

# Factorial Validity and Invariance of an Adolescent Depression Symptom Screening Tool

Ashley Long, PhD\*; J.D. DeFreese, PhD†; Allison Bickett, PhD‡; David Price, MD‡

\*Mount Olive Family Medicine Center, NC; †Center for the Study of Retired Athletes, University of North Carolina at Chapel Hill; ‡Department of Family Medicine, Atrium Health, Charlotte, NC

**Context:** Depression is among the most common mental health disorders in youth, results in significant impairment, and is associated with a higher risk of suicide. Screening is essential, but assessment tools may not account for the complex interrelatedness of various demographic factors, such as sex, socioeconomic status, and race.

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

**Design:** Retrospective cohort design.

**Setting:** Data were obtained from a secure database collected at a free, comprehensive, mass preparticipation physical examination event hosted by a large health care system.

**Patients or Other Participants:** Participants were 683 high school athletes (Black = 416, White = 267). The independent variables were somatic and affective factors contributing to the construct of depression measured by the PHQ-A and participant race (Black or White).

**Main Outcome Measure(s):** (1) Factors upon which the construct of depression is measured and (2) measurement invariance between Black and White participants.

**Results:** A 2-factor model, involving affective and somatic components, was specified and exhibited adequate fit to the data (comparative fit index >0.90). All items exhibited moderate to high squared multiple correlation values ( $R^2 = 0.10$ – $0.65$ ), suggesting that these items resonated relatively well with participants. The 2-factor model demonstrated noninvariance between Black and White participants (root mean square error of approximation =  $0.06$ – $0.08$ ).

**Conclusions:** Overall, the structure of the PHQ-A was supported by a 2-factor model in adolescent athletes, measuring both affective and somatic symptoms of depression. However, a 2-factor PHQ-A structure was not fully invariant for the adolescents sampled across participant groups, indicating that the model functioned differently between the Black and White participants sampled.

**Key Words:** adolescent athletes, mental health screening, psychology

## Key Points

- The Patient Health Questionnaire-Adolescent 2-factor model of somatic and affective symptoms was supported in a group of adolescent athletes.
- The Questionnaire elicited different response patterns between Black and White adolescent athletes.

Depression is the most common mental illness and a leading cause of disability worldwide. It is a major contributor to the overall global burden of disease<sup>1</sup> and often emerges during adolescence. During the recent COVID-19 pandemic, athletes have demonstrated increases in perceived stress and dysfunctional psychobiosocial states,<sup>2</sup> and the full effects are likely not yet realized. The American Academy of Pediatrics recommended an annual universal screening for depression in youth  $\geq 12$  years at health maintenance visits,<sup>3</sup> as early screening and treatment are vital to addressing the significant health consequences and long-term health care costs. Although screening is recommended, many adolescent patients do not complete annual wellness visits and seek medical care only when they are sick. An opportunity exists to engage healthy adolescents in mental health screening through the sport preparticipation examinations (PPEs) that are mandatory in 50 states.<sup>3</sup> Many PPE forms for American high school athletes include some form of mental health screening.

The American Medical Society for Sports Medicine suggested that early recognition provides an opportunity for preventing the more serious effects of depression.<sup>4</sup> The Patient Health Questionnaire-Adolescent (PHQ-A) is a broadly used, well-validated depression screening tool,<sup>5,6</sup> and yet, little information exists 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 PPEs. Conceptualizations of depression among Black adolescents may differ from those among their White counterparts.<sup>7</sup> Without an understanding of the function of the PHQ-A in a racially diverse adolescent athlete population, we risk inappropriate interpretations and conclusions from its results.

The standard of care during the adolescent sports PPE now includes an assessment of mental health (eg, 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 among racial groups. We aimed to examine the (1) the factor structure of the measure of depression on the PHQ-A in a group of adolescent athletes and (2) measurement invariance between Black and White patients on the PHQ-A.

## METHODS

### Setting

We conducted a retrospective review of PPE screening. Adolescent athletes were allowed to register for the PPE online 5 weeks before the event. Parents or guardians of students  $\leq 17$  years of age or students themselves if  $\geq 18$  years were instructed to provide the PPE medical history and demographic information of the participant. The online form is populated in 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. The PPE activities were performed in a specific order for all participants, including an assessment of vital signs, vision test, review of the athlete's medical history and depression screening (PHQ-A), sport-specific medical and musculoskeletal examination, electrocardiogram, and clerical check out. Sports medicine physicians, cardiologists, registered dietitians, 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. Each participant completed the form electronically and individually by reading and responding to the items on a laptop computer with a trained administrator available for questions. Health information obtained during the PPE was uploaded via secure servers to a secure platform, was password protected, and was accessible only to those who met specific criteria, including the health care system's employed athletic trainers, physicians, and privilege-granted researchers. The platform was compliant with federal privacy, storage, and transmission standards.

### Participants

A *participant* was defined as any individual engaged in a school-sanctioned high school sport who attended the PPE event in person. Students participating in the event attended a public, charter, or private school that had a contract for athletic training services with the health care system. A total of 897 participants attended the event, of whom 683 met the inclusion criteria for the current study based on their self-identified race or ethnicity. The ages of the athletes ranged from 13 to 18 years (mean =  $15.6 \pm 0.99$  years). A total of 480 (70.3%) of participants were male, and 203 (29.7%) were female. Because we aimed to examine the construct regarding Black and White students, 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 female adolescents ( $n = 125$ , 18.3%), Black male adolescents ( $n$

= 291, 42.6%), White female adolescents ( $n = 78$ , 11.4%), and White male adolescents ( $n = 189$ , 27.7%). Female adolescents accounted for 30% of the Black participants and 29.2% of the White participants, making the Black and White groups similar in the percentage of each sex. Participants provided their own transportation to their school on the morning of the event and then were bussed to the site of the PPE. This study was approved by our institution's institutional review board 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 et al<sup>6</sup> in 2008 to screen depressive symptoms among adolescents. The PHQ-A can be used to assess the core criteria of depression according to the *Diagnostic and Statistical Manual of Mental Disorders*, fourth edition,<sup>8</sup> 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. Each question is scored on a scale of 0 (*not at all*) to 3 (*nearly every day*). The scores are totaled and can range from 0 to 27. Depressive symptoms are categorized as 0 to 4, *negligible*; 5 to 9, *mild*; 10 to 14, *moderate*; 15 to 19, *moderately severe*; or 20 to 27, *severe*. Scores of  $\geq 10$  are often designated as meeting criteria for depression and warrant the individual's referral to a mental health provider. Any indication of suicidal thoughts or plans constitutes an emergency, and an intervention plan should be enacted immediately. The PHQ-A demonstrated excellent internal reliability (Cronbach  $\alpha = 0.89$ ) and satisfactory specificity (84%–95%), sensitivity (68%–95%), and likelihood ratio (6.0–13.6).<sup>5,6</sup> Demographic information was obtained from the electronic PPE form.

### Data Analysis

Before performing structural analyses, we analyzed the data for missing values, outliers, and violations of multivariate assumptions.<sup>9</sup> The 3 main data-analytic approaches used to examine the structure and invariance of the PHQ-A were exploratory factor analysis (EFA), confirmatory factor analysis (CFA), and multigroup confirmatory factor analysis (MGCFAs). To ensure that EFA and CFA were conducted with unique participants, 2 random subsamples of participants were created using a random-split command. Group difference analysis (ie,  $\chi^2$ ) was performed to ensure the subsamples did not differ in sex, race, or PHQ-A diagnostic classification (Table 1). All statistical analyses and EFA analyses were completed via SPSS (version 25; IBM Corp),<sup>10</sup> whereas all CFA and MGCFAs were completed via the AMOS software extension to SPSS (version 26).<sup>11</sup>

Consistent with best practice, we calculated factor analyses using maximum likelihood estimation and fixing 1 loading value to 1.0 for each latent variable. In accord with best practice,<sup>12</sup> we applied multiple fit indices, path coefficients, and modification indices to assess and compare model fit. Specifically, the following fit indices were used:  $\chi^2$  goodness of fit, goodness-of-fit index, Tucker-Lewis

**Table 1. Demographic Information by Randomly Selected Subsample**

Characteristic	Subsample, % (No.)	
	1 (n = 333)	2 (n = 350)
Sex		
Male	70 (234)	70 (246)
Female	30 (99)	30 (104)
Race		
Black	59 (198)	62 (218)
White	41 (135)	38 (132)
Patient Health Questionnaire-Adolescent depression classification		
None or minimal	84 (281)	81 (283)
Mild	11 (37)	17 (58)
Moderate	2 (8)	2 (8)
Moderately severe	2 (5)	1 (1)
Severe	1 (2)	0 (0)

index (TLI), comparative fit index (CFI), and root mean square error of approximation (RMSEA). We interpreted  $\chi^2$  significance cautiously, as it is sensitive to sample size; small differences may be significant with increasing sample sizes<sup>13</sup> and have been shown to reject good-fitting models.<sup>14</sup> The goodness-of-fit index is analogous to  $R^2$  and larger values represent better fit. Based on the recommendations of Hu and Bentler,<sup>15</sup> TLI and CFI values  $>0.95$  were considered relatively good fit and values  $>0.90$  were considered adequate fit; RMSEA values  $<0.06$  were also desired and  $<0.08$  is considered reasonable.<sup>16,17</sup> We also examined Akaike information criterion (AIC) values. Although no specific guidelines for significant differences in AIC values have been agreed upon, differences have been used to decide among competing models.<sup>12</sup>

First, an EFA with maximum likelihood estimation was conducted on the first subsample. Second, a CFA was conducted on the other subsample. Finally, using the factor structure revealed by EFAs and CFAs, we computed an MGCFA on the full sample to test sample invariance between Black ( $n = 416$ ) and White ( $n = 267$ ) participants. This MGCFA was determined using a series of tests of measurement invariance that increase in stringency guided by the recommendations of Muthén and Christofferson.<sup>18</sup> They include examining the baseline model with the fewest possible parameter constraints (eg, factor loadings, variances, and covariances freely estimated), a second model (moderately strict) placing more constraints on parameters (ie, factor loadings and intercepts), and a final model (strictest) setting all parameters (ie, including error terms) equal across groups. Models were compared on the aforementioned fit indices with  $\chi^2$  difference tests conducted to compare fit indices across the Black and White groups. However, they were interpreted with caution, as sample size may cause a well-fitting model to be rejected.

## RESULTS

### Preliminary Data Screening

As expected, the data were not normally distributed. It is widely accepted in the psychological sciences that ordinal data should be treated as continuous. Likert-type data in a depression screening tool are not normally distributed, and certain items, such as the assessment of suicide, will not follow a normal distribution or be good predictors of

moderate to low levels of depression.<sup>19</sup> Therefore, we accepted that the data were not normally distributed and used the exception to normal distribution, as have many others<sup>13,20–23</sup> who examined the factor analysis of depression screening tools. The proportion of missing data from completed surveys was  $<0.1\%$  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 subsamples. Bivariate correlations among all PHQ-A items were in theoretically expected directions and of small to moderate magnitude for both the Black and White subsamples.<sup>24</sup>

### Exploratory Analysis and CFA

Previous researchers assumed or demonstrated that the 9-item Patient Health Questionnaire (PHQ-9) loads on 1 factor (depression),<sup>5,15–17</sup> whereas others provided evidence for 2 factors (somatic and affect).<sup>22,25,26</sup> Thus, we conducted an EFA with maximum likelihood estimation and varimax rotation using AMOS software on the 9 items from the PHQ-A to confirm the best model. The EFA results suggested that a 2-factor model exhibited the best fit to the data. Three items loaded on factor 1 (ie, affect): items 1, 6, and 9. Five items loaded on factor 2 (ie, somatic): 3, 4, 5, 7, and 8. Item 2 was complex, as it loaded equivalently on both factors (Table 2). Accordingly, a 2-factor structure, with item 2 allowed to load on both the affect and somatic factors, was further tested via follow-up CFA.

To confirm that the 2-factor model was an adequate fit for the data, we then performed a CFA with maximum likelihood estimation. A first-order, 2-factor model was specified and exhibited an acceptable fit to the data according to the CFI and RMSEA fit indices ( $\chi^2_{25} = 63.89$ ,  $P < .001$ , TLI = 0.84, CFI = 0.91, normed fit index [NFI] = 0.87, RMSEA = 0.07 [0.05–0.09]). Item 2 was allowed to load on both affect (estimate = 0.14,  $P = .40$ ) and somatic (estimate = 0.59,  $P < .05$ ) factors, but it only loaded significantly on the somatic factor. Removing the item 2 path to the affect factor did not improve model fit; thus, it was retained in the final model. All items exhibited moderate to high squared multiple correlation values ( $R^2 = 0.10–0.65$ ), suggesting that these items resonated relatively well with participants. The full EFA and CFA results are provided in Table 3.

### Multigroup Confirmatory Factor Analysis

Data from all participants were included in the MGCFA that tested a 2-factor (ie, affect and somatic) model and allowed item 2 to load on both factors. Three hierarchic levels of measurement invariance were examined. In the baseline model, all factor loadings and thresholds were freely estimated across Black and White participants. Factor loadings were similar across models except that item 2 did not significantly load on the affect factor in the Black participant model. However, a fit indices comparison indicated that the 2-factor structure of the PHQ-A was not



**Table 2. Patient Health Questionnaire-Adolescent Means, Standard Deviations, and Item Correlations Among Black (n = 416) and White (n = 267) Adolescents<sup>a</sup>**

Item	Adolescents, Mean $\pm$ SD		Item									
	Black	White	1	2	3	4	5	6	7	8	9	
1. Feeling down, depressed, irritable, or hopeless	0.15 $\pm$ 0.43	0.11 $\pm$ 0.38	NA	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 $\pm$ 0.75	0.16 $\pm$ 0.48	0.21	NA	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 $\pm$ 0.82	0.39 $\pm$ 0.69	0.32	0.19	NA	0.49	0.36	0.29	0.18	0.22	0.27	
4. Poor appetite, weight loss, or overeating	0.50 $\pm$ 0.76	0.42 $\pm$ 0.65	0.34	0.35	0.51	NA	0.25	0.26	0.33	0.23	0.21	
5. Feeling tired or having little energy	0.31 $\pm$ 0.70	0.18 $\pm$ 0.57	0.33	0.18	0.40	0.31	NA	0.35	0.39	0.40	0.12 <sup>b</sup>	
6. Feeling bad about yourself, that you are a failure, or that you have let yourself or your family down	0.18 $\pm$ 0.55	0.15 $\pm$ 0.48	0.64	0.19	0.28	0.32	0.28	NA	0.20	0.30	0.32	
7. Trouble concentrating on things like schoolwork, reading, or watching television	0.28 $\pm$ 0.63	0.24 $\pm$ 0.56	0.29	0.24	0.41	0.27	0.21	0.31	NA	0.42	0.06 <sup>b</sup>	
8. Moving or speaking too slowly or being fidgety or restless	0.13 $\pm$ 0.42	0.13 $\pm$ 0.46	0.27	0.21	0.38	0.23	0.19	0.26	0.40	NA	0.28	
9. Thoughts that you would be better off dead or hurting yourself	0.05 $\pm$ 0.23	0.01 $\pm$ 0.14	0.53	0.19	0.27	0.24	0.24	0.45	0.16	0.16	NA	

<sup>a</sup> Correlations for the Black sample are below the matrix diagonal, whereas those for the White sample are above the matrix diagonal.

<sup>b</sup> Indicates nonsignificant values. All other values were significant at  $P < .01$ .

fully invariant across the Black and White samples ( $\chi^2 = 47.97$ ,  $P < .001$ ,  $\Delta\text{NFI} = 0.029$ ,  $\Delta\text{TLI} = 0.019$ ,  $\text{RMSEA} = 0.06$ ,  $\text{AIC} = 295.56$ ) in this baseline model. In the moderately strict model, factor loadings were similar across models. All factor loadings and intercepts were constrained to be equal across Black and White participants. Again, a fit indices comparison indicated that the 2-factor structure of the PHQ-A was not fully invariant across the Black and White samples ( $\chi^2 = 31.19$ ,  $P < .001$ ,  $\Delta\text{TLI} = -0.004$ ,  $\Delta\text{NFI} = 0.019$ ,  $\text{RMSEA} = 0.07$ ,  $\text{AIC} = 340.72$ ) in this moderate model. Finally, in the strictest testing model, all parameters were constrained to be equal across groups. Factor loadings were similar across models, but item 2 did not significantly load on the affect factor for either group. The fit indices comparison again indicated that the 2-factor structure of the PHQ-A was not fully invariant across the Black and White samples ( $\chi^2 = 162.15$ ,  $P < .099$ ,  $\Delta\text{TLI} = 0.093$ ,  $\Delta\text{NFI} = 0.099$ ,  $\text{RMSEA} = 0.08$ ,  $\text{AIC} = 496.91$ ) in this fully invariant model. Overall, MGCFA model testing suggested that a 2-factor PHQ-A structure, allowing item 2 to load on both factors, was not fully invariant across the

Black (n = 416) and White (n = 267) participant groups, implying that the model functioned differently in the Black and White participants sampled. The full MGCFA results are shown in Table 4.

## DISCUSSION

As the incidence of depression increases and sports medicine providers aim to use population-appropriate screening tools for patients with depressive symptoms, it is important that we understand the structure and function of the PHQ-A in adolescents. Considering the logistical constraints often present at the PPE, the PHQ-A offers practical advantages for athletic trainers and others tasked with screening athletes for depression. Its brevity, well-researched design, and construct based upon the *DSM-4* criteria for depression<sup>8</sup> make it a valuable tool for the sports medicine setting. Furthermore, it was designed for people ages 12 to 17, which is the age range of many athletes. These advantages make the PHQ-A a useful depression

**Table 3. Exploratory and Confirmatory Factor Analysis Results by Subsample**

Item	Exploratory Factor Analysis (n = 333)		Confirmatory Factor Analysis (n = 350)	
	Factor 1 (Affect)	Factor 2 (Somatic)	Factor 1 (Affect)	Factor 2 (Somatic)
1. Down or depressed	0.785	0.427	1.00	
2. Little interest or pleasure	0.267	0.346	0.14 <sup>a</sup>	0.59
3. Sleep trouble	0.320	0.643		1.00
4. Appetite change	0.286	0.638		0.89
5. Tired and 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 and speech	0.238	0.561		0.27
9. Thoughts of death	0.578	0.254	0.23	
Fit indices				
$\chi^2$	54.69			63.89
Tucker-Lewis index				0.84
Comparative fit index				0.91
Normed fit index				0.87
Root mean square error of approximation				0.07
Akaike information criterion				121.89

<sup>a</sup> Indicates nonsignificant values. All other values were significant at  $P < .05$ .

**Table 4. Multigroup Confirmatory Factor Analysis (N = 683)**

Factor Item	Model (Black Adolescents = 416, White Adolescents = 267)					
	Baseline		Moderate		Fully Invariant	
	Black	White	Black	White	Black	White
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
Affect						
2	0.08 <sup>a</sup>	0.46	0.23	0.23	0.15 <sup>a</sup>	0.15 <sup>a</sup>
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
Fit indices comparisons						
$\chi^2$	47.97	$P < .001$	31.19	$P < .001$	162.15	$P < .001$
$\Delta$ Tucker-Lewis index	0.019		-0.004		0.093	
$\Delta$ Normed fit index	0.029		0.019		0.099	
Root mean square error of approximation	0.06		0.07		0.08	
Akaike information criterion	295.56		340.72		496.91	

<sup>a</sup> Indicates nonsignificant values. All other values were significant at  $P < .05$ .

screening tool as mental health emerges as a component equal to physical health in considering overall health.

Because previous researchers have demonstrated both 1- and 2-factor loading on the PHQ-A, we explored both possibilities. Our results indicated that a 2-factor model, including affect and somatic, was an adequate fit for the data. Adolescent athletes appeared to express their depressive symptoms through both factors. Overall, all items resonated well with participants. These findings align with those from other investigators<sup>25-27</sup> who demonstrated the best fit of the data on a 2-factor structure (somatic and affective). Depression is diagnosed based on both emotional and physical symptoms in adolescent and adult patients. Using the PHQ-A in the adolescent patient population will be beneficial for assessing both somatic and affective factors, as demonstrated by our analysis. Assessing both factors is consistent with the psychiatric framework of the 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.

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 race given previous indications that racial differences may exist in the expression of depressive symptoms. The literature is clear that depression assessment tools are not appropriate across races<sup>28</sup> and do not account for the complex interrelatedness of various demographics, including sex, socioeconomic status, and race.<sup>29,30</sup> Black male adolescents, in particular, are well represented in high school athletics and more likely than White male adolescents to participate in the 5 most common high school sports.<sup>31</sup> Nevertheless, the conceptualization of depression among Black adolescents has been shown to vary from other populations studied.<sup>7</sup> Furthermore, diagnosis and treatment were shown to be inequitable. Racial differences such as exposure to perceived daily stress, financial stress, neighborhood stress, and racial

discrimination stress increased the risk of depressive symptoms and led to a linear relationship between the accumulation of stressors and the risk for depressive symptoms in Black teens as they emerged into adulthood.<sup>32</sup>

Although our data demonstrated a 2-factor structure, a further analysis indicated that the 2-factor structure was not invariant between the Black and White participants sampled. The  $\chi^2$  statistics were significant across models, reflecting that the models were not invariant between groups. Additionally, we found modest differences in fit indices across models and groups. Specifically, the changes in the NFI and CFI were  $>0.01$  in the fully invariant model, indicating noninvariance. The RMSEA was not at an acceptable level in any model. Thus, 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.

In the 2-factor model, 3 items loaded on the affect and 5 on the somatic factor. One item, *little interest or pleasure in doing things*, was complex and loaded equivalently on both the somatic and affective factors. This finding deserves further exploration. The inability to feel pleasure, or *anhedonia*, is a complex, poorly understood core symptom of depression.<sup>33</sup> Some researchers<sup>22,34</sup> noted that the item loaded heavily on the affective factor, but little is known about the specific differences in anhedonia between Black and White adolescent athletes. Treadway and Zald<sup>33</sup> pointed out that many have hypothesized on the role of dopamine, but empirical evidence is elusive. The authors argued that anhedonia has not been adequately specified and that further investigation into the multiple components of reward behavior would assist us in a deeper understanding.

Looking more closely, we observed that the item *little interest or pleasure in doing things* loaded 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. The expression of depression may be expressed as more somatic in Black

adolescents and more affective in White adolescents.<sup>7</sup> Screening symptoms, particularly for the purposes of meeting diagnostic criteria, tend to favor the affective over the somatic, thereby inflicting potential bias.<sup>35</sup> For example, Lu et al<sup>7</sup> examined differences in Black adolescents from nonsport populations in their conceptualization of depression via the Center for Epidemiologic Studies Depression Scale. The authors described Black adolescents as more likely to express their depression symptoms as physical discomfort and suggested that clinicians should further consider the unique expression of depression among Black adolescents.<sup>7</sup> Our findings demonstrated that 5 items loaded on the somatic versus 3 on the affective factor. If Black adolescents are more likely to experience and express their depression somatically, the PHQ-9 offers more opportunity to do so. Specifically, continued development and evaluation of the PHQ-A measure in adolescent athlete populations of various racial groups 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 provides an opportunity for the 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 measures both somatic and affective components of depression. However, the differences between races deserve further investigation, as do the locations of the parameters that differ across groups. Our data demonstrated a difference between Black and White adolescents regarding *little interest or pleasure in doing things*. When using 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 evaluation of the perception and expression similarities and differences between adolescent athletes of different races is needed to continue to improve the quality of assessment tools. Until then, a direct comparison of total scores between races may be misleading.

## Limitations

Our work had several limitations. First, Black and White athletes were not represented equally in our sample, possibly contributing to some differences across models. Our sample was also specific to an urban population in the southeastern United States, which may limit 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 participate in their sport, which could have resulted in response bias and skewed outcomes. Athletes may also have been concerned about the information being shared with stakeholders in authority, such as their coach, athletic trainer, or team physician, potentially affecting their ability to play or resulting in perceived negative consequences from those stakeholders. Finally, the PPE event necessitated that athletes procure their own transpor-

tation to and from the school to receive free bus transport to the PPE event, which was held on a weekend. Some student-athletes may not have been able or willing to find such transportation and thus were not included in the study. For that reason, our results may have been inadvertently biased to exclude some individuals from a low socioeconomic background.

## CONCLUSIONS

Increased mental health screening can be achieved by capitalizing on the opportunity that exists in the mandatory sports PPE. Our findings 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 appeared to serve as a multidimensional assessment of depression in our participants as a whole, but construct bias was evident between Black and White adolescent athletes. A prospective analysis controlling for confounding variables is needed. These results 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 somatic and affective components of depression in adolescent athletes, but Black and White participants may interpret the questions differently and, thus, a comparison of total scores between groups may not be appropriate. Racially sensitive depression screening tools for adolescents deserve further investigation.

## REFERENCES

1. GBD 2017 Disease and Injury Incidence and Prevalence Collaborators. Global, regional, and national incidence, prevalence, and years lived with disability for 354 diseases and injuries for 195 countries and territories, 1990-2017: a systematic analysis for the Global Burden of Disease Study 2017. *Lancet*. 2018;392(10159):1789-1858. doi:10.1016/S0140-6736(18)32279-7
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;1:79-91. doi:10.1080/1612197X.2020.1802612
3. Zuckerbrot RA, Cheung A, Jensen PS, Stein REK; GLAD-PC STEERING GROUP. Guidelines for Adolescent Depression in Primary Care (GLAD-PC): part I. Practice preparation, identification, assessment, and initial management. *Pediatrics*. 2018;141(3):e20174081. doi:10.1542/peds.2017-4081
4. Chang C, Putukian M, Aerni G, et al. Mental health issues and psychological factors in athletes: detection, management, effect on performance and prevention: American Medical Society for Sports Medicine Position Statement—executive summary, *Br J Sports Med*. 2020;54(4):216-220. doi: 10.1136/bjsports-2019-101583
5. Kroenke K, Spitzer RL, Williams JB. 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
6. Johnson JG, Harris ES, Spitzer RL, Williams JB. The patient health questionnaire for adolescents: validation of an instrument for the assessment of mental disorders among adolescent primary care

- patients. *J Adolesc Health*. 2002;30(3):196–204. doi:10.1016/s1054-139x(01)00333-0
7. Lu W, Lindsey MA, Irsheid S, Nebbit VE. Psychometric properties of the CES-D among Black adolescents in public housing. *J Soc Social Work Res*. 2017;8(4):595–619. doi:10.1086/694791
  8. American Psychiatric Association. *DSM-IV: Diagnostic and Statistical Manual of Mental Disorders*. American Psychiatric Press Inc: 1994
  9. Tabachnick BG, Fidell LS. *Using Multivariate Statistics*. 6th ed. Pearson; 2012.
  10. *IBM SPSS Statistics for Windows, Version 25.0*. IBM Corporation; 2017.
  11. Arbuckle JL. *AMOS, Version 23*. SPSS; 2014.
  12. Kline RB. *Principles and Practice of Structural Equation Modeling*. 2nd ed. Guilford Press; 2005.
  13. Keum BT, Miller MJ, Inkelas KK. Testing the factor structure and measurement invariance of the PHQ-9 across racially diverse US college students. *Psychol Assess*. 2018;30(8):1096–1106. doi:10.1037/pas0000550
  14. Cheung GW, Rensvold RB. Evaluating goodness-of-fit indexes for testing measurement invariance. *Struct Equ Model*. 2009;9(2):233–255. doi:10.1207/S15328007SEM0902\_5
  15. Hu L-T, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. *Struct Equ Model*. 1999;6:1–55. doi:10.1080/10705519909540118
  16. Crane PK, Gibbons LE, Willig JH, 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. doi:10.1080/09540120903483034
  17. Baas KD, Cramer AO, Koeter MW, van de Lisdonk EH, van Weert HC, Schene AH. Measurement invariance with respect to ethnicity of the Patient Health Questionnaire-9 (PHQ-9). *J Affect Disord*. 2011;129(1–3):229–235. doi:10.1016/j.jad.2010.08.026
  18. Muthén B, Christofferson A. Simultaneous factor analysis of dichotomous variables in several groups. *Psychometrika*. 1981;46(4):407–419. doi:10.1007/BF02293798
  19. Norman G. Likert scales, levels of measurement and the “laws” of statistics. *Adv Health Sci Educ Theory Pract*. 2010;15(5):625–632. doi:10.1007/s10459-010-9222-y
  20. Harry ML, 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 Rep*. 2021;4:100121. doi:10.1016/j.jadr.2021.100121
  21. 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. doi:10.3389/fpsy.2016.00178
  22. Elhai JD, Contractor AA, 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. doi:10.1016/j.psychres.2012.05.018
  23. Borgogna NC, Brenner RE, McDermott RC. Sexuality and gender invariance of the PHQ-9 and GAD-7: implications for 16 identity groups. *J Affect Disord*. 2021;278:122–130. doi:10.1016/j.jad.2020.09.069
  24. Cohen J. *Statistical Power Analysis for the Behavioral Sciences*. 2nd ed. Lawrence Erlbaum Associates; 1988.
  25. Granillo MT. Structure and function of the Patient Health Questionnaire-9 among Latina and Non-Latina white female college students. *J Soc Soc Work Res*. 2012;3(2):80–93. doi:10.5243/jsswr.2012.6
  26. 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. doi:10.1016/j.jpsychores.2012.12.012
  27. 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. doi:10.1016/j.jad.2017.05.020
  28. Ofonedu ME, Percy WH, Harris-Britt A, Belcher HME. Depression in inner city African American youth: a phenomenological study. *J Child Fam Stud*. 2013;22(1):96–106. doi:10.1007/s10826-012-9583-3
  29. Chang TE, Weiss AP, Marques L, et al. Race/ethnicity and other social determinants of psychological well-being and functioning in mental health clinics. *J Health Care Poor Underserved*. 2014;25(3):1418–1431. doi:10.1353/hpu.2014.0138
  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 RW, Perrin EM, Coyne-Beasley T, Peterson CJ, Skinner AC. Reported sports participation, race, sex, ethnicity, and obesity in the US adolescents from NHANES Physical Activity (PAQ-D). *Glob Pediatr Health*. 2015;2:2333794X15577944. doi:10.1177/23337944
  32. Estrada-Martinez LM, Caldwell CH, Bauermeister JA, Zimmerman MA. 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. doi:10.1007/s10964-012-9778-3
  33. Treadway MT, Zald DH. Reconsidering anhedonia in depression: lessons from translational neuroscience. *Neurosci Biobehav Rev*. 2011;35(3):537–555. doi:10.1016/j.neubiorev.2010.06.006
  34. 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):e0199235. doi:10.1371/journal.pone.0199235
  35. Kerr LK, Kerr LD. Screening tools for depression in primary care. *West J Med*. 2001;175(5):349–352. doi:10.1136/ewjm.175.5.349

Address correspondence to Ashley Long, PhD, Mount Olive Family Medicine Center, 201 North Breazeale Avenue, Mount Olive, NC 28365. Address email to ashleylong111@gmail.com.