CROSS-ETHNIC MEASUREMENT INVARAINCE OF THE SCARED AND CES-D IN AN ETHNICALLY DIVERSE YOUTH SAMPLE

By

LAURA C. SKRINER

A thesis submitted to the Graduate School-New Brunswick

Rutgers, The State University of New Jersey

In partial fulfillment of the requirements

For the degree of

Master of Science

Graduate Program in Psychology

Written under the direction of

Brian C. Chu, Ph.D.

and approved by

_____

_____

_____

_____

New Brunswick, New Jersey

January, 2013

ABSTRACT OF THE THESIS

Cross-Ethnic Measurement Invariance of the SCARED and CES-D in an Ethnically Diverse Youth Sample

By LAURA C. SKRINER

Thesis Director:
Lee Jussim

This study evaluated the cross-ethnic measurement invariance of two common screening measures of anxiety and depressive symptoms in youth. The measurement invariance of the Screen for Childhood Anxiety and Related Emotional Disorders (SCARED) and the Center for Epidemiologic Studies Depression Scale (CES-D) was tested across 881 African American (*n* = 396), Latino (*n* = 185), Non-Hispanic White (*n* = 166), and Asian/Indian (*n* = 134) youths in the seventh grade. The measures were administered as part of a grade-wide screening aimed to identify youth with elevated anxiety and depressive symptoms. The five-factor model of the SCARED and the four-factor model of the CESD best represented the data for all ethnic groups. Results provided support for scalar equivalence of the SCARED across all four ethnic groups. Results provided support for scalar equivalence of the CES-D across African American (AA), Non-Hispanic White (NHW), and Asian/Indian youths, and partial invariance for Hispanic youths. Factor mean differences across groups were identified for both measures. In particular, Hispanic youth reported greater levels of anxiety and depression in certain domains than NHW, AA, and Asian youth. Our findings support the use of the SCARED and CESD as tools for measuring anxiety and depressive symptoms in ethnically diverse youth in the U.S. Results are

discussed with respect to the importance of establishing measurement invariance for screening measures of anxiety and depressive symptoms prior to comparing symptom levels across ethnic.

Table of Contents

PAGE

ABSTRACT ii

LIST OF TABLES x

SECTIONS

1 INTRODUCTION 1

Literature Review and Background 1

The Present Study 5

2 METHODS 7

Participants 7

Procedures 7

Measures 8

Analytic Plan 9

Power 13

3 RESULTS 13

SCARED 13

CES-D 19

4 DISCUSSION 22

Limitations 27

Summary and Implications 28

REFERENCES 30

List of Tables

PAGE

Table 1. Reliability of SCARED and CESD Full and Subscale Scores (Chronbach's Alpha).....9

Table 2a. Fit Indices for CFA's testing for configural invariance of 5-factor model of the SCARED.....14

Table 2b. Fit Indices for CFA's testing for configural invariance of 4-factor model of the CESD 14

Table 3a. Fit indices for multiple group CFAs evaluating level of measurement invariance for the SCARED 15

Table 3b. Fit indices for multiple group CFAs evaluating level of measurement invariance for the CESD 15

Table 4. Standardized Factor loadings from CFA showing scalar invariance for the SCARED 17

Table 5. Factor loadings from CFA showing scalar invariance for the CESD 21

Introduction

Literature Review and Background

Anxiety and depressive disorders are among the most prevalent psychiatric conditions affecting children and adolescents (Costello, Mustillo, Erkanli, Keeler, & Angold, 2003) with lifetime prevalence rates of up to 29% for anxiety and 25% for depressive disorders by the end of adolescence. Although anxiety disorders often develop earlier than episodes of depression, there is substantial overlap between anxiety and depressive symptoms and anxiety during youth is a risk factor for the development of depressive disorders (Stark & Laurent, 2001; Angold & Costello, 1993). Screening of anxiety and depressive symptoms simultaneously is important in order to identify youth who may benefit from preventative intervention or require treatment.

Questionnaires are commonly used and offer an efficient method for screening anxious youths across diverse settings. Accurate and reliable screening is complicated by potential cross-ethnic and cross-cultural differences in response to screening measures. However, the majority of measures have been developed, normed, and validated using majority (NHW) samples. As a consequence, we have limited knowledge of how well these measures assess anxiety and depression in other ethnic groups.

Available data comparing multiethnic with NHW youths on prevalence rates and symptom presentation of anxiety is limited and results are inconsistent. For example, some studies have reported Latino youth to be more likely to meet criteria for separation anxiety disorder (Ginsburg & Silverman, 1996) and show elevated somatic symptoms of anxiety (e.g., Varela, Vernberg, Sanchez-Sosa, Riveros, Mitchell, & Mashunkashey,

2004) relative to youth from other ethnic groups. Several studies suggest AA youths' experience of somatic and panic symptoms may be greater and levels of social and school anxiety may be lower as compared to NHW youths (e.g., Kingery, Ginsburg, Alfano, 2007; Neal & Ward Brown, 1994). In contrast, Ferrell, Beidel, and Turner (2004) found few differences in the clinical presentation of social phobia in AA and NHW youth. Of the few available studies, Asian American college students have been found to report higher levels of social anxiety (e.g., Okazarki, 1997) and youth to exhibit greater rates of test anxiety (Pang, 1991) compared to their NHW counterparts.

Regarding depression, some studies have indicated no differences in rates of depression among several racial and ethnic minority youth (Sagrestano, Paikoff, Holmbeck, & Fendrich, 2003; Siegel, Aneshensel, Taub, Cantwell, & Driscoll, 1998), some have shown lower rates of depression in ethnic minority youth (Allen & Mitchell, 1998; Nettles & Pleck, 1994), and others have shown higher levels of depressive symptoms among AA, Latino, and Asian American youth compared with their NHW counterparts (Okazaki, 1997; Perreira, Deeb-Sossa, Harris, & Bollen, 2005; Roberts, Roberts, & Chen, 1997,). There is research to suggest that depression may manifest in ethnic minority groups somewhat differently compared to NHWs. For instance, AAs, Latinos, and Chinese Americans specifically may experience and report somatic rather than psychological or emotional symptoms (Choi, 2002; Brown, Schulberg, & Madonia, 1996).

In order to accurately identify and interpret mean differences, the *measurement invariance* of questionnaires needs to be examined to determine whether observed differences are based on psychometrically sound assessment tools or may be artifacts of

measurement bias. If measures are non-invariant across ethnic groups, prevalence estimates could be inaccurate (Vandenberg & Lance, 2000) and standard cutoff scores and group comparisons could be misleading (Hui & Triandis, 1985). There are several possible reasons that groups may systematically differ on a given measure. Different ethnic groups may conceptualize a construct differently, use varying symptoms to identify it, and may interpret the possible responses or response scale differently (Crockett, Randall, Shen, Russell, and Driscoll, 2005). Cultural norms may be associated with differences in stigma and attitudes towards mental health (U.S. Department of Health and Human Services, 2001) and might influence whether certain items are endorsed that could lead to psychological labels. For example, there is evidence that Korean Americans are less likely to endorse positive items on depression measures than White or Latino Americans (Jang, Kwag, & Chirigoba, 2010). Such response bias could lead to differences in scores unrelated to the construct of interest. Even if total scores are similar between groups, if the factor structure underlying the construct is divergent across groups, obtaining the same total score in one group may not bear the same meaning in another group. For example, research suggests that Latinos tend to report more somatic symptoms related to internalizing problems than NHW (U.S. Department of Health and Human Services, 2001). If Latino's conceptualize internalizing problems differently, an instrument normed on NHWs could fail to accurately capture relevant symptoms in Latinos. Despite the possibility of measurement nonequivalence, relatively little research has addressed this issue for anxiety and depression measures among ethnic minority youth. Yet, accurate recognition and early detection of youth disorders depend on reliable and validated procedures sensitive to differences across populations.

The Screen for Child Anxiety Related Emotional Disorders (SCARED; Birmaher et al., 1999) is a 41-item measure developed to screen DSM-IV-TR youth anxiety disorders. Initial validation data of the SCARED supported the five proposed subscales: General Anxiety (GAD), Somatic symptoms/Panic (SS), Separation Anxiety (SAD), Social Phobia (SoP), and School Phobia (Sch). The SCARED has demonstrated good psychometric properties across diverse settings and clinical utility as a screening tool (Hale, Crocetti, Raaijmakers, and Meeus, 2011). The five-factor structure has been largely replicated, however, there is evidence that the factor structure may vary across ethnic groups. In a study of 111 AA high school students, Boyd, Ginsburg, Lambert, Cooley, & Campbell (2003) failed to confirm the five-factor structure and follow-up exploratory factor analysis (EFA) identified only three factors, suggesting that the underlying structure of anxiety may vary between AA and NHW youths. Gonzalez, Weersing, Warnich, Scahill, and Woolstone (2012) examined measurement invariance between AA and NHW using CFA and although the five-factor structure was replicated, factor loadings were not equivalent. In a study of 515 multiethnic youth (ages 8-12) presenting to a primary care office, Wren et al., (2007) reported results from EFA showing differences in factor structure for Hispanic youth compared to the rest of the sample. However, Wren et al. did not test for measurement invariance (i.e. did not conduct CFA) making it difficult to interpret the results.

The Center for Epidemiologic Studies – Depression Scale (CES-D; Radloff, 1977) a 20-item measure developed to assess affective and somatic symptoms of depression. Initial EFA of the CES-D in a sample of NHW and AA adults yielded a four-factor structure corresponding to Negative Affect (NA), Positive Affect (PA), Somatic

Symptoms (SS), and Interpersonal Symptoms (IP). Although initially developed for adults, the CES-D has been widely used with youth in both clinical and research settings to screen for risk of depression and to track depressive symptoms across time and throughout treatment interventions and has shown psychometric properties similar to those reported with adults (Roberts, Andrews, Lewinsohn, & Hops, 1990; Prescott et al., 1998).

The four-factor solution has been largely replicated in NHW adolescents although some studies have supported three- or two-factor models often with lack of differentiation between the NA and SS subscales (e.g. Crocket et al., 2005; Edman, 1999). A handful of studies have addressed the issue of measurement invariance for ethnic/racial minority groups in the U.S. using CFA procedures. Russell, Crockett, Shen, and Lee (2008) reported evidence of partial invariance for NHW and Filipino adolescents but the four-factor structure was not supported for Chinese Americans. Crockett and colleagues (2005) reported invariance of factor loadings between NHW and Mexican American youth but non-invariance for European-, Cuban-, and Puerto Rican-American youth. In the only study to our knowledge that tested the structure of the CES-D across multiple major ethnic groups, Perreira and colleagues (2005) reported non-invariance for AAs, Hispanics, and Asians. However, the authors did not fully describe results of tests for partial invariance and it is unclear whether the four-factor model may have shown partial invariance for some of the ethnic subgroups.

The Present Study

In sum, there has been only one study investigating the measurement invariance of the SCARED and few studies testing invariance of the CES-D among multiethnic

youth samples. The primary goal of the present study was to examine the cross-ethnic structure and validity of the SCARED and CES-D for NHW, Latino, AA, and Asian American youth in middle school by testing for measurement invariance. A secondary goal was to compare ethnic groups on symptom cluster (factor) means if invariance was established. This investigation adds to existing knowledge regarding these measures in several ways. First, ours is the first study to examine measurement invariance of the SCARED across several major ethnic groups and the second to do so for the CES-D. Second, we make methodological advances compared with previous studies by using an estimator that takes into account the categorical nature of the response scales for these measures. Estimators based on normal theory have yielded biased indices of model fit and standard errors of parameter estimates when used with ordered-categorical variables (West, Finch, & Curran, 1995; Muthen & Kaplan, 1985). The majority of previous factor analytic studies of the SCARED and CES-D did not account for the categorical nature of items and as a result, findings may be misleading. We are aware of only one study (Motl, Dishman, Birnbaum, Lytle, 2005) that used an estimator that accounted for the categorical nature of the items. Motl et al. (2005) tested for invariance of the CES-D in a predominantly NHW sample of adolescents. We extend the findings of Motl and colleagues by using similar methodology in a multiethnic sample. Third, we examined measures of anxiety and depression in the same multiethnic sample whereas previous studies tend to focus on one or the other.

Methods

Participants

Participants were 881 7th-grade youth (45% female; age range: 11 – 14; *M* = 12.5, *SD* = .44) at a suburban/urban middle school in the northeast United States. The middle school enrolls approximately 1100 economically and ethnically diverse 7th and 8th grade students where 32% of youth eligible for free or reduced lunch. Of participating youth, 396 (44.9%) self-identified as African American, 185 (21%) Hispanic, 166 (18.8%) White/Non-Hispanic, and 134 (15.2%) Asian/Pacific Islander/Indian. Ethnic identity was collected as part of the school's routine enrollment procedures; only broad ethnic categories were collected.

Procedures

Youth participated in a grade-wide screening aimed at identifying anxiety and depression symptoms. Screenings occurred in two waves over two consecutive years. All English-speaking students in the 7th grade whose parents provided consent and who attended school the day of the screening were eligible to participate. Students who were enrolled in English as a second language and students who require significant academic assistance (i.e. receiving special education for the majority of their classes) due to learning disabilities were not eligible.

In year one, 427 of 497 eligible youth completed the screening (45 opted out, 25 agreed but were absent). In year two, 454 of 524 eligible youth completed the screening (39 opted out, 31 agreed but were absent). Participants completed a 35-minute battery of questionnaires, including the two for this study, during a gym or health class period. Students were handed questionnaires and recorded their answers on a Scantron card.

Verbal instructions were provided to complement the written instructions, and research assistants read each question verbatim to enhance verbal comprehension and minimize reading difficulties. Additional research assistants were available to answer student questions during the assessment.

Measures

Screen for Child Anxiety Related Emotional Disorders (SCARED; Birmaher et al., 1999; Birmaher et al., 1997). The SCARED is a 41-item self-report measure with each item rated for severity of anxiety symptoms in the past three months on a 0 "Not True or Hardly Ever True" to 2 "Very True or Often True" scale. Subscales and total score are calculated by summing item responses (total score range 0-82). Total scores above 25 are considered indicative of an anxiety disorder. Reliability estimates in this sample are presented in Table 1. Previous research indicates strong discriminant and convergent validity (for review see Muris, Merkelbach, Ollendick, King, & Bogie, 2002).

Center for Epidemiologic Studies-Depression Scale (CES-D; Radloff, 1977). The CES-D includes 20 items designed to assess depressed mood, feelings of worthlessness/guilt, sense of help/hopelessness, psychomotor retardation, loss of appetite, and sleep disturbance. Respondents indicate the frequency with which symptoms have been experienced during the past week on a 0 "Rarely (less than 1 day)" to 3 "Most of the time (5–7 days)" scale. The four PA items are reverse-scored, and a composite is calculated by summing item responses (range 0–60). Total scores ≥ 21 predict significant presence of depression over the next year (Rushton, Forcier, Schectman, 2002). Reliability estimates in this sample are presented in Table 1.

Table 1.
Reliability of SCARED and CESD Full and Subscale Scores (Chronbach's Alpha)

Scale (# of items)	Full Sample	NHW	AA	Hisp	Asian
SCARED					
Total Scale (41)	**.92**	**.90**	**.92**	**.93**	**.90**
Generalized Anxiety (9)	.77	.71	.77	.80	.78
Separation Anxiety (8)	.73	.70	.73	.76	.67
Panic/Somatic (13)	.82	.76	.82	.83	.83
Social Phobia (7)	.79	.80	.80	.79	.70
School Phobia (4)	.65	.71	.63	.63	.61
CESD					
Total Scale (20)	**.84**	**.83**	**.84**	**.83**	**.83**
Depressed Affect (7)	.77	.77	.84	.82	.77
Positive Affect (4)	.61	.72	.60	.70	.60
Somatic Symptoms (7)	.56	.55	.55	.55	.56
Interpersonal Problems (2)	.60	.64	.72	.71	.60

Note. NHW = non-Hispanic White; AA = African American; Hisp = Hispanic.

Analytic Plan

Single and multiple group CFA models were estimated to assess cross-ethnic measurement invariance of the SCARED and CES-D between NHW, AA, Hispanic (Hisp), and Asian youths. Data were ordered categorical given the Likert-scale nature of the response variables. One approach to estimation with categorical variables is to ignore the metric and treat the data as continuous but nonnormal and adjust the standard errors and chi-square. However, when the number of categories is less than five and data are highly skewed, as in the current study, it is recommended that methodology accounting for the ordinal nature of the data be used (for summary see Byrne, 2012). Rather than maximum likelihood, a robust weighted least squares estimator was used (Brown, 2006) which requires a different likelihood ratio test than the typical chi-square to be computed for evaluating nested models. Models were analyzed using Mplus version 6.11 (Muthén & Muthén, 1998-2011).

As with maximum likelihood, all available data were used in the estimation without pairwise or listwise deletion. Missing data was minimal for both the SCARED and CES-D. For the SCARED, 90.4% (n = 796) of the 881 participants had complete data, 7.6% cases were missing one to two of the 41 items, and the remaining 2.0% of cases were missing three to eight items. The largest proportion of missing data for any one item on the SCARED was 1.1%. For the CES-D, 88.2% (n = 777) of participants had complete data, 9.5% were missing one to two of the 20 items, and 2.3% were missing between three and nine items. The largest proportion of missing data for any one item on the CES-D was 1.9%. Although we employed estimation methods robust to non-normal data, we evaluated uni- and multi-variate normality and linearity for both measures. All items of the SCARED were positively skewed. This was expected as participants were from a non-clinical sample. Similarly, all but one item of the CES-D showed positive skew; item 7, "I felt that everything I did was an effort," showed kurtosis, with similar proportions of youth endorsing each of the four answer choices. Based on this finding and issues with this item during initial CFA models (detailed later), item 7 was ultimately removed from all analyses.

Analyses proceeded in multiple steps following the outline from van de Schoot, Lugtig, and Hox (2012). The first step is to specify a CFA that reflects how the construct is theoretically operationalized and to fit this model for each group separately to test for *configural invariance* (i.e., whether the same CFA is valid in each group). For the SCARED, this model consisted of five latent factors (the subscales) each indicated by their corresponding observed variables (41 items). We also tested a single-factor model for each group and subsequently compared the single- with the five-factor model for each

ethnic group (to replicate performance of a total score and test for unidimensionality). For the CES-D, the configural model consisted of four latent factors (the subscales) each indicated by their corresponding observed variables (20 items). As with the SCARED, single-factor models were tested and compared with the four-factor models for each ethnic group.

If configural invariance is found, measurement invariance across ethnic groups can be tested using a set of multiple group CFAs (Millsap & Yun-Tein, 2004; Vandenberg & Lance, 2000; van de Schoot et al., 2012). In the first step, *factor loadings* are constrained equal across groups but intercepts are allowed to differ (*metric invariance*). This tests whether respondents across groups attribute the same meaning to the latent construct (e.g., general anxiety, negative affect) under study. The second step is to run a model where only *intercepts (thresholds)* are constrained equal, but factor loadings are allowed to differ; this tests whether the meaning of the levels of underlying items (intercepts) are equal in both groups. One point must be clarified. With categorical indicators, *thresholds* (or initial scale points) are the focus of analyses instead of indicator intercepts (which are used with continuous variables; Byrne, 2012). Thresholds correspond to the underlying probability distribution of the categories rather than the observed values. The third step is to run a model where the loadings *and* intercepts are constrained to be equal, which tests for full *scalar* (or measurement) invariance. Scalar invariance implies that the meaning of the construct (the factor loadings), and the levels of the underlying items (intercepts) are equal in all groups (Muthén & Asparouhov, 2002). Consequently, groups can be compared on their scores on the latent variable. Chi-square difference tests and several fit indices were used to compare the relative fit of

models. If there is evidence of non-invariant measurement parameters (e.g., unequal factor loadings) at any step in the process, tests of invariance will proceed in the context of partial measurement invariance. Although partial invariance models are considered exploratory, when a small number of parameters are non-invariant and the criteria for partial invariance are met, between-group comparisons of the latent means can still be conducted (Byrne, Shavelson, & Muthén, 1989).

Model fit is often indicated by a nonsignificant χ^2 test, and comparisons between nested models often evaluated with the chi-square difference test ($\Delta \chi^2$). Because the χ^2 and $\Delta \chi^2$ are strongly influenced by sample size, their use is impractical for these data (Dimitrov, 2010) and for tests of measurement invariance (Cheung & Rensvold, 2002). Accordingly, we relied on descriptive fit indices including the comparative fit index (CFI), Tucker-Lewis Index (TLI), and root-mean-square error of approximation (RMSEA) to assess model fit. Fit is considered adequate if the CFI and TLI values are > .90, and better if they are > .95. For the RMSEA, values < .05 indicate good fit, whereas values within the range of 0.05 to 0.08 indicate adequate or fair model fit (Hu & Bentler, 1999). Comparisons among competing models relied on the fit indices listed above and followed recommendations of Cheng and Rensvold (1999). Specifically, a change of 0.01 or less in the CFI or TLI, a change in RMSEA that falls within the 90% confidence interval of the most recent and less constrained model, and acceptable fit of the most recent model indicate measurement equivalence (i.e. supports the more constrained model). We also considered the Akaike Information Criterion (AIC; Akaike, 1987) and Bayesian Information Criterion (BIC; Raftery, 1993) for model comparison, with the smallest value representing the best fit of the hypothesized model.

Power

Various recommendations for the minimum sample size necessary for conducting factor analysis have been proposed. Recommendations fall into one of two categories: one based on the absolute number of cases in the sample and the other based on the subject-to-variable (STV) ratio. Although there is no consensus regarding a minimum for absolute sample size, several researchers have recommended using no less than a sample of 100 (e.g., Gorsuch, 1983; Kline, 1979). Recommendations for STV ratio typically range from 2:1 to 10:1, with 5:1 cited as one of the most frequently recommended ratios (e.g., Bryant and Yarnold, 1995; Gorsuch, 1983, Kline, 1979). Based on these recommendations, factor analyses for each ethnic group appear to be adequately powered. Absolute sample sizes range from 132 to 393, and STV ratios for each ethnic group range from 6.6 to 19.65 for the CES-D and from 3.23 to 9.58 for the SCARED.

Results

SCARED

To test for configural invariance, CFA models were fit separately for each group. A model where all 41 items loaded on one factor (one-factor model) was compared with a model where the 41 items loaded on the original five-factor model in each group (see Table 2a). In all groups, the five-factor model showed adequate to good fit to the data (CFI and TLI ranged from .91 to .96 and RMSEA were all under .04), and showed superior fit over the one-factor model. Further, all factor loadings were moderate to large, in the expected direction, and statistically significant for all four groups. Factor intercorrelations were positive and nearly all significant for all groups; the intercorrelation between the Sch and SoP factors was nonsignificant for Asians. These

results suggest a common five-factor structure of the SCARED in all groups and support configural invariance.

Table 2a.
Fit indices for CFAs testing for configural invariance of 5 factor model of SCARED

	NHW		AA		Hisp		Asian	
	1-factor	5-factor	1-factor	5-factor	1-factor	5-factor	1-factor	5-factor
CFI	0.83	0.91	0.91	0.96	0.90	0.96	0.82	0.91
TLI	0.82	0.91	0.90	0.95	0.88	0.96	0.82	0.90
RMSEA	0.05	0.04	0.04	0.03	0.05	0.03	0.05	0.04

Table 2b.
Fit Indices for CFAs testing for configural invariance of 4-factor model of the CESD

	1-factor	4-factor	1-factor	4-factor	1-factor	4-factor	1-factor	4-factor
CFI	0.91	0.97	0.91	0.98	0.87	0.95	0.94	0.97
TLI	0.90	0.96	0.89	0.97	0.85	0.94	0.93	0.97
RMSEA	0.70	0.04	0.70	0.04	0.09	0.06	0.50	0.04

Note. CFA = Confirmatory factor analyses; NHW = non-Hispanic White; AA = African American; Hisp = Hispanic; CFI = Comparative Fit Index; TLI = Tucker-Lewis Index; RMSEA = Root Mean Square Error of Approximation.

Next, we used multiple group CFA to test for measurement invariance. Results are summarized in Table 3a. We first constrained factor loadings to be equal but allowed the thresholds to differ between groups (Model 1). This more restrictive metric invariance model was a good fit to the data. The second model (Model 2), where only the intercepts were set to be equal and factor loadings were unconstrained across groups (i.e. allowed to differ), also showed good fit to the data. Finally, a model with both factor loadings and thresholds constrained to be equal across all groups was estimated (Model 3). Model 3 also showed good fit to the data.

Table 3a.
Fit indices for multiple group CFAs evaluating level of measurement invariance for the SCARED

Model (constrained parameters)	$\Delta\chi^{2a}$	*df*	CFI	TLI	RMSEA (95% CI)	AIC	BIC
Model 1(factor loadings)	384.20*	246	.947	.940	.031 (.027-.035)	56828.09	58654.45
Model 2 (thresholds)	195.84*	123	.943	.940	.032 (.028-.035)	56936.50	58174.79
Model 3 (factor loadings and thresholds)			.942	.942	.031 (.028-.035)	56832.49	57482.72

Table 3b.
Fit indices for multiple group CFAs evaluating level of measurement invariance for the CESD

Model (constrained parameters)	$\Delta\chi^{2a}$	*df*	CFI	TLI	RMSEA (95% CI)	AIC	BIC
Model 1(factor loadings)	175.79	164	.936	.926	.059 (.05-.037)	32910.78	34115.89
Model 2 (thresholds)	120.32*	54	.959	.960	.044 (.04-.05)	32752.04	33431.11
Model 3 (factor loadings and thresholds)			.946	.951	.048 (.04-.05)	32767.60	33174.08
Model 4 (Model 3 modified)[b]			.951	.956	.046 (.04-.05)	32764.64	33185.48

Note. CFI = comparative fit index; TLI = Tucker-Lewis Index; RMSEA = root-mean-square error of approximation; AIC = Akaike Information Criterion; BIC = Bayesian Information Criterion.

[a] Likelihood ratio test in which nonsignifcant supports invariance. Note that given the estimator used for analysis with the categorical indicators (weighted least square with mean and variance correction estimation), the chi-square must be corrected because the difference in chi-square values is no longer chi-square.

[b] Model 4 is identical to Model 3 except that in Model 4 the factor loadings for items 9, 10, and 13 were allowed to be freely estimated for Hispanic group due to evidence of non-invariance.

* $p < .0001$

Model fit for the two larger, less restricted models (Model 1 and 2) were then compared with model fit for the most restrictive model (Model 3). Comparison of Model

1 with Model 3 yielded a significant likelihood ratio test (labeled $\Delta\chi^2$ in Table 3) indicating that Model 1 is a better fit to the data than Model 3. However, as mentioned previously, the likelihood ratio test is impractical for evaluating model fit with large samples and when multiple constraints are placed on the model. When looking at fit indices Model 3 appears to be the better fitting model; ΔCFI was < .01, which supports the more restricted Model 3, TLI is slighter larger for Model 3 and BIC slightly lower, and RMSEA is identical between the two models. AIC is the only fit index that supports Model 1. Given that Model 3 is shown to fit the data better or equally as well on four of five fit indices and is the more parsimonious model, Model 3 is the superior model. Comparison of Model 2 with Model 3 also yielded a significant likelihood ratio test, suggesting that Model 2 is a better fit to the data. However, Δ CFI was < .01, TLI was slightly larger and RMSEA slightly lower for Model 3, AIC and BIC values for Model 3 were lower for Model 3 compared to Model 2. These results indicate that Model 3 fits the data better than Model 2. Together, these findings indicate that among all three models, Model 3 – the most restrictive model – offers the best trade-off between model fit and model complexity, supporting full scalar invariance of the SCARED across all four ethnic groups. Factor loadings for this model are presented in Table 4. Consistent with the configural models, all factor loadings and intercorrelations for this final model (Model 3) were significant with the exception of the intercorrelation between the Sch and SoP factors for Asians. Factor intercorrelations involving the SoP factor tended to be somewhat smaller for the Asian group and those involving the SAD factor tended to be somewhat smaller or the NWH compared with the other three groups.

Table 4.
Standardized Factor loadings from CFA showing scalar invariance for the SCARED

Variable	NHW	AA	Hispanic	Asian
General Anxiety				
5 Others like me	.65	.58	.66	.56
7 Nervous	.67	.65	.54	.73
14 As good as others	.50	.71	.66	.75
21 Things work out	.47	.54	.53	.64
23 Worrier	.56	.55	.71	.65
28 Others tell me I worry	.53	.58	.63	.49
33 Future	.65	.60	.62	.63
35 How well do things	.61	.56	.56	.52
37 Things in past	.63	.69	.71	.73
Separation Anxiety				
4 Sleep away from home	.70	.60	.57	.57
8 Follow mother	.40	.38	.29	.43
13 Worry sleeping alone	.75	.74	.69	.58
16 Nightmares about parents	.66	.65	.66	.73
20 Nightmares about me	.76	.74	.78	.82
25 Afraid alone in house	.68	.67	.50	.75
29 Away from family	.56	.50	.55	.41
31 Worry parents	.72	.67	.58	.68
Panic / Somatic				
1 Hard to breathe	.48	.59	.56	.53
6 Feel like passing out	.83	.70	.65	.75
9 Look nervous	.57	.56	.54	.57
12 Going crazy	.65	.73	.76	.71
15 Thing not real	.71	.63	.67	.58
18 Heart beats fast	.66	.51	.55	.57
19 Shaky	.67	.68	.62	.69
22 Sweat a lot	.44	.57	.54	.46
24 Frightened no reason	.81	.75	.66	.77
27 Choking	.76	.70	.74	.69
30 Panic attacks	.73	.55	.52	.71
34 Throwing up	.66	.69	.68	.78
38 Dizzy	.62	.65	.77	.76
Social Phobia				
3 Don't like to be with new people	.60	.52	.41	.59
10 Nervous with new people	.73	.67	.62	.61
26 Hard to talk to new people	.70	.70	.70	.70
32 Shy with new people	.80	.77	.83	.77
39 People watching me	.71	.73	.70	.69
40 Nervous in places with new people	.79	.72	.69	.52
41 Shy	.65	.66	.74	.58
School Phobia				
2 Headaches at school	.64	.51	.39	.51
11 Stomachaches at school	.62	.55	.54	.84
17 Worry about school	.84	.79	.72	.79
36 Nervous about school	.96	.92	.89	.84

Note: CFA = Confirmatory factor analysis; NHW = non-Hispanic White; AA = African American
**Unstandardized values for invariant loadings across all four groups are all equivalent to the standardized values listed for the non-Hispanic White group.

With the measurement invariance of the SCARED five-factor modeled established, ethnic groups could now be compared on latent factor means. Analyses involved building on the multiple group CFA identified as the best fitting model in the previous analyses (Model 3). In a series of models, latent factor means for the four ethnic groups were compared by fixing one latent group mean to zero while allowing the means for the other three groups to be freely estimated (Byrne, 2012). The group with the mean fixed at zero served as the reference group against which the other groups were compared. Factor mean estimates and mean differences are z-scores. Results showed several significant (p-values < .05) group differences in means of the SAD, SS, and Sch factors, whereas there were no differences among the four ethnic groups on the GAD and SoP factor means. Specifically, Hisp youth appear to report higher levels of SAD than NHW (ΔM = .38), AA (ΔM = .29), and Asian (ΔM = .35) youth; Hisp youth also report higher levels of School Phobia than NHW (ΔM = .27), AA (ΔM = .15), and Asian (ΔM = .36); AA and Hisp youths report higher levels of SS than NHW (ΔM = .20; ΔM = .24, respectively); and AA youths report higher School Phobia compared to Asian (ΔM = .20) youth.

Although statistically significant, differences were small and may not all be clinically meaningful. To put these results in some context, 34% of the entire sample scored above the clinical cutoff on the SCARED, with a total-score observed mean of 21.33 (*SD* = 12.86). Looking at each group separately, 27.1% of NHW (total score *M* = 18.72, *SD* = 10.55), 34.4% of AA (total score *M* = 21.56, *SD* = 13.33), 41.8% of Hisp (total score *M* = 24.27, *SD* = 14.10), and 30.8% of Asian (total score *M* = 19.84, *SD* = 11.44) youth scored above the clinical cutoff. For a subscale example, Hisp youth scored

approximately 1.5 points higher on the SAD subscale, on average (M = 5.13, SD = 3.51), than all other groups whose means ranged from 3.35 to 3.96 (SD range = 2.68-3.14).

CES-D

The first set of CFA models testing for configural invariance of the CES-D fit the data well, however, item 7 ("I felt that everything I did was an effort") failed to load significantly on its factor (Somatic Symptoms) across all four ethnic groups. Byrne (2012, p. 78) recommends that non-significant parameters can be considered unimportant to the model; and in the interest of parsimony, should be deleted from the model. Based on Byrne's recommendation together with evidence of kurtosis from initial data screening, item 7 was removed from further analyses and subsequent CFA's were based on the remaining 19 observed indicators. It must be noted that although we continued using a CFA framework, these analyses are considered "exploratory" due to the removal of item 7. Separate one- and four-factor CFA models for each group were fit to the data. Table 2b shows the fit indices for the models. In all groups, the four-factor model showed good fit to the data (CFI and TLI ranged from .94-.98 and RMSEA were all under .06), and showed superior fit over the one-factor model. Further, all factor loadings were moderate to large, in the expected direction, and statistically significant for all four groups. These results suggest a common four-factor structure of the CES-D in all groups and support configural invariance. Factor loadings and factor intercorrelations were positive and significant for all groups.

Next we used multiple group CFA to test for measurement invariance. Results are summarized in Table 3b. Fit indices suggested good fit to the data for the model with only factor loadings constrained to be equal across groups (Model 1; metric invariance),

the model with only thresholds set to be equal (Model 2), and the most restrictive model where both factor loadings and thresholds were constrained to be equal (Model 3). Although Model 3 showed good fit to the data, the intercorrelation between the NA and SS factors in the Hispanic group exceeded one, indicating that while scalar invariance was supported across the NHW, AA, and Asian groups, the model was somewhat misspecified in the Hisp group. Using modification indices as a guide, constraints on CES-D items 9, 10, and 13 were released for the Hisp group, allowing the factor loadings to be freely estimated. This modified model (Model 4) showed good fit to the data.

Model fit for Models 1 and 2 were then compared with model fit for Model 4. Comparison of Model 1 with Model 4 yielded a nonsignificant likelihood ratio test indicating Model 4, the more restrictive model, fit the data better than Model 1. Further, CFI, TLI, and RMSEA favored Model 4 compared to Model 1. Comparison of Model 2 with Model 4 yielded a significant likelihood ratio test; however, review of fit indices support Model 4 as the better fitting model. Specifically, ΔCFI and ΔTLI were less than .01, supporting invariance, and the BIC for Model 4 was lower than for Model 2. These results indicate that among the three models, Model 4 – the most restrictive model – fit the data best. These findings support full scalar invariance among NHW, AA, and Asian groups and partial scalar invariance for the Hisp group. As with the single-group models, loadings and factor intercorrelations in Model 4 were positive and significant for all groups. Standardized loadings from Model 4 are presented in Table 5. Factor intercorrelations were generally moderate to large across groups. Factor intercorrelations were mostly similar across groups. For the Hisp group, the correlations between NA and PA and between SS and PA were somewhat lower and the correlation between NA and

SS was slightly higher than for the other groups. The correlation between NA and IP was slightly higher for both Hisp and AA youth.

Table 5.
Factor loadings from CFA showing scalar invariance for the CESD

Variable	NHW	AA	Hisp	Asian
Depressed Affect				
3 Blues	.75	.67	.63	.78
6 Down	.73	.72	.64	.70
9 Failure	.87	.71	**.79**[a]	.75
10 Fearful	.60	.60	**.73**[a]	.63
14 Lonely	.82	.78	.71	.83
17 Crying	.78	.78	.73	.76
18 Sad	.83	.82	.75	.81
Positive Affect				
4 Good	.60	.45	.46	.46
8 Hopeful	.57	.48	.39	.48
12 Happy	.80	.74	.69	.78
16 Enjoyed	.87	.74	.78	.88
Somatic Symptoms				
1 Bothered	.55	.60	.56	.67
2 Eating	.56	.49	.45	.57
5 Mind	.70	.55	.55	.81
11 Sleep	.64	.56	.50	.57
13 Talked	.45	.37	**.52**[a]	.44
20 Get going	.68	.66	.63	.68
Interpersonal Problems				
19 Unfriendly	.99	.88	.82	.99
15 Disliked	.70	.72	.64	.74

Note. CFA = Confirmatory factor analysis; NHW = non-Hispanic White; AA = African American; Hisp = Hispanic. Factor loadings are from Model 4, where loadings (unbolded) and thresholds are constrained equal across groups; bolded items (9, 10, and 13) were allowed to be freely estimated for Hispanic group due to evidence of non-invariance. Unstandardized values for invariant loadings across all four groups are all equivalent to the standardized values listed for the non-Hispanic White group Unstandardized values are identical across groups because loadings are constrained to be equal in this model.
[a]For Hispanics, unstandardized loadings for item 9 = 1.30; for item 10 = .94; and for item 13 = .94.

Factor means of the four latent variables (NA, PA, SS, and IP) were then compared across ethnic groups. Although three item factor loadings were non-invariant for the Hisp group, valid inferences can still be made about differences in latent factor

means when at least two loadings (and thresholds) are constrained equal (Byrne et al., 1989). In this case, 17 loadings are held equal for the Hisp group. Analyses followed the same procedure used with the SCARED. Whereas there were no group differences in means for the PA and IP factors, there were several significant group differences in means of the NA and SS factors. Hisp youth appear to report slightly higher levels of NA than NHW (ΔM = .34), AA (ΔM = .25), and Asian (ΔM = .42) youths. Hisp youths also report higher levels of SS than NHW (ΔM = .27), AA (ΔM = .20), and Asian (ΔM = .45) youths. Asian youth reported slightly lower levels of SS than AA (ΔM = .25) youth.

To put these results in context, we calculated proportion of youth scoring above the cutoff value of 21 or greater. Item 7 was retained for these analyses in order to use the recommended cutoff, which is based on the 20-item CES-D. Of the entire sample, 19.7% scored above the screening cutoff, with a total-score mean of 13.29 (*SD* = 8.99). Looking at each group separately, 12.8% of NHW (total score *M* = 11.08, *SD* = 7.75), 19.4% of AA (total score *M* = 13.73, *SD* = 9.17), 31.3% of Hispanic (total score *M* = 16.21, *SD* = 9.61), and 13.1% of Asian (total score *M* = 10.69, *SD* = 7.56) youth scored above the screening cutoff. Looking at the observed scores for the SS subscale as a subscale example, Hisp youth had a mean score of 7.13 (*SD* = 3.74) while means for the other ethnic groups were lower ranging from 5.02 to 6.2 (range of *SD*s = 3.23-3.58).

Discussion

The primary goal of the present study was to investigate the cross-ethnic measurement equivalence of two frequently used measures for screening anxiety (the SCARED) and depression (the CES-D) across non-indicated samples of NHW, AA, Hisp, and Asian youths in middle school. A secondary goal was to test for invariance of

the latent mean structures if measurement invariance was established. To this end, we examined configural, metric, and scalar invariance, as well as invariance of factor means for both measures. Results provided support for full scalar equivalence of the SCARED across all four ethnic groups. Results also provided support for full scalar equivalence of the CES-D across NWH, AA, and Asian youths, and partial invariance for Hisp youths. These results show that the multi-dimensional factor structures developed with largely NHW populations are also applicable to other major ethnic groups in the U.S. Factor mean differences across groups were identified for both measures.

The original five-factor structure of the SCARED was replicated in NHW, AA, Hisp, and Asian youths, establishing configural invariance across ethnic groups indicating that symptoms included in the SCARED cluster in the same way across group. We subsequently found evidence for invariance of factor loadings (metric invariance), item thresholds, and finally of both loadings and thresholds simultaneously (full scalar invariance) for the five-factor model across all groups. These results suggest that youth across the four ethnic groups appear to attribute the same meaning to the constructs of GAD, SAD, SS, SoP, and Sch as measured by the SCARED and to interpret the individual items and the scoring scale the same. By establishing full scalar invariance, it is likely that SCARED scores represent the same construct in these groups, indicating that we can compare ethnic groups with confidence, making straightforward interpretations of latent means and correlations. These results differ somewhat from those of Gonzalez et al.'s (2012) investigation of measurement invariance between NHW and AA youth although general conclusions and recommendations for use and interpretation of the SCARED are similar. Gonzalez et al. (2012) did replicate the 5-factor structure

finding support for configural invariance, but reported partial metric invariance between the two groups based on youth report.

Comparison of latent means revealed that Hisp youth experience more anxiety in certain domains than NHW, AA, and Asian youth. Specifically, Hisp youth reported higher levels of SAD and of School Phobia than NHW, AA, and Asian youths, and higher levels of SS than NHW youth. AA youth reported higher levels of SS than NHW and higher levels of School Phobia than Asian youth. These findings are consistent with previous literature (e.g., Varela et al., 2008) suggesting that Hispanic youth are more likely to meet diagnostic criteria for SAD and tend to report more somatic symptoms of anxiety than NHWs. Based on the current results, these tendencies appear to extend to AA and Asian youth. Previous research also suggests that somatic symptoms may be stronger indicators of anxiety for AAs than NHWs (Lambert et al., 2004; Gonzalez et al., 2012), which was also reflected in the current findings. Clinically, increased attention to somatic symptoms in Hisp and AA youth may lead to greater detection of emotional distress in these groups. If screening attempts do not make this calibration, Hisp and AA youth with clinical anxiety may be overlooked. At the same time, because the current sample was non-clinical, future research needs to investigate the level of clinical interference associated with the identified clusters of heightened symptoms. In contrast to reports of Hispanics and AAs experiencing more worry and general anxiety than NHWs (Silverman et al., 1995; Ginsberg & Silverman, 1996; Varela et al., 2004) and of Asian youth reporting more SoP than NHWs (Okazarki, 1997), there were no differences among groups on the GAD or SoP factors.

We then examined the structure of the CES-D and the four-factor structure CES-D was replicated across NHW, AA, Hisp, and Asian youths, establishing configural invariance. However, results suggested that Item 7 ("I felt that everything I did was an effort") did not add information to the model and was thus removed from analyses. We continued with tests of invariance although analyses were now considered exploratory. Full metric invariance was established across the NHW, AA, and Asian groups indicating that these groups attribute the same meaning to the constructs of NA, PA, SS, and IP. For the Hisp group, three of the 20 items had non-invariant factor loadings. Items 9 and 10 from the NA factor and item 13 from the SS factor were non-invariant for Hisp youth, indicating that Hisp youth appear to attribute somewhat different meaning to these items. Similar results with these same items have been reported in previous CFA studies (Crockett et al., 2005; Radloff, 1991). The properties of these items should continue to be evaluated in future research including whether removing them altogether could improve the cross-ethnic validity of the CES-D. Support for invariance of item thresholds was found across all four ethnic groups, suggesting that the meaning of the levels of the underlying items are equal across groups. Finally, when factor loadings and thresholds were simultaneously constrained equal, evidence of full scalar invariance for NHW, AA, and Asian youth, and partial scalar invariance for Hisp youth was found. A lack of full scalar invariance for the Hisp group could increase the risk of classification error when the CES-D is used as a screening tool. However, given that only a small number of factor loadings were non-invariant – three out of 20 – the likelihood and amount of misclassification may be small and differences in factor means can still be meaningfully interpreted (Byrne et al., 1989). Overall, these results suggest we can compare ethnic

groups with confidence, making straightforward interpretations of latent means and correlations.

Previous CFA studies with the CES-D have yielded somewhat varied results. Most studies report evidence supporting the four-factor structure (Crockett et al., 2005; McArdle et al., 2001; Motl et al., 2005; Roberts et al., 1990; & Russell, 2008), but a few others have favored either three-factor (Dick, Beals, Keane, & Mason, 1994) or two-factor models (Crockett et al., 2005; Edman, 1999) in some racial/ethnic groups. Congruent with our findings, Crockett et al. (2005) reported partial metric invariance and partial scalar equivalence between Anglo- and Mexican-American adolescents; however, Crockett et al. reported non-invariance among Cuban- and Puerto-Rican- American youth. Our results are consistent with those of Russell et al. (2008), who reported partial scalar invariance between NHW and Filipino-American adolescents and with Hales et al. (2006) who reported invariance of factor structure and loadings across NHW and AA female adolescents. However, Russell et al. only found scalar invariance for eight of the 20 CES-D items while we found evidence of full scalar invariance across Asian, NHW, and AA youth, and Russell et al.'s results did not support invariance for Chinese-Americans. Importantly, current results are consistent with those of Motl et al. (2005), which to our knowledge is the only other study testing for measurement invariance of the CES-D in adolescents that accounted for the categorical nature of the data. Motl et al. reported evidence in support of Radloff's (1997) four-factor structure as well as invariance of the CES-D longitudinal over nearly a two-year period and across genders. Our results extend those of Motl et al., which had a primarily NHW sample, by establishing scalar invariance across NHW, AA, and Asian youth and partial scalar

invariance for Hisp youth. Based on previous and current results it appears that at the broad level, NHW, AA, Hisp, and Asian youth share a similar frame of reference regarding depression.

Comparison of latent factor means revealed that Hisp youth reported higher levels of NA and SS than NHW, AA, and Asian youth. Asian youth reported lower levels of SS than AA. The higher reported levels of NA and SS in the Hisp group is consistent with previous epidemiological studies suggesting that Hispanic/Latino youth tend to have higher rates of depressive symptoms and disorders than other ethnic groups in the U.S. even when socio-economic status is taken into account (e.g., Roberts et al., 1997; Roberts & Sohban, 1992; Siegel et al., 1997). Results are also consistent with the finding of non-invariance of several items from the NA and SS scales. In contrast to reports of Asian Americans reporting higher levels of depression than their NHW counterparts (e.g., Okazaki 1997), Asian youth in the current sample did not report elevated depressive symptoms compared with NHW, Hisp, or AA youth. There were no significant differences among groups on PA or IP factors.

Limitations

All results must be interpreted in light of methodological limitations. The present study relied on convenience samples, and future studies should test whether these results are generalizable to the general population. Further, the sample represented an ethnically diverse, but still English-speaking, American sample. Requiring English language skills may mute potential differences across culturally diverse and recent immigrant groups. In addition, the current sample was not able to discriminate ethnic groups beyond broad categories and differences in factor structure or potential item bias among ethnic

subgroups may have been obscured in the current study. Given previous research indicating potential differences in the structure of the CES-D in some ethnic subgroups, it is important for future research on these measures to identify youth beyond broad ethnic/racial categories.

Summary and Implications

In sum, our findings support the cross-ethnic validity of the SCARED and CES-D and their use as tools for assessing anxiety and depressive symptoms in ethnically diverse youth in middle school. We have provided evidence of full scalar invariance of the SCARED and of full or partial scalar invariance of a 19-item version of the CES-D across four ethnic groups. In other words our results indicate that NHW, AA, Hisp, and Asian youth interpret questions on the SCARED and on the CES-D similarly and that these measures are assessing the same constructs across these groups. These results bolster confidence in previous and future research utilizing these measures in ethnically/racially diverse samples of youth in the U.S. and indicate that observed differences in mean anxiety and depressive symptoms are not likely due to differential measurement properties across ethnic groups in this sample. Studies testing the measurement equivalence of assessment tools are especially important in light of the field's focus on evidence-based assessment.

Results also indicate that certain groups of youth, Hispanic youth in particular, experience more anxiety and depression in certain domains than NHW, AA, and Asian youth. The mechanisms by which such differences evolve have yet to be identified. Cultural phenomena are one possibility. For instance, it has been proposed that strong family orientation in Hispanic culture may explain greater separation anxiety, and

negative stigma for mental health problems may explain somatic expression of emotional distress (Varela et al., 2004; U.S. Department of Health and Human Services, 2001). Research is needed to investigate such hypotheses. Although the current study did not use an epidemiologically representative sample, the observed latent mean differences identified for both measures are concerning in light of service disparities across ethnicities in youth. Our findings indicate that Hispanic youth report higher levels of certain clusters of symptoms, yet recent research reveals that Hispanic youth are less likely than NHWs to be diagnosed with mood disorders (Nguyen, Huang, Arganza, & Liao, 2007; Mak & Rosenblatt, 2002). A full examination of the potential discrepancies between rates of internalizing problems and actual psychiatric diagnoses and services received is beyond the scope of this study. Still, our findings highlight how valid measures are needed to guide clinical practice and mental health policy. By identifying valid differences in clinical characteristics and rates of psychological problems we can work towards ensuring that ethnic minority youth are accurately diagnosed and receive appropriate care.

References

Akaike, H. (1987). Factor analysis and AIC. *Psychometrika, 52*, 317-332.

Allen, L., & Mitchell, C. (1998). Racial and ethnic differences in patterns of problematic and adaptive development: An epidemiological review. In V. C. McLoyd & L. Steinberg (Eds.), *Minority Adolescents: Conceptual, Methodological, and Theoretical Issues* (pp. 29-54). Mahwah, NJ: Lawrence Erlbaum.

Angold, A., & Costello, E. (1993). Depressive comorbidity in children and adolescents: empirical, theoretical, and methodological issues. *American Journal of Psychiatry, 150*(12), 1779-1791.

Birmaher, B., Brent, D. A., Chiappetta, L., Bridge, J., Monga, S., & Baugher, M. (1999). Psychometric properties of the Screen for Child Anxiety Related Emotional Disorders (SCARED): A replication study. *Journal of the American Academy of Child and Adolescent Psychiatry, 38*(10), 1230-1236.

Birmaher, B., Khetarpal, S., Brent, D., Cully, M., Balach, L., Kaufman, J., & Neer, S. M. (1997). The Screen for Child Anxiety Related Emotional Disorders (SCARED): Scale construction and psychometric characteristics. *Journal of the American Academy of Child & Adolescent Psychiatry, 36*(4), 545-553.

Boyd, R. C., Ginsburg, G. S., Lambert, S. F., Cooley, M. R., & Campbell, K. D. M. (2003). Screen for Child Anxiety Related Emotional Disorders (SCARED): Psychometric properties in an African-American parochial high school sample. *Journal of the American Academy of Child and Adolescent Psychiatry, 42*(10), 1188-1196.

Brown, C., Schulberg, H. C., & Madonia, M. J. (1996). Clinical presentations of major depression by African Americans and whites in primary medical care practice. *Journal of Affective Disorders, 41*(3), 181-191.

Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. Bollen & J. S. Long (Eds.), *Testing Structural Equation Models* (pp. 136-162). United Kingdom: Sage Publications.

Bryant, F. B., & Yarnold, P. R. (1995). Principal components analysis and exploratory and confirmatory factor analysis. In L. G. Grimm & R. R. Yarnold (Eds.), *Reading and understanding multivariate statistics* (pp. 99-136). Washington, D.C.: American Psychological Association.

Byrne, B.M. (2012). Structural equation modeling with Mplus: Basic concepts, applications, and programming. New York, NY: Routledge.

Byrne, B. M., Shavelson, R. J., & Muthen, B. (1989). Testing for the equivalence of factor covariance and mean structures: The issue of partial measurement invariance. *Psychological Bulletin, 105*(3), 456-466.

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal, 9*(2), 233 - 255.

Choi, H. (2002). Understanding adolescent depression in ethnocultural context. *Advances in Nursing Science, 25*(2), 71-85.

Costello, E. J., Mustillo, S., Erkanli, A., Keeler, G., & Angold, A. (2003). Prevalence and development of psychiatric disorders in childhood and adolescence. *Archives of General Psychiatry, 60*(8), 837-844.

Crockett, L. J., Randall, B. A., Shen, Y.L., Russell, S. T., & Driscoll, A. K. (2005). Measurement equivalence of the Center for Epidemiological Studies Depression Scale for Latino and Anglo adolescents: A National Study. *Journal of Consulting and Clinical Psychology, 73*(1), 47-58.

Dick, R.W., Beals, J., Keane, E. M., & Manson, S. M. (1994). Factorial structure of the CES-D among American Indian adolescents. *Journal of Adolescence*, *17*, 73-79.

Dimitrov, D. M. (2010). Testing for factorial invariance in the context of construct validation. Measurement and Evaluation in Counseling and Development, 43, 121-149.

Edman, J. L., Danko, G. P., Andrade, N., McArdle, J. J., Foster, J., & Glipa, J. (1999). Factor structure of the CES-D (Center for Epidemiologic Studies Depression Scale) among Filipino-American adolescents. *Social Psychiatry and Psychiatric Epidemiology, 34*(4), 211-215.

Ferrell, C. B., Beidel, D. C., & Turner, S. M. (2004). Assessment and treatment of socially phobic children: A cross cultural comparison. *Journal of Clinical Child & Adolescent Psychology, 33*(2), 260 - 268.

Ginsburg, G. S., & Silverman W. K. (1996). Phobic and anxiety disorders in Hispanic and Caucasian youth. *Journal of Anxiety Disorders, 10*, 517–528.

Gonzalez, A., Weersing, V. R., Warnich, E., Scahill, L., & Woolston, J. (2012). Cross-ethnic measurement equivalence of the SCARED in an outpatient of African American and non-Hispanic White youths and parents. *Journal of Clinical Child & Adolescent Psychology, 0,* 1-9.

Gorsuch, R. L. (1983). *Factor analysis* (2nd edition ed.). HIllsdale, NJ: Erlbaum.

Hale, W. W. III, Crocetti, E., Raaijmakers, Q. A., & Meeus, W. H. (2011). A meta-analysis of the cross-cultural psychometric properties of the Screen for Child Anxiety Related Emotional Disorders (SCARED). *Journal of Child Psychology and Psychiatry, 52*(1), 80-90.

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

Hui, C. H., & Traindis, H. C. (1985). Measurement in cross-cultural psychology: A review and comparison of strategies. *Journal of Cross-Cultural Psychology, 16*(2), 131-152.

Jang, Y., Kwag, K. H., & Chiriboga, D. A. (2010). Not saying I am happy does not mean I am not: Cultural influences on responses to positive affect items in the CES-D. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences, 65B* (6), 684-690.

Kingery, J. N., Ginsburg, G. S., & Alfano, C. A. (2007). Somatic symptoms and anxiety among African American adolescents. *Journal of Black Psychology, 33*(4), 363-378.

Kline, P. (1979). *Psychomatrics and psychology*. London: Academic Press.

Lambert, S. F., Cooley, M. R., Campbell, K. D. M., Benoit, M. Z., & Stansbury, R. (2004). Assessing anxiety sensitivity in inner-city African American children: Psychometric properties of the Childhood Anxiety Sensitivity Index. *Journal of Clinical Child & Adolescent Psychology, 33*(2), 248 - 259.

Millsap, R. E., & Yun-Tein, J. (2004). Assessing factorial invariance in ordered-categorical measures. *Multivariate Behavioral Research, 39*, 479–515.

Mak, W., & Rosenblatt, A. (2002). Demographic influences on psychiatric diagnoses among youth served in California systems of care. *Journal of Child and Family Studies, 11*, 165-178.

Motl, R.W., Dishman, R.K., Birnbaum, A.S., Lytle, L.A. (2005). Longitudinal invariance of the Center for Epidemiologic Studies-Depression Scale among girls and boys in middle school. *Educational and Psychological Measurement, 65*, 90-108.

Muthén, B. O., & Asparouhov, T. (2002). Latent variable analysis with categorical outcomes: Multiple-group and growth modeling in Mplus (Mplus Web Note No. 4). Retrieved July 30, 2012, from http://www.statmodel.com/mplus/examples/webnote.html

Muthén, B., & Kaplan, D. (1985). A comparison of some methodologies for the factor analysis of non-normal Likert variables. *British Journal of Mathematical and Statistical Psychology*, *38*, 171-189.

Muthén, L.K., & Muthén, B.O. (1998-2011). Mplus user's guide, Sixth edition. Los Angeles, CA: Muthén & Muthén

Neal, A. M., & Ward Brown, B. J. (1994). Fears and anxiety disorders in African American children. In S. Friedman (Ed.) *Anxiety disorders in African Americans* (pp. 65-75). New York: Springer.

Nettles, S. M., & Pleck, J. H. (1994). Risk, resilience, and development: The multiple ecologies of Black adolescents in the United States. In R. J. Haggerty, N. Garmezy, M. Rutter & L. Sherrod (Eds.), *Stress, risk, and resilience in children and*

adolescents: Processes, mechanisms, and intervention (pp. 147-181). New York: Cambridge University Press.

Nguyen, L., Huang, L. N., Arganza, G. F., & Liao, Q. (2007). The influence of race and ethnicity on psychiatric diagnoses and clinical characteristics of children and adolescents in children's services. *Cultural Diversity and Ethnic Minority Psychology, 13*(1), 18-25.

Okazaki, S. (1997). Sources of ethnic differences between Asian American and White American college students on measures of depression and social anxiety. *Journal of Abnormal Psychology, 106*(1), 52-60.

Pang, V. O. (1991). The relationship of test anxiety and math achievement to parental values in Asian-American and European-American middle school students. *Journal of Research & Development in Education, 24*(4), 1-10.

Perreira, K. M., Deeb-Sossa, N., Harris, K. M., & Bollen, K. (2005). What are we measuring? An evaluation of the CES-D across race/ethnicity and immigrant generation. *Social Forces, 83*(4), 1567-1602.

Prescott, C. A., McArdle, J. J., Hishinuma, E. S., Johnson, R. C., Miyamoto, R. H., Andrade, N. N., . . . Carlton, M. D. (1998). Prediction of major depression and dysthymia from CES-D scores among ethnic minority adolescents. *Journal of the American Academy of Child and Adolescent Psychiatry, 37*(5), 495-503.

Radloff, L. S. (1977). The CES-D Scale: A self-report depression scale for research in the general population. *Applied Psychological Measurement, 1*(3), 385-401.

Radloff, L.S. (1991). The use of the Center for Epidemiological Studies Depression Scale in adolescents and young adults. *Journal of Youth and Adolescents, 20*, 149-166.

Raftery, A.E. (1993). Bayesian model selection in structural equation models. In K.A. Bollen & J.S. Long (Eds.), *Testing structural equation models* (pp. 163-180), Newbury Park, CA: Sage.

Roberts, R. E., Andrews, J. A., Lewinsohn, P. M., & Hops, H. (1990). Assessment of depression in adolescents using the Center for Epidemiologic Studies Depression Scale. *Psychological Assessment: A Journal of Consulting and Clinical Psychology, 2*(2), 122-128.

Roberts, R. E., Lewinsohn, P. M., & Seeley, J. R. (1991). Screening for adolescent depression: A comparison of depression scales. *Journal of the American Academy of Child & Adolescent Psychiatry, 30*(1), 58-66.

Roberts, R. E., Roberts, C. R., & Chen, Y. (1997). Ethnocultural differences in prevalence of adolescent depression. *American Journal of Community Psychology, 25*(1), 95-110.

Roberts, R. E., & Sobhan, M. (1992). Symptoms of depression in adolescence: A comparison of Anglo, African, and Hispanic Americans. *Journal of Youth and Adolescence, 21*(6), 639-651.

Russell, S., Crockett, L., Shen, Y. L., & Lee, S.A. (2008). Cross-ethnic invariance of self-esteem and depression measures for Chinese, Filipino, and European American adolescents. *Journal of Youth and Adolescence, 37*(1), 50-61.

Sagrestano, L. M., Paikoff, R. L., Holmbeck, G. N., & Fendrich, M. (2003). A longitudinal examination of familial risk factors for depression among inner-city African American adolescents. *Journal of Family Psychology, 17*(1), 108-120.

Siegel, J. M., Aneshensel, C. S., Taub, B., Cantwell, D. P., & Driscoll, A. K. (1998). Adolescent depressed mood in a multiethnic sample. *Journal of Youth and Adolescence, 27*(4), 413-427.

Silverman, W. K., La Greca, A. M., & Wasserstein, S. (1995). What do children worry about? Worries and their relation to anxiety. *Child Development, 66*(3), 671-686.

Stark, K. D., & Laurent, J. (2001). Joint factor analysis of the Children's Depression Inventory and the Revised Children's Manifest Anxiety Scale. *Journal of Clinical Child & Adolescent Psychology, 30*(4), 552 - 567.

U.S. Department of Health and Human Services. (2001). Mental health: Culture, race, and ethnicity – A supplement to mental health: A report of the Surgeon General. Rockville, MD: U.S. Department of Health and Human Services, Substance Abuse and Mental Health Services Administration, Center for Mental Health Services.

Van de Schoot, R., Lugtig, P., & Hox, J. (2012). A checklist for testing measurement invariance. European Journal of Developmental Psychology, Developmetrics, 1-7.

Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods, 3*(1), 4-69.

Varela, R. E., Vernberg, E. M., Sanchez-Sosa, J. J., Riveros, A., Mitchell, M., & Mashunkashey, J. (2004). Parenting style of Mexican, Mexican American, and Caucasian-Non-Hispanic families: Social context and cultural influences. *Journal of Family Psychology, 18*(4), 651-657.

West, S. G., Finch, J. F., & Curran, P. J. (1995) Structural equation models with nonnormal variables: Problems and remedies. In R. H. Hoyle (Ed.), Structural

equation modeling: concepts, issues, and applications (pp. 56-75). Thousand Oaks, CA: Sage Publishing.

Wren, F. J., Berg, E. A., Heiden, L. A., Kinnamon, C. J., Ohlson, L. A., Bridge, J. A., . . . Bernal, M. P. (2007). Childhood anxiety in a diverse primary care population: Parent-child reports, ethnicity and SCARED factor structure. *Journal of the American Academy of Child and Adolescent Psychiatry, 46*(3), 332-340.