Factorial Validity and Invariance of an Adolescent Depression Symptom Screening Tool

Ashley S. Long, PhD, ATC
Research Consultant
Atrium Health, Department of Family Medicine
Charlotte, NC, USA
ashleylong111@gmail.com

JD DeFresse, PhD
Assistant Professor
University of North Carolina-Chapel Hill, Center for the Study of Retired Athletes
Chapel Hill, NC, USA
defreese@email.unc.edu

Allison K. Bickett, PhD
Director of Behavioral Health Education
Atrium Health, Department of Family Medicine
Charlotte, NC, USA
Allison.Bickett@atriumhealth.org

David E. Price, MD
Program Director, Primary Care Sports Medicine Fellowship
Professor of Family Medicine
Atrium Health, Department of Family Medicine
Correspondence:

Ashley Long, PhD, LAT, ATC,
Research Consultant
Atrium Health, Department of Family Medicine
CMC-Mercy,
MMP, Suite 400 B
2001 Vail Avenue
Charlotte, NC 28207
(828) 896-7803
Ashleylong111@gmail.com

Readers should keep in mind that the in-production articles posted in this section may undergo changes in the content and presentation before they appear in forthcoming issues. We recommend regular visits to the site to ensure access to the most current version of the article. Please contact the JAT office (jat@slu.edu) with any questions.
Factorial Validity and Invariance of an Adolescent Depression Symptom Screening Tool

Context: Depression is among the most common mental health disorder in youth, results in significant impairment, and is associated with a higher risk of suicide. Screening is essential but assessment tools may not be appropriate across races or do not account for the complex interrelatedness of various demographics including gender, socio-economic status and race.

Objectives: (1) To determine the factor structure of the Patient Health Questionnaire-Adolescent (PHQ-A) for measuring depression in a group of adolescent athletes; and (2) to determine measurement invariance between Blacks and Whites on the PHQ-A.

Design: Retrospective cohort design.

Setting: Data obtained from a secure database collected at a free, comprehensive, mass pre-participation physical exam (PPE) event hosted by a large health care system.

Participants: Participants included 683 high school athletes (Black n=416; White n=267). Independent variables included somatic and affective factors contributing to the construct of depression measured by the PHQ-A and participant race (Black and White).

Main Outcome Measures: (1) Factors upon which the construct of depression is measured and (2) measurement invariance between Blacks and Whites.

Results: A two-factor model, including affective and somatic components, was specified and exhibited an adequate fit to the data (CFI > .90). All items exhibited moderate to high squared multiple correlation...
values ($R^2 = .10-.65$), suggesting that these items resonated relatively well with participants. The two-factor model demonstrated noninvariance Black and White participants (RMSEA = .06-.08).

Conclusions: Overall, the structure of the PHQ-A is supported by a two-factor model in adolescent athletes, measuring both affective and somatic symptoms of depression. A two-factor PHQ-A structure is not fully invariant for the adolescents sampled across participant groups, implying that the model functions differently between Blacks and Whites sampled.

Key Words: adolescent athlete, depression, mental health screening

Abstract Word Count: 300

Manuscript Word Count: 3099

Key Points:

- The PHQ-A’s two-factor model of somatic and affective symptoms is supported in a group of adolescent athletes.
- The PHQ-A is a tool that elicits different response patterns between Black and White adolescent athletes.
Depression is the most common illness and leading cause of disability worldwide. It is a major contributor to the overall global burden of disease\(^1\) and often emerges during adolescence. With the recent pandemic of COVID-19, athletes have demonstrated an increase in perceived stress and dysfunctional psychobiosocial states\(^2\) and the full effects are likely not yet realized. The American Academy of Pediatrics recommends an annual universal screening for depression in youth 12 and over at health maintenance visits\(^3\) as early screening and treatment is vital to addressing the significant health consequences and long-term healthcare cost. Although screening is recommended, access to otherwise healthy adolescent patients is often evasive. An opportunity exists to access healthy adolescents for mental health screening through the sport pre-participation exams (PPEs) which are mandatory in 50 states.\(^3\) Many PPE forms for American high school athletes include some form of mental health screening.

The American Medical Society for Sports Medicine suggests early recognition provides opportunity for prevention of more serious effects of depression and that providers utilize valid screening tools. The PHQ-A is a broadly used, well-validated depression screening tool\(^4,5\) yet, there is little information regarding the influence of race on the construct of the PHQ-A including its factor structure. A clear understanding of the function of the screening tool is necessary for appropriate use, accurate diagnosis, and proper referral in a diverse population of adolescent athletes. Sports Medicine health care providers will likely encounter a racially diverse population in need of mandatory PPE’s.

Conceptualizations of depression among Black adolescents may differ from Whites\(^6\). Without examination of the function of the PHQ-A in a racially diverse adolescent athlete population, we risk inappropriate interpretation and conclusions from its results.

The standard of care during the adolescent sports PPE now includes assessment of mental health (e.g., depression) and sport participation continues to be racially diverse, therefore it is important to confirm that the construct measurement of the PHQ-A is consistent between racial groups. We aimed
to examine (1) the factor structure of the measure of depression on the PHQ-A in a group of adolescent athletes; and (2) measurement invariance between Blacks and Whites on the PHQ-A.

METHODS

Setting: We conducted a retrospective review of PPE screening. Adolescent athletes were allowed to register for the PPE online five weeks prior to the event. Parents or guardians of those students 17 years of age and younger, or students themselves if 18 years or older, were instructed to complete the PPE Medical History and demographic information of the participant. The online form is populated into a secure online portal. Information collected included age, sex, race, school, sport, and significant past medical history. Then, the student-athletes participated in a mass PPE at a large health care system in a metropolitan region of the southeast. PPE activities included a specific order for all participants including assessment of vitals, vision testing, a review of the athlete’s medical history and depression screening (PHQ-A), a sports-specific medical and musculoskeletal exam, an electrocardiogram and clerical check out. Sports medicine physicians, cardiologists, registered dieticians, behavioral health specialists, and athletic trainers were onsite for participants who required follow-up specialized care. The instructions for the PHQ-A were read to students by trained administrators. Then, participants completed the form electronically and individually by reading and responding to the questions on a laptop computer with a trained administrator available for questions. Health information obtained during the PPE was uploaded to secure servers into a secure platform, password protected, and may only be accessed by those who meet specific criteria, including the health care system’s employed athletic trainers, physicians, and privilege-granted researchers. The platform is compliant with federal privacy, storage, and transmission standards.

Participants: Participants were defined as any individual participating in school-sanctioned high school sport who attended the PPE event in person. Students participating in the event attended a public,
charter, or private school with a contract for athletic training services with the healthcare system. There were a total of 897 participants, 683 of whom met the inclusion criteria for the current study based on their self-identified race/ethnicity. Current study athletes' ages ranged from 13 to 18 years, (\(x=15.6, \pm 0.986\)). Participants were 70.3% male \((n=480)\) and 29.7% female \((n=203)\). Since we were aiming to examine the construct regarding Black and White students in particular, only those who self-identified as Black \((n=416; 60.9\%)\) or White \((n=266; 39.1\%)\) were included in the current study. More specifically, the participants were Black Females \((n=125; 18.3\%)\), Black Males \((n=291; 42.6\%)\), White Females \((n=78; 11.4\%)\) and White Males \((n=189, 27.7\%)\). Females accounted for 30% of the Black participants and 29.2% of the White participants, making Black and White groups similar in percentage of each sex. Participants required their own transportation to their school on the morning of the event, then were bussed to the site of the PPE. The study was approved by our Institution's Institutional Review Board (IRB) with a waiver of consent due to the retrospective nature of the study and removal of patient identifying information.

**Outcome Measures:** The primary measures for this investigation were depressive symptoms as measured by the PHQ-A. The PHQ-A was developed by Johnson, Harris, Spitzer, & Williams\(^5\) in 2008 to screen depressive symptoms among adolescents. The PHQ-A can be used to assess the core criteria of depression according to the DSM IV, but is not a substitute for diagnosis. The PHQ-A requires participants to consider the past 2 weeks and answer 9 questions about their perceptions and feelings related to factors known to indicate depression, scoring each question “0” meaning “Not at all” to “3” meaning “Nearly every day”. Scores are totaled and can range between 0-27, with 27 being the most severe level of depressive symptoms. Category designation includes 0-4 as negligible depressive symptoms, 5-9 as mild depressive symptoms, 10-14 as moderate depressive symptoms, 15-19 as moderately severe depressive symptoms, and 20-27 as severe depressive symptoms. Scores of 10 or higher are often designated as meeting criteria of depression and warrant referral to a mental health
provider. Any indication of suicidal thoughts or plans constitute an emergency and a plan should be enacted immediately. Research has shown that the PHQ-A demonstrates excellent internal reliability (Cronbach’s α = 0.89), satisfactory specificity (84-95%), sensitivity (68-95%), and likelihood ratio (6.0-13.6). Demographic information was obtained from the electronic PPE form.

Data Analysis: Prior to structural analyses, data were analyzed for missing values, outliers, and violations of multivariate assumptions. Three main data analytic approaches used to examine the structure and invariance of the PHQ-A in the current study were exploratory factor analysis (EFA), confirmatory factor analysis (CFA), and multi-group confirmatory factor analysis (MGFA). To ensure the EFA and CFA were conducted with unique participants, two random sub-samples of participants were created using a random split command. Group difference analysis (i.e., chi-square) were used to ensure the sub-samples did not differ on gender, race, or PHQ-A diagnostic classification (see Table 1). All statistical analyses and EFA analyses were completed via SPSS 25 while all CFA and MGFA analyses were completed via the AMOS 26 software extension to SPSS.

Consistent with best practice, we conducted factor analyses using maximum likelihood estimation and fixing one loading value to 1.0 for each latent variable. We used, in accord with best practice, multiple fit indices, path coefficients, and modification indices to assess and compare model fit. Specifically, the following fit indices: chi-squared goodness-of-fit, goodness-of-fit index (GFI), Tucker-Lewis index (TLI), comparative fit index (CFI), and root mean square error of approximation (RMSEA) were used. Chi-square significance was interpreted with caution as they are sensitive to sample size and small differences may be found to be significant with increasing sample sizes and have been shown to reject good fitting models. The GFI index is analogous to $R^2$ and larger values represent better fit. Based on the recommendations of Hu and Bentler, TLI and CFI values greater than .95 were considered relatively good fit and values greater than .90 were considered adequate fit; RMSEA values less than .06 were also desired and less than .08 is considered reasonable. We also examined Akaike information...
criterion (AIC) values. Though no specific guidelines for significant differences in AIC values have been agreed upon, differences in AIC values have been used to decide among competing models. First, an EFA with maximum likelihood estimation was conducted on the first sub-sample. Second, the CFA was conducted on the other sub-sample. Finally, using the factor structure revealed by EFA/CFA analyses, an MGCFA was conducted on the full sample to test sample invariance between Black (n = 416) and White (n = 267) participants sampled. This MGCFA was conducted using a series of tests of measurement invariance that increase in stringency guided by the recommendations of Muthén and Christofferson. These include examining the baseline model with the fewest possible parameter constraints (e.g., factor loadings, variances, and covariances freely estimated), a second model (moderately strict) placing more constraints on parameters (i.e., factor loadings and intercepts), and a final model (strictest) setting all parameters (i.e., including error terms) equal across groups. Models were compared on aforementioned fit indices with chi-square difference tests conducted to compare fit indices across the Black and White groups, but interpreted with caution as sample size may yield rejection of a well-fitting model.

RESULTS

Preliminary Data Screening

As we would expect, the data are not normally distributed. It is widely accepted in the psychological sciences to treat ordinal data as continuous. Likert-type data in a depression screening tool will not be normally distributed and certain items, such as assessment of suicide, will not follow a normal distribution or be a good predictor of moderate to low levels of depression. Therefore, we accept that the data will not be normally distributed and will utilize the exception to normal distribution just as many others have who examine the factor analysis of depression screening tools. The
proportion of missing data from completed surveys was less than 0.1% for all items. Accordingly, all measurement and structural models were assessed using all cases.

Descriptive Statistics

Descriptive statistics appear in Table 1. Participants reported relatively low levels of all PHQ-A items relative to the response set options across both the Black and White sub-samples. Bivariate correlations among all PHQ-A items were in theoretically expected directions and of small to moderate magnitude for both the Black and White sub-samples.22

Exploratory and Confirmatory Factor Analysis

Previous research has assumed or demonstrated that the PHQ-9 loads on one factor (depression)4,13–15 while others provide evidence for 2 factors (somatic and affect).20,23,24 Therefore, we conducted exploratory factor analysis with maximum likelihood estimation and varimax rotation through AMOS on the 9 items from the PHQ-A to confirm the best model. Exploratory factor analysis results suggested a two-factor model to exhibit the best fit to the data. 3 items loaded on factor one (i.e, affect). These included items 1, 6, and 9. 5 items loaded on factor two (i.e., somatic). These included items 3, 4, 5, 7, and 8. Item 2 was complex as it loaded equivalently on both factors. (Table 2) Accordingly, a two-factor structure, with item 2 allowed to load on both affect and somatic factors, was further tested via follow-up confirmatory factor analysis.

To confirm the two-factor model was an adequate fit for the data, a confirmatory factor analysis was then performed with maximum likelihood estimation. A first-order, two-factor model was specified and exhibited an acceptable fit to the data according to CFI and RMSEA fit indices ($\chi^2(25) = 63.89, p < .001$, CFI = .91, NFI = .87, TLI = .84, RMSEA = .07 (.05–.09)). Item 2 was allowed to load on both affect (estimate = .14, p = .40) and somatic factors (estimate = .59, p = < .05), but it only loaded significantly on the somatic factor. Removing the Item 2 path to the affect factor did not include model fit; thus, it was
retained in the final model. All items exhibited moderate to high squared multiple correlation values ($R^2 = .10–.65$), suggesting that these items resonated relatively well with participants. See Table 3 for full EFA/CFA results.

**Multi-Group Confirmatory Factor Analysis**

All participants were included in the MGCFA analysis which tested a two factor (i.e., affect, somatic) model allowing Item 2 to load on both factors. Three hierarchic levels of measurement invariance were tested. In the baseline model, all factor loadings and thresholds were freely estimated across Black and White participants. In the baseline model, factor loadings were similar across models with the exception of Item 2 not significantly loading on the affect factor in the Black participant model only. However, fit indices comparison indicated that the two-factor structure of the PHQ-A was not fully invariant across the Black and White samples $\chi^2 = 47.97, p < .001, \Delta NFI = .029, \Delta TLI = .019, \text{RMSEA} = .06, AIC = 295.56$ in this baseline model. In the moderately strict model, factor loadings were similar across models. In the moderately strict model, all factor loadings and intercepts were constrained to be equal across Black and White participants. Again, fit indices comparison indicated that the two-factor structure of the PHQ-A was not fully invariant across the Black White samples $\chi^2 = 31.19, p < .001, \Delta NFI = .019, \Delta TLI = -.004, \text{RMSEA} = .07, AIC = 340.72$ in this moderate model. Finally, in the strictest testing model, all parameters were constrained to be equal across groups. Factor loadings were again similar across models with item 2 not loading significantly onto the affect factor for either group. Fit indices comparison again indicated that the two-factor structure of the PHQ-A was not fully invariant the Black and White samples $\chi^2 = 162.15, p < .099, \Delta NFI = .099, \Delta TLI = .093, \text{RMSEA} = .08, AIC = 496.91$ in this fully invariant model. Overall, MGCFA model testing suggests a two factor PHQ-A structure allowing Item 2 to load on both factors is not fully invariant for the adolescents sampled across Black ($n = 416$) and White ($n = 267$) participant groups, implying that the model functions differently between Blacks and Whites sampled. See Table 4 for full MGCFA results.
DISCUSSION

As the incidence of depression increases and sports medicine providers aim to utilize population-appropriate screening tools for depressive symptoms, it is important that we understand the structure and function of the PHQ-A in adolescents. Since previous research has demonstrated both one- and two-factor loading on the PHQ-A, we explored both possibilities. Our results indicate that a two-factor model, including “affect” and “somatic”, is an adequate fit for the data. Adolescent athletes appear to express their depressive symptoms through both factors. Overall, all items resonated well with participants. This aligns with other researchers who demonstrate the best fit of the data on a two-factor structure (somatic and affective).23–25

Considering the logistical constraints often present at the PPE, the PHQ-A offers practical advantages for athletic trainers and others tasked with screening for depression in athletes. It’s brevity, well-researched design, and construct based upon the DSM-IV criteria for depression makes it a valuable tool for the sports medicine setting. Further, it was designed for people ages 12-17, an age range of many athletes. These advantages make the PHQ-A a useful depression screening tool as mental health emerges as an equal component in overall health. Depression is diagnosed using both emotional and physical symptoms in adolescent and adult patients. Using the PHQ-A on this patient population will be useful in assessing both somatic and affective factors, as demonstrated by our analysis. This is consistent with the psychiatric framework of diagnostic criteria for depression, which is identified by affective disturbance, and supported by cognitive and somatic indicators. Clinicians should feel confident that the PHQ-A is assessing both components in the adolescent population.

Within the two-factor model, 3 items loaded on “affect” and 5 on “somatic.” One item, “Having little interest or pleasure in doing things,” was complex and loaded equivalently on both somatic and affective factors. This finding deserves further investigation. The inability to feel pleasure, or
anhedonia, is a complex, poorly understood, core symptom of depression\textsuperscript{26}. Some researchers have found the item to load heavily on the affective factor\textsuperscript{20,27}, but little is known about the specific differences between Black and White adolescent athletes and anhedonia. Treadway and Zald\textsuperscript{26} point out that many have hypothesized on the role of dopamine, but empirical evidence is elusive. The authors argue that anhedonia has not been adequately specified and further investigation into consummatory and motivational components of reward behavior would assist us in deeper understanding. Although it is beyond the scope of this paper, further investigation into adolescent athlete’s interpretation and experience with the item “little interest or pleasure in doing things” is warranted, particularly differences between races.

United States high school athletes represent a broad racial demographic, therefore a robust instrument that demonstrates validity across racial groups is required. We felt it was important to examine Blacks as a group considering previous literature indicating that racial differences may exist in expression of depressive symptoms. The literature is clear that depression assessment tools are not appropriate across races\textsuperscript{28} and do not account for the complex interrelatedness of various demographics including gender, socio-economic status and race.\textsuperscript{29,30} Black males, in particular, are well represented in high school athletics and more likely than White males to participate in the five most common high school sports.\textsuperscript{31} Nevertheless, conceptualization of depression among Black adolescents is shown to vary from other populations previously studied.\textsuperscript{6} Further, diagnosis and treatment are shown to be inequitable. Racial differences such as exposure to perceived daily stress, financial stress, neighborhood stress, and racial discrimination stress have been shown to increase the risk of depressive symptoms and led to a linear relationship between the accumulation of stressors and risk for depressive symptoms in Black teens as they emerge into adulthood.\textsuperscript{32}

Although our data demonstrate a two-factor structure, further analysis indicated that the two-factor structure was not invariant between Blacks and Whites sampled. The chi-square statistics were
significant across models, indicating the models were not invariant between groups. Additionally, there were modest differences in fit indices across models/groups. Specifically, the change in the NFI and CFI were greater than .01 in the fully invariant model, indicating noninvariance. The RMSEA is not at an acceptable level in any model. Therefore, race appears to be a source of heterogeneity in the factorial structure of the PHQ-A measurement model. The PHQ-A elicits different response patterns between Black and White adolescent athletes.

Looking more closely, we find the item “Having little interest or pleasure in doing things,” loads significantly differently between the racial groups in the baseline model. This item may drive some of the variability between race groups, along with other contributors. It has been accepted in clinical literature that expression of depression may be expressed as more somatic with Blacks and affective with Whites.\(^6\) Screening symptoms, particularly for purposes of meeting diagnostic criteria, tend to favor the affective over the somatic, therefore inflicting potential bias.\(^33\) For example, Lu et al. examined differences in Black adolescents from non-sport populations in their conceptualization of depression via the Center for Epidemiologic Studies Depression Scale (CES-D). The authors describe Black adolescents as more likely to express their depression symptoms as physical discomfort and suggest that clinicians should further consider the unique expression of depression among Black adolescents.\(^6\) Our findings demonstrate that 5 items load on somatic, versus 3 on affect. If Blacks are more likely to experience and express their depression somatically, then there would be more opportunity to do so on the PHQ-9. Specifically, continued PHQ-A measure development and evaluation in Black adolescent athlete populations represents a research line with important implications for the diagnosis and treatment in the increasingly diverse adolescent athlete population. The development of an instrument that gives opportunity for expression of both factors equally would allow for a better comparison of mean scores of the tool.
Clinicians can feel confident that the PHQ-A is measuring both affective and somatic component of depression. The differences between races deserve further investigation. The location of the parameters that differ across groups needs further investigation. Our data demonstrate a difference between Blacks and Whites regarding “little interest or pleasure in doing things.” When utilizing the PHQ-A, clinicians should consider that although they are assessing both somatic and affective parameters, Black and White adolescent athletes experience a construct bias on the PHQ-A. Further investigation into the perceptions and expression similarities and differences between adolescent athletes of differing races is needed in order to continue to improve the quality of assessment tools. Until then, direct comparison of total scores between races may be misleading.

Limitations

There are several limitations to our investigation. First, there are not equivalent numbers of Black and White athletes in our sample, potentially contributing to some differences seen across models. Our sample is also specific to an urban population in the southeastern United States and is limited in generalizability to the larger adolescent athlete population, particularly those living in a rural setting. The mass PPE can be an impersonal experience therefore participants were likely unfamiliar with those administering the PHQ-A and may have been resistant to sharing personal information and feelings. Specifically, the participants may have perceived that reporting symptoms of depression would jeopardize their ability to be permitted to participate in their sport, which could result in response bias and skewed outcomes. Athletes may also have been concerned about the information being directed back to stakeholders in authority such their coach, athletic trainer, or team physician, potentially affecting their ability to play or resulting in perceived negative consequences from those authoritative stakeholders. Finally, the PPE event necessitated that athletes procure their own transportation to and from the school in order to receive free bus transport to the PPE event, which was held on a weekend. It is possible that some student athletes were not able or willing to find such transportation, and thus...
were not included in the study. For that reason, our study may have been inadvertently biased to exclude some individuals from a low socioeconomic background.

**Conclusion**

Capitalizing on the opportunity that exists through the mandatory sports PPE, increased mental health screening can be achieved. The results of this investigation provide important information for clinicians and researchers as to the lack of full structural invariance between Black and White adolescent athletes on the PHQ-A. Clinicians must inquire regarding both somatic and affective components when assessing depressive symptoms in adolescent athletes. The PHQ-A appears to serve as a multidimensional assessment of depression for our participants as a whole but contains construct bias between Black and White adolescent athletes. A prospective analysis controlling for confounding variables is needed. These findings add to the current body of knowledge regarding depressive symptoms and construct validity, informing efforts to establish racially sensitive mental health screening and care.

**Implications and Contributions:** The PHQ-A measures both affective and somatic components of depression for adolescent athletes, but Black and White participants may interpret questions differently and total score comparison between groups may not be appropriate. Racially sensitive depression screening tools for adolescents deserve further investigation.
References

1. Global, regional, and national incidence, prevalence, and years lived with disability for diseases and injuries for 195 countries and territories, 1990-2017: A systematic analysis for the Global Burden of Disease Study 2017. *Glob Heal Metrics*. 2018;392(10159):1789-1858.

2. di Fronso S, Costa S, Montesano C, et al. The effects of COVID-19 pandemic on perceived stress and psychobiosocial states in Italian athletes. *Int J Sport Exerc Physiol*. 2020. doi:10.1080/1612197X.2020.1802612

3. Zuckerbrot R, Cheung A, Jensen P, Stein R. Guidelines for Adolescent Depression in Primary Care (GLAD-PC): Part I. Practice Preparation, Identification, Assessment, and Initial Management. *Pediatrics*. 2018;141(3).

4. Kroenke K, Spitzer RL, Williams JBW. The PHQ-9: Validity of a brief depression severity measure. *J Gen Intern Med*. 2001;16(9):606-613. doi:10.1046/j.1525-1497.2001.016009606.x

5. Johnson J, Harris E, Spitzer R, Williams J. The patient health questionnaire for adolescents. *J Adolesc Heal*. 2002;30(3):196-204.

6. Lu W, Lindsey M, Irsheid S, Nebbit VE. Psychometric Properties of the CES-D Among Black Adolescents in Public Housing. *J Soc Social Work Res*. 2017;8(4).

7. Tabachnick B, Fidell L. *Using Multivariate Statistics*. 6th ed. New York: Pearson; 2012.

8. IBM Corp. IBM SPSS Statistics for Windows, Version 25.0.

9. Arbuckle J. Amos (Version 23.0). 2014.

10. Kline R. *Principles and Practice of Structural Equation Modeling*. 2nd ed. New York: Guilford Press; 2005.
11. Keum B, Miller M, Inkelas K. Testing the Factor Structure and Measurement Invariance of the PHQ-9 across Racially Diverse US College Students. *Psychol Assess*. 2018;30(8):1096-1106.

12. Cheung G, Rensvold R. Evaluating goodness-of-fit indexes for testing measurement invariance. *Struct Equ Model*. 1999;9(2):233-255.

13. Hu L-T, Bentler P. Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Struct Equ Model*. 1999;6:11-55.

14. Crane P, Gibbons L, Willig J, et al. Measuring depression levels in HIV-infected patients as part of routine clinical care using the nine-item patient health questionnaire (PHQ-9). *AIDS Care*. 2010;22(7):874-885.

15. Baas K, Cramer A, Koeter M, van de Lisdonk E, van Weert H, Schene A. Measurement invariance with respect to ethnicity of the patient health questionnaire-9 (PHQ-9). *J Affect Disord*. 2011;129(1):229-235.

16. Muthen B, Christofferson A. Simultaneous factor analysis of dichotomous variables in several groups. *Psychometrika*. 1981;46:407-419.

17. Norman G. Likert scales, levels of measurement and the “laws” of statistics. *Adv Heal Sci Educ*. 2010;15:625-632.

18. Harry M, Coley RY, Waring SC, Simon GE. Evaluating the cross-cultural measurement invariance of the PHQ-9 between American Indian/Alaska Native adults and diverse racial and ethnic groups. *J Affect Disord Reports*. 2021;4.

19. Assari S, Moazen-Zadeh E. Confirmatory Factor Analysis of the 12-Item Center for Epidemiologic Studies Depression Scale among Blacks and Whites. *Front Psychiatry*. 2016;7:178.
20. Elhai J, Contractor A, Tamburrino M, et al. The factor structure of major depression symptoms: a test of four competing models using the patient health Questionnaire-9. *Psychiatry Res.* 2012;199(3):169-173.

21. Borgogna N, Brenner R, McDermott R. Sexuality and gender invariance of the PHQ-9 and GAD-7: Implications for 16 identity groups. *J Affect Disord.* 2021;278:122-130.

22. Cohen J. *Statistical Power Analysis for the Behavioral Sciences.* 2nd ed. Hillsdale, NJ: Lawrence Erlbaum Associates; 1988.

23. Granillo MT. Structure and Function of the Patient Health Questionnaire-9 Among Latina and Non-Latina White Female College Students. 2012;3(2):80-93. doi:10.5243/jsswr.2012.6

24. Chilcot J, Rayner L, Lee W, et al. The factor structure of the PHQ-9 in palliative care. *J Psychosom Res.* 2013;75(1):60-64.

25. Guo B, Kaylor-Hughes C, Garland A, et al. Factor structure and longitudinal measurement invariance of PHQ-9 for specialist mental health care patients with persistent major depressive disorder: Exploratory Structural Equation Modelling. *J Affect Disord.* 2017;219:1-8.

26. Treadway M, Zald D. Reconsidering anhedonia in depression: lessons from translational neuroscience. *Neurosci Biobehav Rev.* 2011;35(3):537-555.

27. Doi S, Ito M, Takebayashi Y, Muramatsu K, Horikoshi M. Factorial validity and invariance of the Patient Health Questionnaire (PHQ)-9 among clinical and non-clinical populations. *PLoS One.* 2018;13(7).

28. Ofonedu M, Percy W, Harris-Britt A, Belcher H. Depression in inner city African American youth: A phenomenological study. *J Child Fam Stud.* 2013;22(1):96-106.
29. Chang T, Weiss A, Luana M, et al. Race/Ethnicity and Other Social Determinants of Psychological Well-being and Functioning in Mental Health Clinics. *J Heal Care Poor Underserved*. 2017;25(3):1418-1431.

30. Sohn H. Racial and Ethnic Disparities in Health Insurance Coverage: Dynamics of Gaining and Losing Coverage over the Life-Course. *Popul Res Policy Rev*. 2017;36(2):181-201. doi:10.1007/s11113-016-9416-y.

31. Turner R, Perrin E, Coyne-Beasley T, Peterson C, Skinner A. Reported Sports Participation, Race, Sex, Ethnicity, and Obesity in the US Adolescents from NHANES Physical Activity (PAQ_D). *Glob Pediatr Heal*. 2015. doi:10.1177/2333794414563799.

32. Estrada-Martinez L, Caldwell C, Bauermeister J, Zimmerman M. Stressors in multiple life-domains and the risk for externalizing and internalizing behaviors among African Americans during emerging adulthood. *J Youth Adolesc*. 2012;41(12):1600-1612.

33. Kerr LK, Kerr LD. Screening tools for depression in primary care. *West J Med*. 2001;175(5):349-352.
Table 1.

Demographic Information by Randomly Selected Sub-Sample

|                     | Sub-Sample 1 | Sub-Sample 2 |
|---------------------|--------------|--------------|
| **Gender**          |              |              |
| Male                | 70%          | 70%          |
| Female              | 30%          | 30%          |
| **Ethnicity**       |              |              |
| Black               | 59%          | 62%          |
| White               | 41%          | 38%          |
| **PHQ-A Classification** |          |              |
| No or Minimal Depression | 84%  | 81%  |
| Mild Depression     | 11%          | 17%          |
| Moderate Depression | 2%           | 2%           |
| Moderately Severe Depression | 2%  | 1%  |
| Severe Depression   | 1%           | 0%           |

Downloaded from http://meridian.allenpress.com/jat/article-pdf/doi/10.4085/1062-6050-343-21/2982846/10.4085_1062-6050-343-21.pdf by guest on 15 December 2021
Table 2.

Means, Standard Deviations, and Item Correlations for the PHQ-A Among Blacks (n = 416) and Whites (n = 267)

| Item                                                                 | BL M | BL SD | WH M | WH SD | 1   | 2   | 3   | 4   | 5   | 6   | 7   | 8   | 9   |
|----------------------------------------------------------------------|------|-------|------|-------|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| 1. Feeling down, depressed, irritable, or hopeless                   | 0.15 | 0.43  | 0.11 | 0.38  | -   | 0.48| 0.47 | 0.46 | 0.41 | 0.69 | 0.24 | 0.35 | 0.34 |
| 2. Little interest or pleasure in doing things                       | 0.38 | 0.75  | 0.16 | 0.48  | 0.21 | -   | 0.28 | 0.35 | 0.28 | 0.38 | 0.25 | 0.21 | 0.20 |
| 3. Trouble falling asleep, staying asleep or sleeping too much       | 0.52 | 0.82  | 0.39 | 0.69  | 0.32 | 0.19| -   | 0.49 | 0.36 | 0.29 | 0.18 | 0.22 | 0.27 |
| 4. Poor appetite, weight loss or overeating                         | 0.50 | 0.76  | 0.42 | 0.65  | 0.34 | 0.35 | 0.51 | -   | 0.25 | 0.26 | 0.33 | 0.23 | 0.21 |
| 5. Feeling tired, or having little energy                            | 0.31 | 0.70  | 0.18 | 0.57  | 0.33 | 0.18 | 0.40 | 0.31 | -   | 0.35 | 0.39 | 0.40 | 0.12 |
| 6. Feeling bad about yourself, that you are a failure, or that you have let yourself or your family down | 0.18 | 0.55  | 0.15 | 0.48  | 0.64 | 0.19 | 0.28 | 0.32 | 0.28 | -   | 0.20 | 0.30 | 0.32 |
| 7. Trouble concentrating on things like schoolwork, reading or watching tv | 0.28 | 0.63  | 0.24 | 0.56  | 0.29 | 0.24 | 0.41 | 0.27 | 0.21 | 0.31 | -   | 0.42 | *0.06 |
| 8. moving or speaking too slowly or being fidgety or restless        | 0.13 | 0.42  | 0.13 | 0.46  | 0.27 | 0.21 | 0.38 | 0.23 | 0.19 | 0.26 | 0.40 | -   | 0.28 |
| 9. Thoughts that you would be better off dead or hurting yourself     | 0.05 | 0.23  | 0.01 | 0.14  | 0.53 | 0.19 | 0.27 | 0.24 | 0.24 | 0.45 | 0.16 | 0.16 | -   |

Notes. Correlations for the Black sample are below matrix diagonal, whereas White correlations are above the matrix diagonal. All values significant at $p < .01$ unless noted; BL = Black sample; WH = White sample.
Table 3.
**Exploratory and Confirmatory Factor Analysis Results by Sub-Sample**

| Item                        | Exploratory Factor Analysis (n = 333) | Confirmatory Factor Analysis (n = 350) |
|-----------------------------|---------------------------------------|---------------------------------------|
|                             | Factor 1 | Factor 2 | Factor 1 | Factor 2 |
|                             | Affect   | Somatic  | Affect   | Somatic  |
| 1. Down/depressed           | 0.785    | 0.427    | 1.00     | -        |
| 2. Little interest/pleasure | 0.267    | 0.346    | 0.14     | 0.59     |
| 3. Sleep trouble            | 0.320    | 0.643    | -        | 1.00     |
| 4. Appetite change          | 0.286    | 0.638    | -        | 0.89     |
| 5. Tired/little energy      | 0.348    | 0.517    | -        | 0.54     |
| 6. Feeling bad about self   | 0.829    | 0.255    | 1.16     | -        |
| 7. Trouble concentrating   | 0.165    | 0.643    | -        | 0.57     |
| 8. Slow or fast movement/speech | 0.238 | 0.561    | -        | 0.27     |
| 9. Thoughts of death        | 0.578    | 0.254    | 0.23     | -        |

**Fit Indices**

|                     | Exploratory Factor Analysis | Confirmatory Factor Analysis |
|---------------------|-----------------------------|-----------------------------|
| $\chi^2$            | 54.69                       | 63.89                       |
| CFI                 | -                           | 0.91                        |
| NFI                 | -                           | 0.87                        |
| TLI                 | -                           | 0.84                        |
| RMSEA               | -                           | 0.07                        |
| AIC                 | -                           | 121.89                      |

*Note. All values significant at $p < .05$ unless noted *; CFI = comparative fit index, NFI = normed fit index; TLI = Tucker-Lewis index, RMSEA = root mean square error of approximation, AIC = Akaike information criterion.*
### Table 4.

**Multi-Group Confirmatory Factor Analysis (N = 683)**

| Factor Item | Baseline Model | Moderate Model | Fully Invariant Model |
|-------------|----------------|----------------|-----------------------|
|             | BL (n=416)     | WH (n=267)     | BL (n=416)            | WH (n=267)            | BL (n=416) | WH (n=267) |
| **Affect**  |                |                |                       |                      |            |            |
| 1           | 1.00           | 1.00           | 1.00                  | 1.00                 | 1.00       | 1.00       |
| 2           | 0.08*          | 0.46           | 0.23                  | 0.23                 | 0.15*      | 0.15*      |
| 6           | 1.11           | 0.98           | 1.05                  | 1.05                 | 1.08       | 1.08       |
| 9           | 0.39           | 0.14           | 0.26                  | 0.26                 | 0.30       | 0.30       |
| **Somatic** |                |                |                       |                      |            |            |
| 2           | 0.45           | 0.26           | 0.42                  | 0.42                 | 0.46       | 0.46       |
| 3           | 1.00           | 1.00           | 1.00                  | 1.00                 | 1.00       | 1.00       |
| 4           | 0.81           | 0.95           | 0.85                  | 0.85                 | 0.87       | 0.87       |
| 5           | 0.58           | 0.80           | 0.65                  | 0.65                 | 0.66       | 0.66       |
| 7           | 0.58           | 0.64           | 0.58                  | 0.58                 | 0.60       | 0.60       |
| 8           | 0.35           | 0.56           | 0.39                  | 0.39                 | 0.40       | 0.40       |

**Fit Indices Comparisons**

| Fit Indices | Baseline Model | Moderate Model | Fully Invariant Model |
|-------------|----------------|----------------|-----------------------|
| $\chi^2$    | 47.97          | 31.19          | 162.15                |
| $\Delta$ NFI| 0.029          | 0.019          | 0.099                 |
| $\Delta$ TLI| 0.019          | -0.004         | 0.093                 |
| RMSEA       | .06            | .07            | .08                   |
| AIC         | 295.56         | 340.72         | 496.91                |

**Note.** All values significant at $p < .05$ unless noted by *; BL = Black sample; WH = White sample; CFI = comparative fit index, NFI = normed fit index; TLI = Tucker-Lewis index, RMSEA = root mean square error of approximation, AIC = Akaike information criterion.