FACTORIAL VALIDITY AND INVARIANCE OF THE CENTER FOR EPIDEMIOLOGIC STUDIES DEPRESSION (CES-D) SCALE IN A SAMPLE OF BLACK AND WHITE ADOLESCENT GIRLS

Derek P. Hales, MA; Rod K. Dishman, PhD; Robert W. Motl, PhD; Cheryl L. Addy, PhD; Karin A. Pfeiffer, PhD; Russell R. Pate, PhD

Meaningful comparison of depression symptoms requires that the measurement instrument has equivalent measurement properties among racial and ethnic groups. We tested the factorial validity and invariance of the Center for Epidemiologic Studies-Depression (CES-D) Scale among Black (n=610) and White (n=452) adolescent girls in the 12th grade. The invariance analyses were conducted by using LISREL 8.70 with maximum likelihood estimation and Satorra-Bentler scaled chi-square statistics and standard errors. The hypothesized second-order model (first-order factors: depressed affect, somatic and retarded activity, interpersonal, and positive affect; second-order factor: depression) demonstrated good overall fit in both groups. Comparison of nested models supported the between-group invariance of the overall factor structure, first- and second-order factor loadings, first-order factor variances, and the second-order factor variance. Item uniquenesses were not invariant. Our results support the hypothesis that a meaningful comparison of composite CES-D scores can be made between Black and White girls in the 12th grade. (*Ethn Dis*. 2006;16:1–8)

Key Words: African American, Confirmatory Factor Analysis, Depression Symptoms, Measurement, Psychometrics

From the Department of Exercise Science, University of Georgia, Athens, Georgia (DPH, RKD); Department of Kinesiology, University of Illinois, Champaign, Illinois (RWM); Department of Epidemiology and Biostatistics (CLA) and Department of Exercise Science (KAP, RRP), University of South Carolina, Columbia, South Carolina.

Address correspondence and reprint requests to Rod K. Dishman, PhD; Department of Exercise Science; The University of Georgia; Ramsey Student Center, 300 River Road; Athens, GA 30602-6554; 706-542-9840; 706-542-3148 (fax).

INTRODUCTION

The annual rate of depression among teenagers and young adults in the United States is nearly twice that of adults 25–44 years old,[1] and an estimated 15%–30% of adolescents will experience an episode of depression.[2–4] Moreover, major depressive disorder is occurring at an earlier age than in the past,[5] and it reoccurs in 60% to 70% of adults who experienced major depression as children or adolescents.[2,4–7] Hence, growing public health concern exists regarding the prevention and treatment of adolescent depression.[6] This concern is especially pronounced for girls, who have higher prevalence rates of major depressive episode than boys during late adolescence.[3]

> *. . .major depressive disorder is occurring at an earlier age than in the past,[5] and it reoccurs in 60% to 70% of adults who experienced major depression as children or adolescents.[2,4–7]*

The Center for Epidemiologic Studies Depression (CES-D) Scale[8] is a practical, self-report measure used to assess the presence and severity of depression symptoms in population-based studies. The CES-D consists of 20 items that correspond with seven of nine criterion-based symptoms of major depressive disorder within the fourth edition of the Diagnostic and Statistical Manual of Mental Disorders.[9] The items were selected from a pool of items from validated depression scales and represented the primary components of depressive symptoms. The 20 items were pretested on small "samples of convenience" and this indicated appropriate performance plus minor revisions for clarity and acceptability. The 20-item CES-D was then field tested in three large community samples and found to have acceptable test-retest reliability, internal consistency, and evidence of score validity. Finally, exploratory factor analyses using principal components extraction yielded a four factor structure that was consistent across three groups and reflected the components of "depressed affect," "positive affect," "somatic and retarded activity," and "interpersonal." Although not tested, the theoretical structure of the CES-D implied that this scale consisted of a single second-order factor, namely depression, and this factor caused the interrelationships among the four first-order factors of depressed affect (7 items), somatic and retarded activity (7 items), interpersonal (2 items), and positive affect (4 items).[8] Several exploratory and confirmatory factor analyses have supported this model of the CES-D among adults[8,13,22,25] and young adolescent boys and girls.[26] However, others have reported that one-, two-, and three-factor models fit CES-D data best.[24,27] The CES-D was originally designed for adults, but the core diagnostic symptoms of depression are similar for adults and adolescents, and the CES-D has acceptable validity to screen depression among adolescents.[10–13]

However, information is limited about the factorial validity and invariance of the CES-D among adolescents. Factorial validity is the degree to which a measure conforms to the theoretical definition of the construct it assesses[14–15]; it is explored by testing the fit of a theoretically based measurement model for describing the variances and covariances underlying items on a scale.[14,16] Factorial invariance is the degree to which a construct is measured similarly across points of time or between groups of people; it is explored by testing the comparability of the form and values of parameters within a measurement model across time or between groups.[14,17] Evidence for factorial validity and invariance is necessary before scores on the CES-D are assumed to have the same meaning and are used to compared groups of people.

To our knowledge, no studies have compared the factorial validity and invariance of CES-D scores between Black and White girls in late adolescence, when risk of depression is elevated.[11,18–20] This comparison is necessary because previous researchers have reported a higher prevalence of depression symptoms among Black adolescent girls compared with White girls, but this difference might be attributable to violations of measurement invariance rather than true group differences. Therefore, the purpose of this study was to examine the factorial validity and invariance of the hypothesized four-factor, second-order model of the CES-D in a sample of Black and White girls in the 12th grade.

METHODS

Participants

Participants (N=1062) were Black (n=610) and White (n=452) 12th-grade girls recruited from 22 public high schools in South Carolina. The mean age of the girls was 17.7 (SD=.6) years. The Black girls were slightly older (.2 years), t (1,1060)=4.1, P<.01, and had a higher body mass index (BMI), (26.4 ± 7.2) than White girls (23.8 ± 5.3), t (1,1041)=6.2, P<.01. Parental education levels of Black girls compared with White girls were lower for mothers ($\chi^2_{(5)}$=58.14, P<.001) and fathers ($\chi^2_{(5)}$=110.8, P<.001). A high school diploma was the highest educational degree for 41% of the mothers and 44% of the fathers of Black girls compared to 28% each for the mothers and fathers of White girls. Fourteen percent of fathers and 11% of mothers of Black girls had not completed high school compared to 10% of fathers and 6% of mothers of White girls. Approximately 22% of fathers and 32% of mothers of Black and White girls had vocational training or some college after high school. Thirty-nine percent of the fathers and 35% of the mothers of White girls had a college degree or post-baccalaureate education compared to 11% of the fathers and 15% of the mothers of Black girls. Twenty percent of Black girls (n=124) and 14% of White girls (n=65) had elevated CES-D scores (≥24).[21]

Measures

The 20-item CES-D[8] was used to measure depressive symptoms and was administered by trained data collectors to groups of 6 to 10 girls in a quiet classroom. Test administration was approved by the University of South Carolina Institutional Review Board. Participants and their parent or legal guardian provided written informed consent. Each of the items is rated on a zero to three scale based on the frequency of occurrence "during the past week" [Rarely or none (<1 day)=0, Some or a little (1–2 days)=1, Occasionally (3–4 days)=2, Most or all (5–7 days)= 3]. A summary of CES-D items can be found in Table 1.

Statistics

Analysis and Fit

The factorial validity and invariance analyses were conducted by using maximum likelihood estimation with the Satorra-Bentler scaled chi-square statistic and standard errors in LISREL 8.7.[28–29] The Satorra-Bentler procedure is a normal theory method that adjusts the chi-square statistic and standard errors for the observed multivariate kurtosis of the data.[29] Item/scale descriptive statistics and group comparisons were obtained by using SPSS 13.0.[30] PRELIS 2.5 was used to estimate univariate and multivariate kurtosis.[31] Mardia coefficient of multivariate kurtosis was significant for both Black (z=26.32, P<.001) and White (z=22.16, P<.001) girls, which indicates a violation of multivariate normality.[32] The mean and median for the absolute values of univariate kurtosis estimates were 1.11 and .896, respectively, with a range of .004 to 4.84. Satorra-Bentler scaled chi-squares and standard errors were used to minimize the bias in standard errors and test statistics that can result from excessive multivariate kurtosis.

Multiple indicators were estimated to compare and assess model fit. The Satorra-Bentler scaled χ^2 statistic was used to assess absolute fit of the model to the data. This statistic is very sensitive to sample size and in most covariance modeling calls for rejection of the hypothesized model.[16,33] For this reason, it is reported but not used to draw specific conclusions about model fit. The root mean square error of approximation (RMSEA) is a standardized estimate that represents closeness of fit of population data to the model. It is widely used and is considered one of the most informative fit criteria.[34] Values of the RMSEA ≤.05 reflect close fit of the model.[35–36] The 90% confidence interval (CI) around the RMSEA point estimate is also presented to indicate the possibility of close or exact fit. The

Table 1. Items, means (SD), Cronbach α, and invariant parameter estimates for CES-D

	Black Girls		White Girls		Parameter Estimates	
Scale/Item (item#)	**Mean (SD)**	**α**	**Mean (SD)**	**α**	**Unstandardized (SE)**	**Standardized**
Depressed affect	3.91 (4.44)	.862	3.39 (4.07)	.882	1.00	1.00
Blues (3)	.879 (.998)		.602 (.838)		1.00	.64
Depressed (6)	.593 (.882)		.496 (.781)		1.07 (.046)	.68
Failure (9)	.367 (.751)		.230 (.565)		.67 (.044)	.42
Fearful (10)	.402 (.726)		.365 (.654)		.61 (.046)	.39
Lonely (14)	.577 (.889)		.498 (.792)		1.01 (.051)	.64
Crying (17)	.526 (.892)		.580 (.866)		.94 (.054)	.6
Sad (18)	.562 (.839)		.619 (.780)		1.03 (.048)	.65
Somatic activity	6.56 (3.72)	.656	5.84 (3.31)	.640	.68 (.043)	.92
Bothered (1)	.879 (.868)		.850 (.807)		1.00	.47
Appetite (2)	.762 (.892)		.670 (.774)		.73 (.070)	.34
Mind (5)	1.015 (.963)		1.082 (.873)		1.14 (.085)	.53
Effort (7)	1.682 (.976)		1.186 (.949)		.68 (.096)	.32
Sleep (11)	.800 (.976)		.832 (.948)		1.07 (.093)	.5
Talked (13)	.803 (.964)		.628 (.779)		.96 (.076)	.45
Get going (20)	.618 (.875)		.588 (.716)		1.16 (.084)	.54
Positive affect-R*	3.42 (2.71)	.704	3.59 (2.67)	.800	.51 (.041)	.64
Good (4)	1.053 (1.016)		1.126 (.908)		1.00	.5
Hopeful (8)	.875 (.961)		.934 (.845)		1.01 (.076)	.51
Happy (12)	.787 (.848)		.796 (.815)		1.29 (.087)	.65
Enjoyed (16)	.710 (.883)		.732 (.808)		1.30 (.092)	.66
Interpersonal	1.06 (1.45)	.660	.81 (1.21)	.695	.58 (.051)	.77
Unfriendly (15)	.513 (.803)		.431 (.697)		1.00	.48
Disliked (19)	.549 (.878)		.381 (.687)		1.34 (.11)	.64
Total CES-D score	14.95 (9.993)	.883	13.626 (9.190)	.893		

* Positive affect-R indicates that subscale items were reversed prior to scoring.

Note: Estimate of loading for PA to Effort (7)=−.85 (.094) and −.043. Parameter estimates are from model 5 (all parameters except item uniquenesses held invariant). Error variance of DA constrained to .001 in both groups (see results section).

SD=standard deviation; SE=standard error; CES-D=Center for Epidemiologic Studies Depression.

comparative fit index (CFI) and non-normed fit index (NNFI) test the proportionate improvement in fit by comparing the target model to some baseline model.[37–38] The NNFI is affected by model parsimony (more complex models are penalized), but the CFI is not. Values for the CFI and NNFI ≈.90 are considered acceptable.[36–37] Values ≥.95 indicated good fit.[36] Scaled χ^2 statistics (independence and target model) were used to calculate CFI and NNFI.

Nested models in the invariance analysis were compared based on Satorra and Bentler's corrected χ^2 difference tests[39] and changes in the values of the RMSEA, NNFI, and CFI between nested models Although χ^2 difference tests were conducted and reported, the χ^2 difference test is sensitive to sample size issues and thus conclusions are based primarily on differences in CFI, NNFI, and RMSEA between models. Examining differences in these three measures is superior to interpretations based strictly on χ^2 difference tests.[40] The criterion for the NNFI and CFI was a change of .01 between nested models as this has been reported to work well for testing multigroup invariance.[40] Overlap in the RMSEA point estimates and 90% CIs between two nested models were used to judge meaningful change in fit between models. Parameter estimates, standard errors, z values, and squared multiple correlations were inspected for sign and magnitude. Independent samples t tests with a Bonferroni correction for multiple comparisons (25 comparisons; $P<.002$ significant) were used to compare item and scale means between Black and White girls. Item and scale comparisons between groups were also adjusted for potential confounding effects of parental education with analysis of covariance (ANCOVA), with mother's and father's education levels used as covariates.

Model

The models pictured in Figures 1 and 2 were initially tested for the samples of Black and White girls separately. This test allowed the adequacy of the model to be assessed within each group prior to the multigroup invariance analysis. Each of the models was based on the original four-factor model and subsequent research into the structure of the scale.[26,41–42]

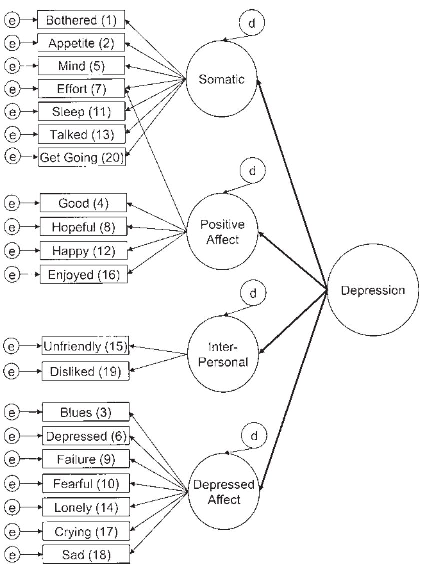

Fig 1. Model depicting hierarchical second-order factor structure including four first-order factors

The correlated factors model (Figure 2) does not include the second-order factor depression and was tested for comparison to the theoretical higher-order model.[43] If the correlated factors model fits substantially better than the second-order model, then the relationships among the factors are not adequately explained by a single second-order factor. If the fit of the two models was similar, the second-order model could be interpreted as more parsimonious and, if theoretically sound, it would be the better of the two models.

The primary departure from the original structure of the CES-D involved allowing item 7, effort, to load on both somatic and positive affect factors. Inclusion of this cross-loading was based on prior findings.[26] In addition, the disturbance term for the depressed affect factor was found to be small (Black=−.001; White=.001) in both groups. Because the value was negative for the sample of Black girls, two tests were conducted to determine if the improper solution could be attributed to sampling fluctuation.[44] The tests included: 1) forming a CI around the negative error variance by using the asymptotic standard errors to see if it included zero[45]; and 2) computing a χ^2_{diff} test between the constrained (setting suspect parameter to .001) and the unconstrained models. Each of these tests indicated that the negative error variance could be the result of sample fluctuation (95% CI=−.027–.024 [includes zero]; χ^2_{diff} (1)=.03 [is nonsignificant]). For this reason the value of the disturbance term for depressed affect was constrained to .001 in all analyses.

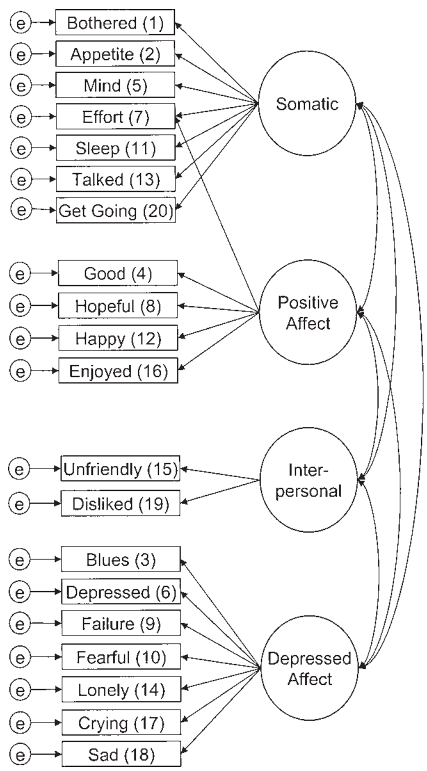

Fig 2. Model depicting correlated first-order factors

The invariance analysis involved testing and comparing six models. Each model (M1 to M6) included previous model restrictions (ie, M3 included restrictions from M2) plus additional constraints, resulting in a series of nested models, model 1 (M1) tested the equivalence of the hypothesized pattern of paths, factor variances, and item uniquenesses across groups. In this model, all hypothesized parameters were freely estimated in the two groups (depressed affect disturbance term was fixed). Model 2 (M2) restricted paths from the first-order factors to the observed items across groups. In model 3 (M3), the paths from the second-order factor (depression) to each of the first-order factors were added to those being held invariant. The variance of the second-order factor was constrained in model 4 (M4), while in model 5 (M5) the disturbance terms for the first-order factors were set equal for Black and White girls. Finally, in model 6 (M6) the observed item uniquenesses (errors) were held equivalent across groups. Item uniquenesses (error terms) reflect both random and systematic variance not explained by the factor. Testing their equivalence is very restrictive and does not need to hold for an instrument to be considered invariant across groups.[34] When model fit is compared across gradually more restrictive models (more parameters constrained to be equal across groups), it can be determined if model fit is affected by constraining groups of parameters to be equal across groups. Equivalence of factor structure and first- and second-order loadings was considered the minimal criteria for concluding invariance across groups.

RESULTS

A summary of model fit indices for Black and White girls and for the multigroup invariance analyses can be found in Table 2. Results for the sample of Black and White girls provide evidence that supports the hypothesized second-order factor model. The RMSEA (90% CI), CFI, and NNFI suggest good fit of the model to the data for both groups. The similarity in the RMSEA, CFI, NNFI, and χ^2 suggests that the fit of the correlated factors model was not different than the second-order model for Black (χ^2_{diff}= 8.01 (3), P=.05) or White (χ^2_{diff}=2.66 (3), P=.45) girls. This finding indicates that allowing all factors to correlate does

Table 2. Fit and comparison of models used to test the factorial validity and invariance of the CES-D in a sample of Black and White girls

Black girls	χ^2	*df*	NNFI	CFI	RMSEA (90% CI)
2nd-order 4-factor model (Figure 1)	374.44	166	.972	.976	.045 (.039–.052)
Correlated factors (Figure 2)	366.17	163	.972	.976	.045 (.039–.051)
White girls					
2nd-order 4-factor model (Figure 1)	261.57	166	.983	.985	.036 (.027–.044)
Correlated factors (Figure 2)	259.09	163	.982	.985	.036 (.028–.044)
Black and White girls					
Independence	22495.13	380			
M1 - All parameters free	636.35	332	.984	.986	.042 (.037–.046)
M2 - 1st-order loadings	674.56	349	.984	.985	.042 (.037–.047)
M3 - M2 + 2nd-order loadings	692.23	352	.983	.985	.043 (.038–.047)
M4 - M3 + Depression variance	695.03	353	.983	.985	.043 (.038–.047)
M5 - M4 + Error variances for 1st-order factors	696.66	356	.984	.985	.042 (.038–.047)
M6 - M5 + Item uniquenesses	931.22	376	.975	.975	.053 (.049–.057)
Model comparisons†	$\Delta\chi^{2*}$	Δdf	*P*		
M1 vs M2	38.21	17	.0023		
M2 vs M3	21.43	3	.0001		
M3 vs M4	2.59	1	.1078		
M4 vs M5	1.46	3	.6927		
M5 vs M6	289.93	20	<.0001		

* Scaled χ^2 is Satorra-Bentler scaled chi-square.
† Model comparisons computed by using scaled chi-square adjustment.

not substantially improve model fit for Black or White girls.

The fit of model M1 for the multigroup analysis was also good (CFI=.99, NNFI=.98, RMSEA [90% CI]=.042 [.037–0.046]). The χ^2_{diff} tests comparing models M1 and M2, M2 and M3, and M5 and M6 were statistically significant. The similarity of fit (see Table 1) for models M1, M2, M3, M4, and M5 (all models with NNFI=.98, CFI=.99, and RMSEA=.042/.043) do, however, support the invariance of the factor structure, first-order loadings, second-order loadings, second-order factor variance, and first-order disturbance terms across Black and White girls. Based on the large χ^2_{diff} value combined with a decrease in CFI and an increase in RMSEA, the item uniquenesses were judged not to be invariant. The standardized and unstandardized path coefficients and standard errors for model M5 are presented in Table 1.

Post-hoc results from our sample indicate that one- and two-factor models fit substantially worse for both Black and White girls compared to the four-factor correlated model and the second-order model. A three-factor second-order model (somatic and depressed affect combined) also fit substantially worse compared to the four-factor correlated model and the four-factor second-order model for White girls. For Black girls the three-factor second-order model produced a nonsignificant negative error variance for the depressed affect-somatic factor. Setting this value to .001 (similar to the four-factor second-order model) or .0 resulted in improper solutions (negative error variances, very large parameter estimates). Hence, the four-factor second-order model represented the best fit for CES-D scores for the Black girls.

Group comparisons were conducted by using independent samples *t* tests with a Bonferroni correction for multiple comparisons. Effect size was also calculated ([mean difference]/[pooled standard deviation]). In addition, differences in the proportion of Black and White girls above commonly used CES-D cutoff scores were examined with a χ^2 test. A summary of item and scale means can be found in Table 1. Mean (SD) scores were slightly higher for Black girls compared to White girls, respectively, for the total CES-D (14.94 [9.99] vs 13.63 [9.19], $t\,[1,1060]=2.2$, $P=.03$) and the somatic (6.56 [3.72] vs 5.84 [3.31], $t\,[1,1060]=3.33$, $P<.001$) subscale. Although statistically significant, the effect sizes of these differences are small (Cohen d=.14 and .20), which suggests that they may not be clinically meaningful. The differences were no longer significant ($P>.21$) after covariance adjustment for the lower education levels of the Black girls' parents. Among Black girls, 38.2% scored ≥16 and 20.3% scored ≥24 on the CES-D. For White girls, 31.4% scored ≥16 while 14.4% scored ≥24. Differences in the proportion of Black and White girls scoring ≥16 ($\chi^2[1]=5.2$, $P=.022$) and ≥24 ($\chi^2[1]=6.3$, $P=.012$) were statistically significant. The mean differences for five items (3, 7, 9, 13, and 19) were statistically significant after correcting for multiple comparisons; differences remained significant for items 3, 7, and 13 and reached significance for items 17 and 18 ($P<.05$) after covariance adjustment for parental education. Effect sizes for these differences ranged from .198 to .515 (median=.206), with all but one <.30. The largest effect size (.515) was for item 7, effort.

Cronbach α coefficient was also calculated for each subscale and for the CES-D as a whole. Values for the subscales ranged from .640 to .882 and were very similar for Black and White girls. The largest difference was for the positive affect scale. The internal consistency of the whole scale was good for both Black (α=.883) and White (α=.893) girls.

DISCUSSION

We examined the factorial validity and invariance of the hypothesized higher-order factor model for CES-D scores in a sample of White and Black adolescent girls. The results support the factorial validity of the second-order model in both groups. The model (Figure 1) fit slightly better for White compared to Black girls. In addition, the first-order and second-order factor loadings, variance of the depression factor, and error variances of depressed affect, positive affect, somatic, and interpersonal factors were invariant between samples of Black and White adolescent girls. Although, item uniquenesses were not invariant, the equivalence of factor structure and loadings indicates that the theoretical relations among items from the CES-D are similar across Black and White girls in this sample.

Factorial invariance of the CES-D has been supported across age and sex.[13,24,26,42] Less is known about whether CES-D scores have measurement equivalence between Blacks and Whites. Previous exploratory factor analyses of CES-D scores among Black and White adults yielded mixed results, supporting similar[46] and different[46–48] factor structures. Previous exploratory and confirmatory work suggested that two- or three-factor models of CES-D scores provide a better fit compared to one- or four-factor models in some racial/ethnic groups.[24,27,49] In the present sample, the four-factor second-order model represented the best fit for CES-D scores for Black and White girls.

Some caution must be taken when interpreting the parameter estimates and standard errors associated with the depressed affect factor because of the improper solution and constraint placed on the negative error variance. Thus, the parameters and standard errors most closely associated with the error variance for the depressed affect factor may be biased.[44,50] Researchers attribute negative error variance in confirmatory factor analysis to sampling fluctuation, model misspecification (no model fits the data), indefiniteness of the model, empiric underidentification, and outlier/influential cases.[44–45] Based on our previous research on the CES-D in another sample of adolescents,[26] we believe that the negative error variance for the depressed affect factor in the four-factor second-order model (see Figure 1) was most likely the result of sampling fluctuation. The effect of the constraint we imposed is unknown, but the size of our sample and the number of indicators for the depressed affect factor likely minimized parameter bias.

A growing body of research exists on shortened versions of the CES-D and their usefulness or appropriateness for certain populations. Several studies have supported the validity and reliability of 10-, 11-, or 5-item CES-D scales. Radloff et al[23] reported the results of a factor analysis conducted on ≈2800 youth and adults (≥18 years) and found the CES-D scores to have a four-factor structure similar to the original, but included only 17 of the original 20 items (items 9, 10, and 13 removed). Although the factorial validity and ability of the revised scale to identify depressed adolescents has not been tested, eliminating items may help to alleviate a possible redundancy problem in the depressed affect and somatic subscales and reduce the item burden of the scale.

Although item uniquenesses were not invariant, the equivalence of factor structure and loadings indicates that the theoretical relations among items from the CES-D are similar across Black and White girls in this sample.

The internal consistency of the CES-D was similar for Black (alpha=.883) and White (alpha=.893) girls. This finding is consistent with previous research.[11,21,23] Even though the factor structure of the CES-D has consistently included two or more factors, we found only one study of adolescents that reported the internal consistency for each subscale. The alpha estimates from our sample were slightly higher than those reported in a high school group by Radloff et al.[23] The largest difference was for the positive affect subscale. Why the internal consistencies are larger in our sample is unclear. Natural parameter fluctuation or differences in the composition of the samples (ie, age, racial distribution, and sex) are possible causes of the observed difference.

The total CES-D score from our sample was lower than those reported from other samples of high school students[21,32] but very similar to those reported by Iwata[18] for a sample of White and Black young adults. The difference in total CES-D score between White and Black girls in our sample was also similar to those found in other studies.[18,47] However, the slightly higher CES-D score in Black girls compared to White girls was not statistically significant after controlling for parental education levels. Nonetheless, the proportion of Black girls at or above commonly used cut scores was significantly higher than the proportion of White girls. Researchers have reported that 28%–59% of adolescent girls score ≥16 and 9%–28% score ≥24 on the CES-D.[18–21,23] The confounding influence of socioeconomic status or the modifying effects of cultural factors on these differing rates requires elucidation.

The means for items 3 (blues), 7 (effort), 9 (failure), 13 (talked), and 19 (dislike) and the somatic subscale were higher among Black than White girls, even after adjusting for lower parental education levels for Black girls. Studies comparing Black and White adults have reported mean differences as well as differential item functioning for similar items and subscales of the CES-D. In their samples of similar size to ours, Iwata et al[18] not only found significant mean differences between Black and White non-Hispanic young adults for items 2, 7, 9, 11, 13, and the somatic subscale, they also detected differential item functioning for 13 of the 20 CES-D items when comparing Black and White participants.

Because of the utility and widespread use of the CES-D, research must continue to move from exploratory to more confirmatory approaches when validating measures of depressive symptoms among adolescents. A better understanding of the factor structure and invariance of the CES-D will allow for more clear comparisons and more accurate estimates of depression risk. For example, the factor structure and invariance of the CES-D should be examined across different subgroups and over a larger age range.

Finally, we must determine why some differences in depression symptoms exist between White and Black girls and whether they reflect true differences in the risk of clinically diagnosed depression. Our findings indicate that the measurement structure of the CES-D is similar across Black and White girls, but they do not rule out individual item bias in the CES-D. Socio-cultural factors, other than parental education, associated with race could influence not only responses to the items on the CES-D but the expression of depressive symptoms in Black and White girls. For example, higher rates of depression symptoms among non-Hispanic Black compared to non-Hispanic White women in the United States are largely explained by lower socioeconomic status.[51–53] Such sociocultural factors, though not assessed in the present study, could influence responses to selected items that assess specific symptoms of depression and require further study in girls.

REFERENCES

1. Kessler RC, McGonagle KA, Zhao S, et al. Lifetime and 12-month prevalence of DSM-III-R psychiatric disorders in the United States. Results from the National Comorbidity Survey. *Arch Gen Psychiatry.* 1994;51(1):8–19.
2. Birmaher B, Ryan ND, Williamson DE, et al. Childhood and adolescent depression: a review of the past 10 years. Part I. *J Am Acad Child Adolesc Psychiatry.* 1996;35(12):1427–1439.
3. Kessler RC, Walters EE. Epidemiology of DSM-III-R major depression and minor depression among adolescents and young adults in the National Comorbidity Survey. *Depress Anxiety.* 1998;7(1):3–14.
4. Lewinsohn PM, Rohde P, Seeley JR. Major depressive disorder in older adolescents: prevalence, risk factors, and clinical implications. *Clin Psychol Rev.* 1998;18(7):765–794.
5. Cross-National Collaborative Group. The changing rate of major depression. *JAMA.* 1992;268(21):3098–3105.
6. Weissman MM, Wolk S, Goldstein RB, et al. Depressed adolescents grown up. *JAMA.* 1999;281(18):1707–1713.
7. National Institute of Mental Health (NIMH). *Depression in Children and Adolescents: A Fact Sheet for Physicians.* Bethesda, Md: NIMH; 2000. NIH Publication No. 00-4744.
8. Radloff LS. The CES-D Scale: a self-report depression scale for research in the general population. *Appl Psychol Meas.* 1977;1:385–401.
9. Okun A, Stein RE, Bauman LJ, Silver EJ. Content validity of the Psychiatric Symptom Index, CES-depression Scale, and State-Trait Anxiety Inventory from the perspective of DSM-IV. *Psychol Report.* 1996;79(3, pt 1):1059–1069.
10. Dierker LC, Albano AM, Clarke GN, et al. Screening for anxiety and depression in early adolescence. *J Am Acad Child Adolesc Psychiatry.* 2001;40:929–936.
11. Garrison CZ, Addy CL, Jackson KL, McKeown RE, Waller JL. The CES-D as a screen for depression and other psychiatric disorders in adolescents. *J Am Acad Child Adolesc Psychiatry.* 1991;30(4):636–641.
12. Garrison CZ, Jackson KL, Marsteller F, McKeown R, Addy C. A longitudinal study of depressive symptomatology in young adolescents. *J Am Acad Child Adolesc Psychiatry.* 1990;29(4):581–585.
13. Roberts RE, Andrews JA, Lewinsohn PM, Hops H. Assessment of depression in adolescents using the Center for Epidemiologic Studies Depression Scale. *Psychol Assess.* 1990;2:122–128.
14. Hoyle RH, Smith GT. Formulating clinical research hypotheses as structural equation models: a conceptual overview. *J Consult Clin Psychol.* 1994;62:429–440.
15. Messick S. Validity of psychological assessment: validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *Am Psychol.* 1995;50:741–749.
16. Bollen KA. *Structural Equations with Latent Variables.* New York, NY: John Wiley & Sons; 1989.
17. Vandenberg RJ, Lance CE. A review and synthesis of the management invariance literature: suggestions, practices, and recommendations for organizational research. *Org Res Methods.* 2000;3:4–69.
18. Iwata N, Turner RJ, Lloyd DA. Race/ethnicity and depressive symptoms in community-dwelling young adults: a differential item functioning analysis. *Psychiatry Res.* 2002;110:281–289.
19. Kubik MY, Lytle LA, Birnbaum AS, Murray DM, Perry CL. Prevalence and correlates of depressive symptoms in young adolescents. *Am J Health Behav.* 2003;27:546–553.
20. Rushton JL, Forcier M, Schectman RM. Epidemiology of depressive symptoms in the National Longitudinal Study of Adolescent Health. *J Am Acad Child Adolescent Psychiatry.* 2002;41:199–205.
21. Roberts RE, Lewinsohn PM, Seeley JR. Screening for adolescent depression - a comparison of depression scales. *J Am Acad Child Adolesc Psychiatry.* 1991;30:58–66.

22. Clark VA, Aneshensel CS, Frerichs RR, Morgan TM. Analysis of effects of sex and age in response to items on the CES-D scale. *Psychiatry Res*. 1981;5:171–181.
23. Radloff, LS. The use of the Center for Epidemiological Studies Depression Scale in adolescents and young adults. *J Youth Adolesc*. 1991;20:149–166.
24. Dick RW, Beals J, Keane EM, Manson SM. Factorial structure of the CES-D among American Indian adolescents. *J Adolesc*. 1994;17:73–79.
25. McArdle JJ, Johnson RC, Hishinuma ES, Miyamoto RH, Andrade NN. Structural equation modeling of group differences in CES-D ratings of native Hawaiian and non-Hawaiian high school students. *J Adolesc Res*. 2001;16:108–149.
26. Motl RW, Dishman RK, Birnbaum AS, Lytle LA. Longitudinal invariance of the Center for Epidemiologic Studies Depression Scale among girls and boys in middle school. *Ed Psychol Meas*. 2005;65:90–108.
27. Edman JL, Danko GP, Andrade N, McArdle JJ, Foster J, Glipa J. Factor structure of the CES-D among Filipino-American adolescents. *Soc Psychiatry Psychiatr Epidemiol*. 1999;34: 211–215.
28. Jöreskog KG, Sörbom D. *LISREL 8: User's Reference Guide*. Chicago, Ill: Scientific Software International; 1996.
29. Satorra A, Bentler PM. Scaling corrections for chi-square statistics in covariance structure analysis. *American Statistical Association: 1988 Proceedings of the Business and Economics Sections*. 1988;308–313.
30. *SPSS 13.0 for Windows*. Chicago, Ill: SPSS, Inc; 2005.
31. Joreskog K, Sorbom D. *PRELIS 2: Users Reference Guide*. Hillsdale, NJ: Lawrence Erlbaum Associates; 1996.
32. Mardia KV. Measures of multivariate skewness and kurtosis with applications. *Biometrika*. 1970;57:519–530.
33. Joreskog KG. Testing structural equation models. In: Bollen KA, Long JS, eds. *Testing Structural Equation Models*. Newbury Park, Calif: Sage Publishers; 1993.
34. Byrne BM. *Structural Equation Modeling with LISREL, PRELIS, and SIMPLIS: Basic Concepts, Applications, and Programming*. Mahwah, NJ: Lawrence Erlbaum Associates; 1998.
35. Browne MW, Cudeck R. Single sample cross-validation indices for covariance structures. *Multivar Behav Res*. 1989;24(4):445–455.
36. Hu LT, Bentler PM. Fit indices in covariance structure modeling: sensitivity to underparameterized model misspecification. *Psychol Methods*. 1998;3:424–453.
37. Bentler PM, Bonett DG. Significance tests and goodness of fit in the analysis of covariance-structures. *Psychol Bull*. 1980;88:588–606.
38. Mcdonald RP, Marsh HW. Choosing a multivariate model - noncentrality and goodness of fit. *Psychol Bull*. 1990;107:247–255.
39. Satorra A, Bentler PM. A scaled difference chi-square test statistic for moment structure analysis. *Psychometrika*. 2001;66:507–514.
40. Cheung GW, Rensvold RB. Evaluating goodness-of-fit indexes for testing measurement invariance. *Struct Equat Model*. 2002;9:233–255.
41. Knight RG, Williams S, McGee R, Olaman S. Psychometric properties of the Center for Epidemiologic Studies Depression Scale (CES-D) in a sample of women in middle life. *Behav Res Ther*. 1997;35:373–380.
42. Sheehan TJ, Fifield J, Reisine S, Tennan H. The measurement structure of the Center for Epidemiologic Studies Depression Scale. *J Person Assess*. 1995;64:507–521.
43. Anderson JC, Gerbing DW. Structural equation modeling in practice: a review and recommended 2-step approach. *Psychol Bull*. 1988;103:411–423.
44. Chen F, Bollen K, Paxton P, Curran PJ, Kirby JB. Improper solutions in structural equation models: causes, consequences, and strategies. *Sociol Methods Res*. 2001;29:468–508.
45. Van Driel OP. On various causes of improper solutions in maximum likelihood factor analysis. *Psychometrika*. 1978;43:225–243.
46. Roberts RE. Reliability of the CES-D Scale in different ethnic contexts. *Psychiatry Res*. 1980;2:125–134.
47. Blazer DG, Landerman LR, Hays JC, Simonsick EM, Saunders WB. Symptoms of depression among community-dwelling elderly African Americans and White older adults. *Psychol Med*. 1998;28:1311–1320.
48. Tran TV. Exploring the equivalence of factor structure in a measure of depression between Black and White women: measurement issues in comparative research. *Res Soc Work*. 1997;7:500–517.
49. Posner SF, Stewart AL, Marin G, Perez-Stable EJ. Factor variability of the Center for Epidemiological Studies Depression Scale (CES-D) among urban Latinos. *Ethn Health*. 2001;6(2):137–144.
50. Gerbing DW, Anderson JC. Improper solutions in the analysis of covariance structures: their interpretability and a comparison of alternate re-specifications. *Psychometrika*. 1987;52:99–111.
51. Centers for Disease Control and Prevention (CDC). Self-reported frequent mental distress among adults–United States, 1993–2001. *MMWR*. 2004;53(41):963–966.
52. Dunlop DD, Song J, Lyons JS, Manheim LM, Chang RW. Racial/ethnic differences in rates of depression among preretirement adults. *Am J Public Health*. 2003;93:1945–1952.
53. US Department of Health and Human Services. *Mental Health: Culture, Race, and Ethnicity. A Supplement to Mental Health: A Report to the Surgeon General*. Rockville, Md: US Dept of Health and Human Services; 2001.

AUTHOR CONTRIBUTIONS

Design and concept of study: Dishman, Pfeiffer, Pate, Hales
Acquisition of data: Addy, Pfeiffer, Pate
Data analysis and interpretation: Hales, Dishman, Motl, Addy, Pate
Manuscript draft: Hales, Dishman, Motl, Addy, Pfeiffer
Statistical expertise: Hales, Dishman, Motl, Addy
Acquisition of funding: Pate
Administrative, technical, or material assistance: Dishman, Addy, Pfeiffer, Pate
Supervision: Dishman, Pate