA = 0.063.

See Table 2 for standardized item factor loadings. High somatic factor loadings were found for fatigue (0.78), loss of energy (0.70), and sleep change (0.70), irrespective of subsample, which is consistent with prior adolescent findings by Osman et al. (2004, 2008) with predominantly Caucasian samples. Like Osman et al. (2008), cognitive-affective factor loadings were notably high on self-dislike (0.81) and self-criticalness (0.81). However, the strong loading for sadness (0.79) is more consistent with other adolescent studies (i.e., Dozois et al., 1998; Osman et al., 2004). Methodological differences between the present study and young-adult BDI-II measurement invariance studies (Hambrick et al., 2010; Whisman et al., 2013) make for difficult comparisons. However, punishment feelings (0.39–0.44), which had the lowest cognitive-affective factor loadings, showed discrimination differential item functioning between White and African American participants in Whisman et al. (2013). As the authors suggested, removing this item may improve cross-racial/ethnic application of the BDI-II.

Measurement Invariance

A series of sequentially restrictive models were fitted to test for measurement invariance with a significant chi-square difference test ($p < .05$) and/or ΔCFI ≤ −0.01 indicating worse comparable fit of the more restrictive model and support of the less restrictive model (Meade et al., 2008). Configural invariance (baseline model) provided excellent fit to the data, CFI = .975, TLI = .972, RMSEA = .052, meaning the constructs had similar patterns of free and fixed loadings across groups (Putnick & Bornstein, 2016). Metric (weak factorial) invariance was subsequently tested whereby factor variances remained freely estimated but factor loadings were held invariant. Fit indices showed the more parsimonious model provided acceptable fit with this constraint as indicated by a nonsignificant chi-square difference test, $\chi^2(36) = 40.18$, and little ΔCFI (0.007). The items therefore loaded onto factors similarly across groups (Putnick & Bornstein, 2016). Scalar (strong factorial) invariance was then tested whereby indicator thresholds were now also held invariant. Although a significant chi-square difference test was observed, $\chi^2(116) = 168.03$, ΔCFI was only −0.004. This more parsimonious model was favored per comparative-fit decision criteria.

This demonstration of measurement invariance indicates that BDI-II symptom comparisons between African American/Black, Caucasian/White, and Hispanic/Latino(a) adolescent inpatients are indeed valid and that observed raw score differences indicate meaningful differences in latent depressive symptom severity (Meredith & Teresi, 2006; Vaughn-Coaxum et al., 2015). Furthermore, this suggests that past studies that used the BDI-II and demonstrated high incidence and prevalence rates, increased functional impairment, and/or high risk for suicide among African American/Black and Hispanic/Latino(a) adolescents relative to Caucasian/White adolescents may reflect true, pressing public health needs. In contrast with adult findings by Brown and colleagues (1999) using the original BDI, Caucasian/White partici-

Table 2
Standardized Item Factor Loadings

Variable	Combined sample		African American/ Black		Caucasian/White		Hispanic/Latino(a)	
	C-A	S	C-A	S	C-A	S	C-A	S
1. Sadness	.79		.83		.82		.73	
2. Pessimism	.69		.70		.67		.69	
3. Past failure	.73		.79		.70		.72	
4. Loss of pleasure	.68		.66		.70		.69	
5. Guilty feelings	.61		.59		.62		.65	
6. Punishment feelings	.43		.44		.48		.39	
7. Self-dislike	.81		.73		.85		.82	
8. Self-criticalness	.81		.71		.87		.84	
9. Suicidal thoughts	.74		.81		.76		.67	
12. Loss of interest	.72		.72		.75		.72	
13. Indecisiveness	.68		.62		.80		.67	
14. Worthlessness	.85		.88		.87		.82	
17. Irritability	.66		.76		.75		.62	
11. Agitation		.70		.70		.73		.70
15. Loss of energy		.70		.79		.69		.64
16. Sleep change		.70		.82		.66		.62
18. Appetite change		.62		.65		.68		.57
19. Concentration		.67		.60		.74		.66
20. Fatigue		.78		.76		.88		.70
21. Loss of interest in sex		.30		.40		.28		.26

Note. Item 10 excluded per Osman, Kopper, Barrios, Gutierrez, and Bagge (2004, 2008). C-A = Cognitive-affective factor; S = Somatic factor.

pants scored higher than African American/Black participants on both cognitive-affective ($p = .011$) and somatic factors ($p = .019$). However, the extent to which these discrepancies are meaningful is unclear. There may have been substantial diagnostic heterogeneity between present study groups. For instance, more Caucasian/White adolescents may have met criteria for major depressive disorder, which could explain their higher latent factor scores.

Notably, demonstration of measurement invariance differs from Vaughn-Coaxum and colleagues (2015), who found Children's Depression Inventory (Kovacs, 1992) scores to be noninvariant across a community sample of White, Black, Latino, and Asian youth. A number of factors beyond the rating scale utilized, including age and recruitment setting, may account for discrepant findings. The inpatient sample in the present study, for example, may have reported a narrower range of symptom ratings than the community sample in Vaughn-Coaxum et al. (2015). Subsample sizes were also substantially larger in Vaughn-Coaxum et al. (2015), which may have improved statistical power to detect differences. Moreover, an item response theory approach to invariance testing was utilized rather than SEM.

Concurrent Validity

See Table 3 for within-group bivariate Pearson correlations of BDI-II scores with other study measures. YSR Somatic Problems uniquely associated ($p < .001$) with BDI-II scores in Hispanic/Latino(a) adolescents, echoing the notion that somatic symptom expression correlates strongly with depressive symptomatology in this racial-ethnic group (Canino, 2004). Bivariate associations across subsamples generally support the concurrent validity of BDI-II scores, implying that comparison of depressive symptom correlates, or processes, between these groups is appropriate and valid (Whisman et al., 2013).

Table 3
Within-Group Bivariate Pearson Correlations Between BDI-II Scores and Other Study Measures

Variable	BDI-II total		
	African American/Black	Caucasian/White	Hispanic/Latino(a)
YSR Affective Prob.	.70**	.73**	.72**
YSR Somatic Prob.	.29	.19	.34**
	($n = 84$)	($n = 89$)	($n = 136$)
MSSI	.29*	.63**	.59**
	($n = 90$)	($n = 92$)	($n = 148$)
INQ Thwarted Belong.	.29*	.28	.28**
	($n = 92$)	($n = 91$)	($n = 141$)
INQ Perceived Burden.	.36*	.43**	.45**
	($n = 62$)	($n = 66$)	($n = 103$)

Note. Number of respondents per group following pairwise deletion. Item 9 (suicide) was excluded in calculating Beck Depression Inventory–II (BDI-II) total scores. YSR Affective Prob. = Youth Self-Report Affective Problems *t* score; YSR Somatic Prob. = Youth Self-Report Somatic Problems *t* score; MSSI = Modified Scale for Suicidal Ideation total score; INQ Thwarted Belong. = Interpersonal Needs Questionnaire Thwarted Belongingness score; INQ Perceived Burden. = Interpersonal Needs Questionnaire Perceived Burdensomeness score.
* $p \leq .005$ and ** $p \leq .001$ after Bonferroni correction for multiple comparisons.

This study was not without limitations. Recruiting 200-plus participants per group would have been preferable (Meade et al., 2008). However, increasing sample sizes would not inherently change the likelihood of detecting measurement noninvariance. This notion stems from an overwhelming reliance among past studies on χ^2, which is heavily influenced by sample sizes, as a primary indicator of comparative model fit (Putnick & Bornstein, 2016). Moreover, sample size is not associated with each step of invariance testing. The present study's prioritization of a strict CFI cutoff was chosen to help mitigate this concern (Hu & Bentler, 1999). Psychiatric diagnoses were lacking, which may have been informative, and could have been used to further evaluate BDI-II score validity. Participants reported co-occurring psychiatric symptoms that may exceed those observed by patients in other treatment settings. This may limit generalizability or be a strength of findings as "pure" depression is often the exception rather than the norm (Garber & Weersing, 2010). Future ethnoracial invariance psychometric testing of the BDI-II with adolescents from other treatment settings is certainly warranted.

Limitations notwithstanding, present study findings suggest that BDI-II symptom ratings are reliable and valid indicators of depression across African American/Black, Caucasian/White, and Hispanic/Latino(a) inpatient adolescents. Demonstrating measurement invariance showed the same depression construct was measured, with similar associations between indicators and latent factors, and similar item-level depressive symptom severity among groups. Paired with findings of strong reliability and concurrent validity, comparing these racial and ethnic groups on depressive symptom ratings and depressive symptom correlates on the BDI-II is meaningful and could well may inform future research, clinical practice, and initiatives to address growing mental health needs among African American/Black and Hispanic/Latino(a) adolescents.

References

Achenbach, T. M., & Rescorla, L. A. (2001). *Manual for the Achenbach system of empirically based assessment school-age forms profiles*. Burlington, VT: Aseba.

Ayalon, L., & Young, M. A. (2003). A comparison of depressive symptoms in African Americans and Caucasian Americans. *Journal of Cross-Cultural Psychology, 34,* 111–124. http://dx.doi.org/10.1177/0022022102239158

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

Bowen, N. K., & Masa, R. D. (2015). Conducting measurement invariance tests with ordinal data: A guide for social work researchers. *Journal of the Society for Social Work and Research, 6,* 229–249. http://dx.doi.org/10.1086/681607

Brown, C., Schulberg, H. C., Sacco, D., Perel, J. M., & Houck, P. R. (1999). Effectiveness of treatments for major depression in primary medical care practice: A post hoc analysis of outcomes for African American and white patients. *Journal of Affective Disorders, 53,* 185–192. http://dx.doi.org/10.1016/S0165-0327(98)00120-7

Canino, G. (2004). Are somatic symptoms and related distress more prevalent in Hispanic/Latino youth? Some methodological considerations. *Journal of Clinical Child and Adolescent Psychology, 33,* 272–275. http://dx.doi.org/10.1207/s15374424jccp3302_8

Chen, F. F., & West, S. G. (2008). Measuring individualism and collectivism: The importance of considering differential components, refer-

ence groups, and measurement invariance. *Journal of Research in Personality, 42,* 259–294. http://dx.doi.org/10.1016/j.jrp.2007.05.006

Dozois, D. J., Dobson, K. S., & Ahnberg, J. L. (1998). A psychometric evaluation of the Beck Depression Inventory–II. *Psychological Assessment, 10,* 83–89. http://dx.doi.org/10.1037/1040-3590.10.2.83

Eaton, D. K., Foti, K., Brener, N. D., Crosby, A. E., Flores, G., & Kann, L. (2011). Associations between risk behaviors and suicidal ideation and suicide attempts: Do racial/ethnic variations in associations account for increased risk of suicidal behaviors among Hispanic/Latina 9th- to 12th-grade female students? *Archives of Suicide Research, 15,* 113–126. http://dx.doi.org/10.1080/13811118.2011.565268

Garber, J., & Weersing, V. R. (2010). Comorbidity of anxiety and depression in youth: Implications for treatment and prevention. *Clinical Psychology: Science and Practice, 17,* 293–306. http://dx.doi.org/10.1111/j.1468-2850.2010.01221.x

González, H. M., Tarraf, W., Whitfield, K. E., & Vega, W. A. (2010). The epidemiology of major depression and ethnicity in the United States. *Journal of Psychiatric Research, 44,* 1043–1051. http://dx.doi.org/10.1016/j.jpsychires.2010.03.017

Graham, J. W. (2009). Missing data analysis: Making it work in the real world. *Annual Review of Psychology, 60,* 549–576. http://dx.doi.org/10.1146/annurev.psych.58.110405.085530

Hambrick, J. P., Rodebaugh, T. L., Balsis, S., Woods, C. M., Mendez, J. L., & Heimberg, R. G. (2010). Cross-ethnic measurement equivalence of measures of depression, social anxiety, and worry. *Assessment, 17,* 155–171. http://dx.doi.org/10.1177/1073191109350158

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

IBM Corp. (2010). *IBM SPSS statistics for windows* (Version 19.0). Armonk, NY: Author.

Kovacs, M. (1992). *The Children's Depression Inventory (CDI) Manual*. North Tonawanda, NY: Multi-Health Systems, Inc.

Lowry, R., Crosby, A. E., Brener, N. D., & Kann, L. (2014). Suicidal thoughts and attempts among U.S. high school students: Trends and associated health-risk behaviors, 1991–2011. *Journal of Adolescent Health, 54,* 100–108. http://dx.doi.org/10.1016/j.jadohealth.2013.07.024

Mather, M., Pollard, K., & Jacobsen, L. A. (2009). *Reports America*. Washington, DC: Population Reference Bureau.

Meade, A. W., Johnson, E. C., & Braddy, P. W. (2008). Power and sensitivity of alternative fit indices in tests of measurement invariance. *Journal of Applied Psychology, 93,* 568–592. http://dx.doi.org/10.1037/0021-9010.93.3.568

Meredith, W., & Teresi, J. A. (2006). An essay on measurement and factorial invariance. *Medical Care, 44*(Suppl. 3), S69–S77. http://dx.doi.org/10.1097/01.mlr.0000245438.73837.89

Milfont, T. L., & Fischer, R. (2010). Testing measurement invariance across groups: Applications in crosscultural research. *International Journal of Psychological Research*. Advance online publication. http://dx.doi.org/10.21500/20112084.857

Miller, I. W., Norman, W. H., Bishop, S. B., & Dow, M. G. (1986). The Modified Scale for Suicidal Ideation: Reliability and validity. *Journal of Consulting and Clinical Psychology, 54,* 724–725. http://dx.doi.org/10.1037/0022-006X.54.5.724

Mosotho, N. L., Louw, D. A., Calitz, F. J., & Esterhuyse, K. G. (2008). Depression among Sesotho speakers in Mangaung, South Africa. *African Journal of Psychiatry, 11,* 35–43. http://dx.doi.org/10.4314/ajpsy.v11i1.30253

Muthén, L. K., & Muthén, B. O. (1998–2012). *Mplus user's guide* (7th ed.). Los Angeles, CA: Author.

Osman, A., Barrios, F. X., Gutierrez, P. M., Williams, J. E., & Bailey, J. (2008). Psychometric properties of the Beck Depression Inventory–II in nonclinical adolescent samples. *Journal of Clinical Psychology, 64,* 83–102. http://dx.doi.org/10.1002/jclp.20433

Osman, A., Kopper, B. A., Barrios, F., Gutierrez, P. M., & Bagge, C. L. (2004). Reliability and validity of the Beck Depression Inventory–II with adolescent psychiatric inpatients. *Psychological Assessment, 16,* 120–132. http://dx.doi.org/10.1037/1040-3590.16.2.120

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. http://dx.doi.org/10.1016/j.dr.2016.06.004

Raykov, T., & Marcoulides, G. A. (2015). A direct latent variable modeling based method for point and interval estimation of coefficient alpha. *Educational and Psychological Measurement, 75,* 146–156. http://dx.doi.org/10.1177/0013164414526039

Schafer, J. L., & Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Methods, 7,* 147–177. http://dx.doi.org/10.1037/1082-989X.7.2.147

Slobodskaya, H. R. (2007). The associations among the big five, behavioural inhibition and behavioural approach systems and child and adolescent adjustment in Russia. *Personality and Individual Differences, 43,* 913–924. http://dx.doi.org/10.1016/j.paid.2007.02.012

Steer, R. A., Ball, R., Ranieri, W. F., & Beck, A. T. (1999). Dimensions of the Beck Depression Inventory–II in clinically depressed outpatients. *Journal of Clinical Psychology, 55,* 117–128. http://dx.doi.org/10.1002/(SICI)1097-4679(199901)55:1<117::AID-JCLP12>3.0.CO;2-A

Steer, R. A., Kumar, G., Ranieri, W. F., & Beck, A. T. (1998). Use of the Beck Depression Inventory–II with adolescent psychiatric outpatients. *Journal of Psychopathology and Behavioral Assessment, 20,* 127–137. http://dx.doi.org/10.1023/A:1023091529735

Van Orden, K. A., Witte, T. K., Gordon, K. H., Bender, T. W., & Joiner, T. E., Jr. (2008). Suicidal desire and the capability for suicide: Tests of the interpersonal-psychological theory of suicidal behavior among adults. *Journal of Consulting and Clinical Psychology, 76,* 72–83. http://dx.doi.org/10.1037/0022-006X.76.1.72

Vaughn-Coaxum, R. A., Mair, P., & Weisz, J. R. (2015). Racial/Ethnic differences in youth depression indicators: An item response theory analysis of symptoms reported by white, black, Asian, and Latino youths. *Clinical Psychological Science, 4,* 239–253. http://dx.doi.org/10.1177/2167702615591768

Wang, Y. P., & Gorenstein, C. (2013). Psychometric properties of the Beck Depression Inventory–II: A comprehensive review. *Revista Brasileira de Psiquiatria, 35,* 416–431. http://dx.doi.org/10.1590/1516-4446-2012-1048

Whisman, M. A., Judd, C. M., Whiteford, N. T., & Gelhorn, H. L. (2013). Measurement invariance of the Beck Depression Inventory–2nd ed. (BDI-II) across gender, race, and ethnicity in college students. *Assessment, 20,* 419–428. http://dx.doi.org/10.1177/1073191112460273

Wilkinson, G. S., & Robertson, G. J. (2006). *WRAT 4: Wide Range Achievement Test*. Lutz, FL: Psychological Assessment Resources.

Received May 10, 2018
Revision received December 6, 2018
Accepted January 8, 2019 ■