

A Measurement Invariance Examination of the Revised Child Anxiety and Depression Scale in a Southern Sample: Differential Item Functioning Between African American and Caucasian Youth

Lindsay Rae Trent, Erin Buchanan, Chad Ebesutani, Chelsea M. Ale, Laurie Heiden, Terry L. Hight, John D. Damon and John Young

Assessment published online 31 July 2012
DOI: 10.1177/1073191112450907

The online version of this article can be found at:
<http://asm.sagepub.com/content/early/2012/07/31/1073191112450907>

Published by:



<http://www.sagepublications.com>

Additional services and information for *Assessment* can be found at:

Email Alerts: <http://asm.sagepub.com/cgi/alerts>

Subscriptions: <http://asm.sagepub.com/subscriptions>

Reprints: <http://www.sagepub.com/journalsReprints.nav>


Permissions: <http://www.sagepub.com/journalsPermissions.nav>

>> [OnlineFirst Version of Record](#) - Aug 1, 2012

[OnlineFirst Version of Record](#) - Jul 31, 2012

[What is This?](#)

A Measurement Invariance Examination of the Revised Child Anxiety and Depression Scale in a Southern Sample: Differential Item Functioning Between African American and Caucasian Youth

Assessment
XX(X) 1–13
© The Author(s) 2012
Reprints and permission:
sagepub.com/journalsPermissions.nav
DOI: 10.1177/1073191112450907
http://asm.sagepub.com


Lindsay Rae Trent¹, Erin Buchanan², Chad Ebesutani³, Chelsea M. Ale⁴,
Laurie Heiden⁵, Terry L. Hight⁵, John D. Damon⁵, and John Young¹

Abstract

This study examined the psychometric properties of the Revised Child Anxiety and Depression Scale in a large sample of youth from the Southern United States. The authors aimed to determine (a) if the established six-factor Revised Child Anxiety and Depression Scale structure could be replicated in this Southern sample and (b) if scores were associated with measurement invariance across African American and Caucasian youth representative of youth from this region of the United States. The established six-factor model evidenced the best fit in comparison to one-, two-, and five-factor models in the total sample ($N = 12,695$), as well as in the African American ($n = 4,906$) and Caucasian ($n = 6,667$) subsamples. Multigroup confirmatory factor analysis also supported measurement invariance across African American and Caucasian youth at the levels of equal factor structure and equal factor loadings. Noninvariant item intercepts were identified, however, indicating differential functioning for a subset of items. Clinical and measurement implications of these findings are discussed and new norms are presented.

Keywords

psychometrics, Revised Child Anxiety and Depression Scale, ethnicity, norms, measurement invariance, confirmatory factor analysis

The Importance of Assessing Anxiety and Depression in Youth

Epidemiological studies suggest that the debilitating effects of anxiety and depression are experienced by approximately 8% to 27% of children and adolescents at some point in their development (Costello, Mustillo, Erkanli, Keeler, & Angold, 2003). Anxiety disorders in childhood have been linked to the development of additional anxiety disorders, substance abuse, depression in adolescence and adulthood, and lower rates of educational attainment (Bittner et al., 2007; Egger, Costello, & Angold, 2003; Grover, Ginsburg, & Jalongo, 2007). These disorders, however, are often overlooked by teachers and parents because of the covert nature of symptoms typically associated with these disorders (Shahar et al., 2006). As a result, children may experience significant interference in academic and social functioning prior to formal, resource-intensive identification of these problems (Muris & Meesters, 2002). For

these reasons, empirically supported assessment tools may be particularly important to identify these issues in youth.

Many well-researched self-report instruments are available for the assessment of youth anxiety and depression (Silverman & Ollendick, 2005). Various types of symptoms associated with anxiety or depression are targeted by these measures; however, several of these instruments were not developed using current nosological guidelines (*Diagnostic and Statistical Manual of Mental Disorders*, fourth edition,

¹The University of Mississippi, University, MS, USA

²Missouri State University, Springfield, MO, USA

³Yonsei University, Seoul, Korea

⁴University of South Florida, St. Petersburg, FL, USA

⁵Mississippi Children's Home Services and CARES School, Jackson, MS, USA

Corresponding Author:

Lindsay Rae Trent, Department of Psychology, University of Mississippi, Peabody Hall, University, MS 38677, USA
Email: lrtrent@olemiss.edu

text revision [*DSM-IV*]; American Psychiatric Association, 2000) and have either questionable or absent data concerning psychometric properties (Silverman & Ollendick, 2005). Considerable evidence also exists to suggest that comorbidity among clinically disordered youth is the norm rather than the exception (e.g., Angold & Costello, 1999; Seligman & Ollendick, 1998; Silverman & Ollendick, 2005). Youth with comorbidity are more severely impaired than those with only one disorder (Seligman & Ollendick, 1998) making measures' ability to discriminate among co-occurring disorders critical (Silverman & Ollendick, 2005). Overall, however, findings of studies assessing the psychometric properties of widely used child self-report measures demonstrate questionable ability to make accurate diagnostic formulations (e.g., Hodges, 1990; Seligman, Ollendick, Langley, & Baldacci, 2004; Silverman & Ollendick, 2005; Stark & Laurent, 2001).

These limitations in extant forms of measurement point toward the need for more detailed psychometric study in the context of contemporary nosological theory. Thus, in an attempt to better inform diagnostic formulations, researchers have begun developing measures that target specific dimensions of psychopathology according to current *DSM-IV* nosology and integrate recent conceptual changes concerning the relationship between anxiety and depressive disorders (i.e., tripartite model; Clark & Watson, 1991). The current study offers an examination of one such measure that is widely known and used across clinical and research settings—the Revised Child Anxiety and Depression Scale (RCADS; Chorpita, Yim, Moffitt, Umemoto, & Francis, 2000).

The RCADS (Chorpita et al., 2000) is a 47-item youth self-report measure of depression (major depressive disorder) and five major anxiety disorders consistent with the *DSM-IV* (social phobia, panic disorder, separation anxiety disorder, obsessive-compulsive disorder, and generalized anxiety disorder). Although the RCADS has received substantial empirical support for the reliability and validity of its scale scores using both Hawaii- and Australia-based samples (Chorpita et al., 2000; Chorpita, Moffitt, & Gray, 2005; de Ross, Gullone, & Chorpita, 2002), a notable feature of the RCADS is that its U.S. normative data were derived from a unique sample of school-based and clinic-referred youths in Hawaii. More than 20 different ethnic minorities were reported as comprising the Hawaii-based normative samples, the majority of whom were multiethnic or Asian (e.g., Japanese American, Chinese American, Filipino). Although this demographic distribution is consistent with the population composition in Hawaii, relatively fewer ethnicities were represented that are more typical in other regions of the country, particularly Caucasian and African American youth. For example, instrument development studies conducted with a school-based sample of 1,641 children and adolescents only included approximately

8% Caucasians ($n = 133$) and a negligible number of African American youth (exact value not reported as African American children were included in the “other” category along with 11 other ethnicities also representing a negligible proportion of the sample; Chorpita et al., 2000). Research in a clinic-referred sample of 513 children and adolescents included approximately 16% Caucasians ($n = 82$) and again a negligible number of African Americans (included in “other” category along with four other ethnicities; Chorpita et al., 2005). Although these studies were important in establishing the RCADS as a clinically useful measure, the degree to which currently available normative data generalize to youth from other regions of the United States is relatively unknown.

Importance of Examining Ethnic Differences

Research has highlighted the importance of examining differences across ethnic groups to inform needed modifications to scale item and structure, as well as how to best administer and interpret scores based on individuals from different cultures and ethnic backgrounds who may express symptoms of psychopathology differently. Failing to consider systematic variations due to cultural and ethnic differences (e.g., response pattern and symptom expression) could lead to inaccurate and thus misleading conclusions, especially given inconsistent research findings assessing psychopathology across ethnicities (Lambert, Cooley-Quille, Campbell, Benoit, & Stansbury, 2004). For instance, several discrepant findings related to depression expression across ethnic groups have been reported in epidemiological research (Nguyen, Kitner-Triolo, Evans, & Zonderman, 2004). Some studies have reported higher prevalence rates of depressive symptoms among African American males compared with Caucasian males (Jones-Webb & Snowden, 1993), whereas others report the opposite pattern (Dunlop, Song, Lyons, Manheim, & Chang, 2003; D. R. Williams et al., 2007) or fail to detect differences across these groups (Berkman et al., 1986). Importantly, it has been suggested that a contributing factor to inconsistent findings is differential responding across these groups because of social and/or cultural differences rather than differences in levels of the construct of interest, which could make comparisons across cultural groups across cultures inappropriate (e.g., African Americans incorporating somatic symptoms into their responses to questions assessing affective symptoms reflective of cultural differences; Ginsburg, Riddle, & Davies, 2006; Tylee & Gandhi, 2005).

Consequently, researchers have begun to examine idiosyncrasies in differential reporting of anxiety and depression symptoms between African American and Caucasian samples. Scott, Eng, and Heimberg (2002) found that African American and Caucasian college students both reported fears (related to generalized anxiety disorder) on

the Penn State Worry Questionnaire (Meyer, Miller, Metzger, & Borkovec, 1990) but endorsed different content domains as measured by the Worry Domains Questionnaire (Tallis, Eysenck, & Mathews, 1992). Differences in reported panic disorder symptoms have also been found between African American and Caucasian adults (e.g., Smith, Friedman, & Nevid, 1999), whereby African Americans reported relatively more concerns related to a fear of dying and symptoms of intense numbing and tingling sensations. Similarly, Carter, Miller, Sbrocco, Suchday, Lewis (1999) examined the factor structure of the Anxiety Sensitivity Index among African American college students and found that the original three-factor structure did not fit well among this sample. Similar results have been found in youth samples as well. For example, in a sample of school-aged children using the Child Anxiety Sensitivity Index (Lambert et al., 2004) the original and three- and four- higher order structures (based on Silverman, Ginsburg, & Goedhart's, 1999 study, which comprised a primarily Caucasian sample) were not supported in the African American youth in their sample. Instead, they found support for only two meaningful factors (physical concerns and mental incapacitation concerns), demonstrating unique expressions of anxiety sensitivity among African American children compared with Caucasian children.

Neal, Lilly, and Zakis (1993) also found differences in the factor structure of the Revised Fear Survey Schedule for Children among a child sample (age 6-12 years). Specifically, they found that the original five-factor solution best fit the data for Caucasian youth, whereas a three-factor solution was superior for African American youth. Similarly, Boyd, Ginsburg, Lambert, Cooley, and Campbell (2003) examined the psychometric properties of the Screen for Child Anxiety Related Emotional Disorders (Birmaher et al., 1999) in an African American sample comprising adolescent high school youth and found that a three-factor solution fit better for African Americans, in contrast to the original five-factor solution supported by primarily Caucasian youth samples in previous studies (Birmaher et al., 1999).

Inadequate representation of African American and Caucasian youth in the samples of previous psychometric examinations of the RCADS has prevented any systematic examinations to determine how the RCADS may perform differently across these ethnic groups. Weems, Cost, Watts, Taylor, and Cannon's (2007) recent study found that African American youth reported significantly higher RCADS Anxiety Total scores compared with Caucasian youth; however, this ancillary analysis was not the focus of their article. More thorough, systematic research and psychometric study is thus needed in this area to elucidate how African American and Caucasian youth's scores may differ when completing the RCADS—including whether certain RCADS items are associated with differential item functioning across these groups.

Examining Measurement Invariance

Based on the aforementioned studies' findings of greater reported anxiety (in some domains) by African Americans relative to Caucasian counterparts, one might conclude that African Americans are at an increased risk for problems in these domains. This conclusion assumes, however, that the raw scores produced by the measures "tap" the underlying construct in the exact same way across groups. This indeed relates to the fundamentally difficult problem faced by both researchers and clinicians alike related to the challenge of being able to know whether differences in raw scores between groups are reflective of (a) actual differences between groups on the targeted underlying constructs or (b) merely differences in reporting styles between groups (even for individuals who are at the same level of the underlying construct). So difficult is this problem that not only is it common for researchers to ignore this issue—often completely—when comparing scores across groups, but it is also difficult to provide recommendations regarding how to properly handle this issue based on recent advances in statistical modeling of latent constructs; indeed, the answer may be one that researchers do not want to hear.

Nonetheless, it is important to remind readers that statistical modeling techniques do currently exist that allow researchers to discriminate between conditions (a) and (b) above. Multigroup confirmatory factor analysis (MG-CFA) and multiple indicators–multiple causes (MIMIC) CFA are the mostly widely used methods for examining measurement invariance across groups. Although MG-CFA is considered the most powerful and versatile approach to examining measurement invariance across groups (Brown, 2006; Steenkamp & Baumgartner, 1998)—including having advantages over MIMIC CFA, such as having the ability to measure all aspects of measurement and population heterogeneity (cf. Brown, 2006)—MG-CFA is still grossly underutilized in applied research (Vandenberg & Lance, 2000), likely contributing to the confusion in the literature and discrepant findings reported across studies pertaining to apparent "group differences" (or the lack thereof).

If researchers are truly interested in comparing scores across groups to understand how groups actually differ on the underlying targeted construct(s), then they must contend with the fact that simply comparing raw scores between groups (without first ensuring that the scores of both groups are associated with measurement invariance) is not sufficient. Such would be akin to wanting to compare between apples and apples (e.g., to determine which group of apples is larger), without first knowing whether or not you are even comparing between (1) apples and apples or between (2) apples and oranges! For the sake of brevity, the steps to make this determination between conditions (a) and (b) under an MF-CFA framework are outlined in the Data Analytic Approach section. The point here, however, is that

requisite steps must first be conducted for researchers to acquire the adequate confidence that they are indeed comparing what they think (and hope to be) comparing.¹

For these reasons, as well as the reasons stated above outlining the lack of consistency in the literature regarding how ethnic group scores' differ from each other, more studies that adequately examine measurement invariance across ethnic subgroups are needed to advance clinical knowledge and assessment procedures—particularly in areas across the United States that are becoming increasingly more multicultural and diverse. One geographical area amenable to studies of this sort is the American South, given its high percentage of African American inhabitants (e.g., in Mississippi 44% of children are African American; U.S. Census Bureau, 2010) and disproportionate representation of known risk factors for psychopathology. In 2009, Alabama, Arkansas, Kentucky, Mississippi, and West Virginia were estimated to have the highest poverty rates in the United States (U.S. Census Bureau, 2010), which is a known risk factor for psychopathology as well as a predictor of associated risk factors (e.g., maladaptive parenting behaviors, out-of-home placement, etc.; Costello, Compton, Keeler, & Angold, 2003; Samaan, 2000). The outcomes of such studies are also especially clinically salient to these areas, as it remains unclear whether current assessment tools can appropriately be directly applied to these populations.

The Present Study

In the present study, therefore, we aimed to determine (a) whether the previously established factor structure for the RCADS could be replicated in a sample of Southern youth comprising ethnic diversity typical of the region (i.e., Caucasian and African American) and (b) whether scores obtained on the RCADS were associated with measurement invariance across these youths. We were also particularly interested in whether African American and Caucasian youths' reports on the various RCADS anxiety and depression items were associated with *differential item functioning* across these two predominant ethnic groups in the South, as this would suggest the need for ethnic-specific normative data to aid in the interpretation of scale scores. Given previous studies finding differences between African American and Caucasian youth reports on internalizing problems (Boyd et al., 2003; Lambert et al., 2004; Neal et al., 1993; Weems et al., 2007), we hypothesized that African American and Caucasian youths' RCADS scores would be associated with some degree of differential reporting, demonstrating the need for new normative data specific to these ethnic groups.

Method

Participants

The present sample was derived from children and adolescents in Grades 2 to 12 in public schools across the state of

Mississippi ($N = 12,802$; median grade = 7) who completed a battery of questionnaires that included the RCADS. Survey responses were first examined for missing data, and youth having more than 10% missing data ($n = 107$; 0.8%) were excluded from the analysis. Among the 12,695 remaining participants, 11,718 (92.3%) had no missing data, 799 (6.3%) had one missing item, 122 (1%) were missing two items, 34 (0.3%) had three missing items, 13 (0.1%) were missing four items, and 9 (0.1%) were missing five items. We imputed missing data using the Missing Value Analysis module of SPSS 18.0, whereby missing data patterns are analyzed and data are imputed via a maximum likelihood algorithm (Little & Rubin, 1987).

The sample comprised 6,477 (51%) females and 6,218 (49%) males. With respect to youth ethnicity, 6,667 (52.5%) were Caucasian, 4,906 (38.6%) were African American, 323 (2.5%) were Latino/Hispanic, 227 (1.8%) were Asian, and 523 (4.1%) were other. Forty-nine youth (0.4%) did not report ethnicity data. With respect to youths per grade, there were 839 (6.6%) second graders, 847 (6.7%) third graders, 984 (7.8%) fourth graders, 991 (7.8%) fifth graders, 1,756 (13.8%) sixth graders, 2,064 (16.3%) seventh graders, 1,758 (13.8%) eighth graders, 1,431 (11.3%) ninth graders, 718 (5.7%) tenth graders, 689 (5.4%) eleventh graders, and 618 (4.9%) twelfth graders. Based on the school districts surveyed in the present study (i.e., Bolivar, Coahoma, Hancock, Harrison, Hinds, Jackson, Jefferson, Madison, Pontotoc, and Simpson counties in Mississippi), we took a weighted average of the most recent and available U.S. Census Bureau data (U.S. Census Bureau, 2010) to approximate family income of our participants. Approximately 22% of families from the districts surveyed fell below the poverty line, with mean family income being \$40,500.

Procedure

Data for this study were collected as part of a broader school-based mental health screening initiative in Mississippi (the Behavioral Vital Signs Project [BVS]). The BVS administers scientifically supported mental health screenings to youth in Grades 2 to 12 and provides feedback to schools concerning students' mental health, including information pertaining to internalizing and externalizing behavior, loneliness, and hazardous or risky behavior (such as drug use). The study used a passive consent procedure that was approved by the Mississippi Department of Education, the University Mississippi Institutional Review Board, and each school involved in the BVS. Members of the project staff distributed assessment materials to each classroom, and teachers were informed of study procedures. Teachers read a brief set of instructions and handed out assessment packets to their students, who provided answers via Scantron brand optical forms. All questionnaires were completed anonymously.

Measures

Revised Child Anxiety and Depression Scale (RCADS). The RCADS (Chorpita et al., 2000; Chorpita et al., 2005) is a 47-item, youth self-report questionnaire with subscales corresponding to separation anxiety disorder (SAD), social phobia (SP), generalized anxiety disorder (GAD), panic disorder (PD), obsessive-compulsive disorder (OCD), and major depressive disorder (MDD). The RCADS also yields a Total Anxiety scale (sum of all anxiety subscales) and a Total Internalizing scale (sum of all scales). Items are rated on a 4-point Likert-type scale from 0 = *never* to 3 = *always*. The factor structure, reliability, and validity of the RCADS scales have been supported in both school-based and clinic-referred samples (Chorpita et al., 2000; Chorpita et al., 2005).

Data Analytic Approach

Confirmatory factor analysis. We first conducted CFA using AMOS version 18 (Arbuckle, 1995) and maximum likelihood estimation to examine whether the six-factor structure of the RCADS was supported based on the full sample. We evaluated model fit via various fit indices, including the χ^2/df ratio (Bryant & Yarnold, 1995), root mean square error of approximation (RMSEA; Steiger, 1990), comparative fit index (CFI; Bentler, 1990), and the Tucker-Lewis index (TLI; Tucker & Lewis, 1973). A good fitting model has a small χ^2/df ratio (closer to zero), an RMSEA of .05 or less (.08 marginal, .10 poor fit; Browne & Cudeck, 1993), and TLI/CFI values greater than .90. The chi-square (χ^2) difference test was used to examine whether the six-factor model fit significantly better than competing models, including the following: (a) a one-factor model of general negative affect, (b) a two-factor model of depression and anxiety (collapsing all anxiety items into a single “broad anxiety” factor), and (c) a five-factor model whereby MDD and GAD subscales are collapsed into a “distress” factor, as recently proposed (Lahey et al., 2008; Watson, 2005).

Multigroup confirmatory factor analysis. We then conducted MG-CFA using AMOS version 18 to determine whether the best fitting model of the RCADS was associated with measurement invariance across African American ($n = 4,906$) and Caucasian ($n = 6,667$) youth. We did not examine measurement invariance between any other ethnic groups because of insufficient sample sizes. Although MG-CFA has been noted as a somewhat underutilized approach in applied research (Vandenberg & Lance, 2000), it is widely accepted as the most powerful and versatile approach to examining measurement invariance (Brown, 2006; Steenkamp & Baumgartner, 1998). This method offers advantages over other approaches, such as MIMIC CFA, including the ability to measure all aspects of measurement and population heterogeneity (i.e., factor loadings, intercepts, residual variances, factor variances and covariances, and latent means; cf. Brown, 2006).

The procedures of MG-CFA involve applying CFA to more than one group simultaneously by analyzing separate input matrices specific to each group and placing constraints on various parameters. Although various terms have been used to describe the same MG-CFA invariance tests in the literature, we have used the terms suggested by Brown (2006) in the present article because these terms are more descriptive and easily interpretable.

Variation also exists regarding the order in which these parameters should be tested when conducting MG-CFA. For example, some have argued that invariance tests should begin by constraining all parameters to be equal across groups. Instead, however, we followed published standards for a stepwise approach, whereby the least restrictive model (i.e., equal form) was examined first followed by increasingly restrictive models with more parameters constrained. This approach was chosen because it facilitates identification of the parameter types associated with noninvariance (Brown, 2006).

Single-group solutions. Taking this stepwise approach, we began our MG-CFAs by conducting single-group solutions in our African American and Caucasian subsamples. These single group solutions produce a set of fit indices specific to each subsample. Examination of these fit indices reveals whether the posited (six-factor RCADS) structure provided an acceptable fit in each group. We assessed the degree of model fit for these single-sample solutions using the same fit indices and benchmarks noted above. Adequate fit of these single-group solutions would support the notion that both (African American and Caucasian) groups' scores on the RCADS reflect the same number of factors and would allow for additional tests to be conducted; hence the stepwise approach. Otherwise, scores from both groups would be considered to represent a different number of factors, precluding the ability to conduct further MG-CFA tests (Brown, 2006).

Equal form (or configural invariance). After assuring that the single-sample solutions fit well for both groups, we conducted the test of “equal form” (or “configural invariance”), which examines the equality of both groups' factor structure (i.e., whether the number of factors and pattern of indicator-factor loadings are identical across groups). This test produces only one set of fit indices, which are assessed via the same benchmarks noted above.

Metric invariance. If equal form of the RCADS was supported across groups (as evidenced by fit indices meeting the benchmarks for “good fit”), we then conducted the test of metric invariance. Metric invariance refers to whether factor loadings (i.e., the relationship between items and factors) are equivalent across groups—in other words, whether indicators evidence comparable relationships to the latent construct across African Americans and Caucasian youth in this sample. Metric invariance was supported if (a) fit indices achieved benchmarks for good fit and (b) model fit did not degrade significantly relative to the fit

Table 1. Fit Statistics for the Confirmatory Factor Analytic Models Based on the Full Sample ($N = 12,695$)

Model	χ^2	df	χ^2/df	p	RMSEA [90% CI]	CFI	TLI	Difference from six-factor	
								$\Delta\chi^2$	Δdf
Six-factor (MDD, GAD, SP, SAD, OCD, PD)	23342.35	1019	22.91	<.001	.041 [.041, .042]	.90	.89		
Five-factor (MDD/GAD, SP, SAD, OCD, PD)	26827.44	1024	26.20	<.001	.045 [.044, .045]	.88	.88	3485.09	5
Two-factor (anxiety/depression)	35391.32	1033	34.26	<.001	.051 [.051, .052]	.84	.84	12048.96	14
One-factor (general negative affect)	38014.92	1034	36.77	<.001	.053 [.053, .054]	.83	.82	14672.57	15

Note. RMSEA = root mean square error of approximation; CI = confidence interval; CFI = comparative fit index; TLI = Tucker–Lewis index; MDD = major depressive disorder; GAD = generalized anxiety disorder; SP = social phobia; SAD = separation anxiety disorder; OCD = obsessive–compulsive disorder; PD = panic disorder.

of the equal form solution previously tested. The fit of the “metric invariance” model was considered significantly degraded relative to the fit of the “equal form” model if $\Delta CFI > .01$ (Cheung & Rensvold, 2002). Although a significant chi-square test is often used to examine whether a (nested) model is significantly degraded, chi-square tests are sensitive to sample size (Brown, 2006) and our large sample would thus make it likely to always find a significant degrade in model fit based on the chi-square difference test. We therefore used $\Delta CFI > .01$ given that this test is less susceptible to confounding influences from large sample sizes (Medsker, Williams, & Holahan, 1994; Bentler, 1990).

Scalar invariance. If metric invariance of the RCADS items was supported across African American and Caucasians in our sample, we then tested scalar invariance across groups. This examines whether indicator intercepts are equal across groups. If scalar invariance is supported, then individuals on the same levels of the latent construct will produce the same raw score regardless of group membership. Scalar noninvariance, on the other hand, is similar to differential item functioning (McDonald, 1999), whereby individuals who are on the same level of the latent construct yield different raw scale scores. Support for scale invariance was demonstrated if (a) fit indices achieved the benchmarks for good fit and (b) model fit was not degraded significantly relative to the fit of the metric invariance solution, as mentioned above (i.e., $\Delta CFI > .01$).

As noted above, this MG-CFA stepwise approach is such that each test can only be conducted if all previous tests support invariance (e.g., metric invariance tests can only be conducted and meaningfully interpreted if both single-sample and equal form solutions are supported). However, it is notable that it is rare for all items to evidence measurement invariance at all levels, particularly in cross-cultural psychometric research (Cheung & Rensvold, 1999). When the omnibus test of measurement invariance is not supported beyond the level of equal form (i.e., configural invariance), tests of *partial* measurement invariance may be conducted to examine whether the overall lack of invariance is due to *all* parameters being noninvariant across

groups, or due to just a *subset* of items being noninvariant across groups (Byrne, Shavelson, & Muthén, 1989). For example, a scale may fail the omnibus tests of metric invariance (suggesting noninvariance of factor loadings across groups) due to just a few factor loadings being different across groups as opposed to all factor loadings being different across groups.

Various methods have been proposed for handling cases of partial measurement invariance. First, the scale may be abandoned completely. Alternately, the noninvariant items may simply be deleted from the scale. It is notable, however, that Cheung and Rensvold (1999) discouraged this approach of deleting noninvariant item because of the potential compromise in construct validity as well as violating the underlying theory of the scale and its items. Another strategy recommended—and used in the present study—is to treat partial measurement invariance as sufficient support for measurement invariance if the following three criteria are met: (a) the proportion of the noninvariant items must be small relative to the invariant items, (b) the noninvariant items should be meaningfully related to the constructs for all groups, and (c) equal form must hold (Cheung & Rensvold, 1999). Relatedly, when proceeding with tests of measurement invariance under conditions of partial measurement invariance, the constraints of all noninvariant items must be relaxed and carried through all remaining tests (Brown, 2006; Byrne et al., 1989).

Results

Confirmatory Factor Analysis

The fit indices for the full sample associated with the various models appear in Table 1. Results indicated that the six-factor model evidenced the best set of fit indices that also met benchmark for good fit (e.g., CFI = .90, RMSEA = .041) compared with the competing models. The six-factor model also fit the data (based on the full sample) significantly better than the alternate five-factor, $\Delta\chi^2(5) = 3485.09$, $p < .001$; two-factor, $\Delta\chi^2(14) = 12048.96$, $p < .001$; and one-factor models, $\Delta\chi^2(15) = 14672.57$, $p < .001$. All standardized

Table 2. Fit Statistics for the Multigroup Confirmatory Factor Analytic Six-Factor Model

Six-factor model	χ^2	df	χ^2/df	RMSEA [90% CI]	TLI	CFI	Δ CFI	Δ df
Single-group solutions								
Caucasians (n = 6,667)	14911.88	1019	14.63	.045 [.045, .046]	.89	.893	—	—
African Americans (n = 4,906)	8637.98	1019	8.48	.039 [.038, .039]	.89	.893	—	—
Equal form	23549.79	2038	11.55	.030 [.030, .031]	.89	.893	—	—
Metric invariance	23726.15	2079	11.41	.030 [.030, .030]	.89	.892	-0.001	41
Scalar invariance	26133.25	2126	12.29	.031 [.031, .032]	.88	.881	-0.012	47
Scalar (partial) invariance	25264.21	2121	11.91	.031 [.030, .031]	.88	.885	-0.007	42

Note. RMSEA = root mean square error of approximation; CI = confidence interval; CFI = comparative fit index; TLI = Tucker–Lewis index.

factor loadings for the six-factor model were statistically significant ($ps < .05$, supporting each item as adequately tapping each factor. The loadings for Generalized Anxiety questions ranged from .52 to .80 (Cronbach's $\alpha = .83$), Major Depressive questions .45 to .62 (Cronbach's $\alpha = .83$), Obsessive Compulsive questions .54 to .63 (Cronbach's $\alpha = .76$), Panic questions .43 to .68 (Cronbach's $\alpha = .83$), Separation Anxiety questions .50 to .67 (Cronbach's $\alpha = .76$), and finally Social Phobia questions .50 to .69 (Cronbach's $\alpha = .83$).

Multigroup Confirmatory Factor Analysis Between Caucasians and African Americans

Single-sample solutions. The fit indices associated with the single-sample solutions for the Caucasian and African American subsamples appear in Table 2. Results demonstrate that the six-factor model fit well in both the Caucasian (e.g., RMSEA = .045) and African American (e.g., RMSEA = .039) youth samples. These results support the general six-factor structure in both ethnicity groups (i.e., reports from both groups map onto the problem areas of MDD, GAD, SAD, OCD, PD, and SP).

Test for equal form. Placing equality constraints on the factorial structure across groups led to adequate fit indices, supporting the item cluster patterns associated with the six-factor RCADS model. Specifically, the χ^2/df ratio was fairly low for a large data set (11.55), RMSEA suggested excellent fit (.03), and CFI was close to good fit at .89. Overall, these results suggest that both African American and Caucasian youths provided scores on the RCADS that evidenced similar item cluster patterns mapping well onto the six-factor structure of the RCADS.

Test for metric invariance. Following support for equal form across groups, we then tested for metric invariance across African Americans and Caucasians by imposing equality constraints on all factor loadings across groups, as shown in Table 2. This test did not result in a significant CFI drop in model fit relative to the equal form solution, thereby supporting metric invariance across groups. Fit indices in general also supported adequate model fit ($\chi^2/df = 11.41$, RMSEA = .03, and CFI = .89). Overall, these results

indicated that the general item–factor relationships were the same across African American and Caucasian youths.

Test for scalar invariance. Scalar invariance was examined across groups by imposing equality constraints on item intercepts. Importantly, Δ CFI was greater than .01, suggesting that at least one item intercept was noninvariant across groups (Brown, 2006). We therefore examined for *partial* scalar invariance, which involved identifying noninvariant item intercepts based on the procedures outlined in Cheung and Rensvold (1999). Specifically, we constrained item intercepts one at a time and conducted a series of invariance tests. Through this process, we identified five noninvariant items (RCADS Items 6, 8, 23, 42, and 44) that significantly degraded model fit when specifying each intercept pair to be equal across groups. The test for partial scalar invariance was then employed by allowing the item intercepts associated with these five items to be freely estimated across groups, and constraining the remaining 42 item intercepts to be equivalent across groups. The fit indices associated with the test of these set of constraints also appear in Table 2. Results supported partial scalar invariance given that Δ CFI did not exceed .01. Other fit indices also supported good fit (RMSEA = .03, $\chi^2/df = 11.91$). Overall, these results in support of *partial* scalar invariance revealed that the item intercepts for those five items were noninvariant across Caucasian and African American youths.² Therefore, African American and Caucasian youths who are at the same level of the latent construct reported different raw scale scores on those items.

Normative Data

We calculated means and standard deviations to provide normative data for the RCADS scales based on sex (boys, girls) and grade (2–12) following the same groupings of sex by grade levels as the original RCADS normative tables (cf. Chorpita et al., 2000; Chorpita et al., 2005) with the addition of a slightly younger age group in the current sample's Grade 2 participants. We also calculated and presented these normative data by ethnicity subgroups (i.e., African Americans, Caucasians, all ethnicities) given that African Americans and Caucasians evidenced some degree

Table 3. Number of Youth Falling in the Borderline and Clinical Ranges for Each of the RCADS Scales Based on Hawaii and Southern Normative Data

RCADS scale	Hawaii norms	Southern norms (ethnic specific)	Southern norms (not ethnic specific)
GAD			
Borderline elevation: n (%)	866 (6.8)	1,427 (11.2)	1,398 (11.0)
Clinical elevation: n (%)	436 (3.4)	815 (6.4)	803 (6.3)
OCD			
Borderline elevation: n (%)	1,183 (9.3)	1,557 (12.3)	1,815 (14.3)
Clinical elevation: n (%)	640 (5.0)	986 (7.8)	933 (7.3)
SAD			
Borderline elevation: n (%)	2,021 (15.9)	2,068 (16.3)	1,759 (13.9)
Clinical elevation: n (%)	1,318 (10.4)	1,248 (9.8)	1,155 (9.1)
PD			
Borderline elevation: n (%)	1,857 (14.6)	1,332 (10.5)	1,171 (9.2)
Clinical elevation: n (%)	1,268 (10.0)	792 (6.2)	675 (5.3)
SP			
Borderline elevation: n (%)	733 (5.8)	1,550 (12.2)	1,503 (11.8)
Clinical elevation: n (%)	359 (2.8)	844 (6.6)	7,44 (5.9)
MDD			
Borderline elevation: n (%)	2,199 (17.3)	1,237 (9.7)	1,126 (8.9)
Clinical elevation: n (%)	1,489 (11.7)	607 (4.8)	528 (4.2)
Anxiety total			
Borderline elevation: n (%)	1,202 (9.5)	1,627 (12.9)	1,436 (11.3)
Clinical elevation: n (%)	714 (5.6)	877 (6.9)	772 (6.1)
Total scale			
Borderline elevation: n (%)	1,341 (10.6)	1,508 (11.9)	1,363 (10.7)
Clinical elevation: n (%)	809 (6.4)	776 (6.1)	686 (5.4)
All scales			
Borderline elevation: n (%)	4,093 (32.2)	4,031 (31.8)	3,885 (30.6)
Clinical elevation: n (%)	2,846 (22.4)	2,605 (20.5)	2,465 (19.4)

Note. RCADS = Revised Child Anxiety and Depression Scale; GAD = Generalized Anxiety Disorder subscale; OCD = Obsessive-Compulsive Disorder subscale; SAD = Separation Anxiety subscale; PD = Panic Disorder subscale; SP = Social Phobia subscale; MDD = Major Depressive Disorder subscale; "Anxiety total" refers to combined *T*-score of the five anxiety subscales; "Total scale" refers to combined *T*-score of all six subscales; "All scales" refers to the number of youth across all eight RCADS scales who were categorized in at least one clinical/borderline range of elevation. Borderline elevation was defined by $70 > T \geq 65$; clinical elevation was defined as $T > 70$.

of differential item functioning for certain items. Because of spatial limitations these tables are not presented here but can be accessed online at <http://www.olemiss.edu/depts/psychology/people/faculty/young.html>.

We compared rates of scoring meeting cutoff criteria for clinical (i.e., *T*-score > 70) and borderline (i.e., *T*-score > 65) elevations on the various RCADS subscales within our sample (Chorpita et al., 2000; Chorpita et al., 2005) based on three different RCADS normative data sets: (a) the original RCADS normative data derived from Chorpita et al.'s (2000) sample of Hawaii youth (specific to grade and gender), (b) the newly acquire normative data in the present study that is specific to the Southern region only (also specific to grade and gender), and (c) the same newly acquire normative data in the present study specific to the Southern region—but further divided by ethnic group membership (i.e., African Americans, Caucasians,

all ethnicities). The purpose of making this comparison was to assess whether usage of different normative information would result in different rates of borderline and clinical elevations. As seen in Table 3, there were differences in the number of children exceeding clinical cutoffs both at the region-specific level (i.e., Hawaii sample vs Southern sample norms) and ethnic-specific (i.e., total Southern sample vs norms divided by group membership). The practical, "real-world" implication of these findings is that usage of different norms would differentially inform diagnostic formulations and diagnostic outcomes for a substantial number of children (Table 3)—thus affecting appropriate prescription and allocation of mental health services. Therefore, this examination provided support for the usage of both region-specific and ethnic-specific normative information when scoring RCADS *T*-scores to inform clinical decision making.

Discussion

The present study extends support for the psychometric properties of the RCADS to a new sample comprising primarily African American and Caucasian youth from relatively low socioeconomic status backgrounds residing in rural areas of the Southeastern United States. Through confirmatory factor analytic methods, we demonstrated that the originally proposed six-factor structure of the RCADS (Chorpita et al., 2000) was supported in the present large school-based sample of youth. The six-factor structure also evidenced significantly better model fit compared with competing one-, two-, and five-factor structures, providing further support of the RCADS as a measure of MDD and specific anxiety subtypes in accordance with *DSM-IV* nosology. Combining the MDD and GAD indicators to form a single “distress” factor in the five-factor model was associated with significantly degraded model fit relative to the originally posited six-factor RCADS structure (positing MDD and GAD as separate factors), questioning recent findings that MDD and GAD constitute the same “distress” factor (Lahey et al., 2008). Future research should further examine the relationship between MDD and GAD symptoms in other samples to better elucidate the validity of this purported distress factor.

We also conducted MG-CFA across the most representative ethnic groups in this region of the United States to determine the degree of measurement invariance across these groups, as well as the need for ethnic-specific normative data to allow for more precise scale score interpretations among African American and Caucasian youth. We found that the RCADS items were invariant across African American and Caucasian youth at the levels of factor structure and factor loadings. Constraining the item-intercepts to equality resulted in significantly degraded overall model fit, however, suggesting some degree of differential item functioning. The presence of differential item functioning can be problematic as it can preclude the ability to make direct comparisons of raw scores of individuals across groups. This problem occurs because items associated with differential item functioning are endorsed systematically differently from individuals of different groups, despite individuals being on the same level of the latent construct. After performing a series of partial invariance tests of item intercepts, we found that the SAD, PD, and GAD subscales of the RCADS were invariant across all metric parameters, including item intercepts. This finding suggests that raw scores on these three anxiety subscales may be compared directly across African American and Caucasian youth given that these scores reflect youths’ level on the latent construct in the same way across groups. The remaining anxiety RCADS subscales contained the following noninvariant items: a social phobia item (*I feel worried when I think someone is angry with me*), and three OCD items (*I can’t*

seem to get bad or silly thoughts out of my head; I have to do some things over and over again (like washing my hands, cleaning or putting things in a certain order); I have to do some things in just the right way to stop bad things from happening). One MDD item was also found to have noninvariant intercepts (*Nothing is much fun anymore*).

With respect to these five items associated with differential item functioning, two points are worth noting. First, it is important to make clear that although we found these five items associated with differential item functioning across African American and Caucasian youth, the present study does not answer the question of *why* these groups differed systematically on these items. At this point, we may only speculate. For instance, it is possible that the noninvariance associated with depression Item 6 (*Nothing is much fun anymore*) may be related to differences in activity engagement styles between Caucasian and African American youth—even among those who are depressed. It is also possible that the noninvariance associated with social phobia Item 8 (*I feel worried when I think someone is angry with me*) is related to differences between Caucasian and African American youth with respect to how they react to and process situations when others are angry at them. And finally, it is possible that the noninvariance associated with three OCD items (OCD Item 23, *I can’t seem to get bad or silly thoughts out of my head*; OCD Item 42, *I have to do some things over and over again (like washing my hands, cleaning or putting things in a certain order)*; OCD Item 44, *I have to do some things in just the right way to stop bad things from happening*) may represent systematic differences in cultural norms (e.g., superstitions) and behavioral reaction styles between Caucasian and African American youth—similar to M. T. Williams, Turkheimer, Schmidt, and Oltmanns’s (2005) finding of differential item functioning between African American and Caucasian adults for eight items of the Padua Inventory for Obsessive Compulsive Disorder, specifically those related to contamination fears and checking.

Further research, guided by theoretically based hypotheses pertaining to *why* these items are associated with differential performance—perhaps also including manipulations to these items’ content to test noted hypotheses—is thus needed to determine the actual reasons *why* these items are associated with noninvariance across these ethnic groups. Second, although some researchers view the presence of differential item functioning as problematic and recommend discarding such items, it has been argued that full measurement invariance is overly stringent (Millsapp & Kwok, 2004), and that comparison of latent factor means (while controlling for differential item functioning) is still feasible. In fact, several researchers have contended that it is preferable to retain such items over discarding the differentially performing items or abandoning the scale completely (Cheung & Rensvold, 1999), particularly when the criteria are met for

partial measurement invariance (Byrne et al., 1989). It has been suggested that differential responding can actually be used as a source of important cross-cultural information (Cheung & Rensvold, 1999) leading to both research and clinical advances, such as better informing the etiological and treatment research specific to various subgroups. Given that these partial metric invariance criteria were met in the present study, we retained all 47 RCADS items.

Identification of noninvariant item intercepts warranted the provision of ethnic-specific normative data to aid in the accurate interpretation of scale scores through the derivations of *T*-scores. Given our identification of items associated with differential item functioning across the African American and Caucasian youth, these normative data were based on age, gender and major ethnic groups. A comparison of rates of clinical and borderline elevations revealed differences when different sets of normative information were used. Normative data were also provided for the total sample, which should be used for determination of subscale elevations in children belonging to ethnicities other than Caucasian or African American.

Researchers often assume measurement invariance when making cross-group comparisons based on raw scores without first explicitly examining relevant measurement invariance parameters (Borsboom, 2006). Relatively few studies exist that have explicitly investigated the contribution of measurement bias in prevalence rates and symptom expression across youth samples (Meredith, & Teresi, 2006). Without such measurement invariance tests, it is difficult to know the degree to which reported statistics are interpretable given that the same raw score from individuals belonging to different groups does not reflect the same position on the latent construct when differential item functioning is present (Hofmans, Pepermans, & Loix, 2009). In fact, Borsboom 2006 has cautioned that findings concerning group differences should be interpreted with caution when diagnoses are based on instruments with uninvestigated measurement invariance. Unfortunately, epidemiological research has also often reported differences across Caucasian and African American youth with respect to prevalence rates and symptom severity of depression, social phobia, and OCD. Many of these studies have used the presence of diversity within normative samples as being evidence of measurement equivalence; however, these are not adequate tests of measurement invariance. We encourage future research to conduct more explicit tests of measurement invariance to better understand these issues.

Although research focused on adolescent group differences in psychopathology is sparse, the majority of extant studies have demonstrated equivocal results (e.g., (Beidel, Turner, & Morris, 1999); Beidel, Turner, Hamlin, & Morris, 2000; McLaughlin, Hilt, & Nolen-Hoeksema,

2007). Such inconsistent findings may not be surprising, particularly given the lack of metric invariance tests conducted on these measures. Differences previously observed may in part be attributed to differential item functioning that has gone unexamined (and thus undetected). In one of the few available studies that specifically examined differential item functioning across ethnic groups, differential item functioning was identified for six items on the Center for Epidemiologic Studies Depression Scale, with African American youth endorsing somatic symptoms more frequently than affective symptoms compared with Caucasians (Iwata, Turner, & Lloyd, 2002). Until now, no studies have been conducted that have examined ethnic measurement bias issues in OCD assessment in youth samples, and the manifestation of OCD in African Americans is not well understood (K. E. Williams, Chambless, & Steketee, 1998). Interestingly, Muris & Meesters (2002) found that the RCADS OCD items did not load on the expected factor in their sample and thus recommended discarding the OCD scale altogether. However, since Muris et al. sample was based on 1748 youth from the Netherlands, it is possible that differential item functioning may account for the OCD factor not being supported.

Studies investigating similar issues in adult samples, however, have identified noninvariance of OCD measures at various levels when comparing responses of African American and Caucasians (Thomas, Turkheimer, & Oltmanns, 2000; Ritsher, Stuenkel, Hellman, & Guardino, 2002). In one such study (M. T. Williams et al., 2005), the authors identified differential item functioning on eight items of the Padua Inventory for Obsessive Compulsive Disorder, specifically those related to contamination fears and checking. Given that the RCADS OCD scale was associated with the greatest degree of differential item functioning, further research should be conducted in this area to understand how OCD and its markers differentially relate to African American and Caucasian youths.

Also notable is that the present findings suggest that metric invariance may also differ as a function of specific disorders within a disorder class. For example, some anxiety subscales (i.e., OCD, SP) were associated with differential item functioning whereas other anxiety subscales were not (i.e., SAD, PD, GAD). These findings demonstrate that measurement invariance cannot simply be assumed across an entire disorder class and that item performance can differ even across similar and related problem areas. Future research should also be conducted to understand reasons for noninvariant items across African American and Caucasian children in these particular problem domains. Understanding reasons for differential reporting of symptoms by problem domain can help us better understand etiology and result in improved treatment approaches that are more culturally sensitive.

In a similar fashion, it was apparent that mean scores obtained on the depression, GAD, OCD, SAD, and SP subscales in the current study were on average lower than those obtained in the original RCADS normative sample in Hawaii (Chorpita et al., 2000). Our sample, however, obtained higher scores than the Hawaiian sample on the panic disorder subscale on average. The reason for these group differences is unknown and no conclusions about actual group differences on these latent constructs can be made until invariance examinations are conducted across these groups (Meredith, 1993). The differences in score obtained across these samples, however, speak further to the importance of region-specific normative data and psychometric exploration in this new sample of Southern youth. This is particularly salient in regard to clinical and research endeavors conducted in the Southeast, given that treatment decisions and scientific conclusions are likely to be considerably different when applying regional norms.

Despite the noted limitations and areas for future research, the present study extended support for the RCADS to a new population of U.S. children and adolescents. We hope that the newly available normative data specific to youth from this (Southeastern) region of the United States may lead to more precise inferences drawn regarding, for example, whether youth are clinically elevated on the various anxiety and depression RCADS dimensions relative to their specific reference group. Since sample size limitations precluded analysis of other ethnic groups represented in the data set (e.g., Asian and Hispanic), normative data for the *total sample* should be used for these and all other ethnic groups that are not African American or Caucasian. In the end, the information provided in this study may not only inform scientific research—such as comparing rates of psychopathology across groups—but may also have direct clinical utility and application by improving the accuracy of diagnostic determinations and thus treatment planning decisions made at the individual patient level. To assist in this process, we have developed a scoring program that scores and reports *T*-score data of *individual* RCADS forms. This (Excel-based) program is available for free download at the aforementioned webpage. This will hopefully inform clinical practice, particularly in the South, by assisting in the interpretation of scale score responses from youth from this large region of the United States.

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

Notes

1. Because of both ambiguity in language found in the literature and the common omission of researchers to consider measurement invariance issues when comparing scores across groups, it is worth repeating that when researchers indicate that they are interested in examining the degree to which “scores” differ across groups, this is actually seldom the case. Rather, in the majority of cases, researchers are likely actually interested in examining the degree to which groups fall at different levels of the *underlying construct* targeted by the scores. Indeed, because of issues related to measurement invariance (such as differential item functioning), it is possible for two groups to fall at significantly different locations on the underlying latent construct while (a) having the exact same raw scores across groups and also (b) having different raw scores across groups—depending on their respective reporting styles on the measure(s) used to “tap” the underlying construct(s).
2. Item 6 (Depression), *Nothing is much fun anymore*; Item 8 (Social Phobia), *I feel worried when I think someone is angry with me*; Item 23 (OCD), *I can't seem to get bad or silly thoughts out of my head*; Item 42 (OCD), *I have to do some things over and over again (like washing my hands, cleaning or putting things in a certain order)*; Item 44 (OCD), *I have to do some things in just the right way to stop bad things from happening*.

References

- American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders* (4th ed., text revision). Washington, DC: Author.
- Angold, A., & Costello, J. E. (1999). Comorbidity. *Journal of Child Psychology and Psychiatry & Allied Disciplines*, *40*, 57-87.
- Arbuckle, J. L. (1995). *AMOS 6.0 user's guide*. Chicago, IL: SPSS.
- Beidel, D. C., Turner, S. M., & Morris, T. L. (1999). Psychopathology of childhood social phobia. *Journal of the American Academy of Child and Adolescent Psychiatry*, *26*, 759-765.
- Beidel, D. C., Turner, S. M., Hamlin, K., & Morris, T. L. (2000). The Social Phobia and Anxiety Inventory for Children (SPAI-C): External and discriminative validity. *Behavior Therapy*, *31*, 75-87.
- Bentler, P. M. (1990). Comparative fit indices in structural models. *Psychological Bulletin*, *107*, 238-246.
- Berkman, L. F., Berkman, C. S., Kasl, S., Freeman, D. H., Leo, L., Ostfield, A. M., . . . Brody, J. A. (1986). Depressive symptoms in relation to physical health and functioning in the elderly. *American Journal of Epidemiology*, *124*, 372-388.
- Birmaher, B., Brent, D., 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*, 1230-1236.

- Bittner, A., Egger, H., Erkanli, A., Costello, E., Foley, D., & Angold, A. (2007). What do childhood anxiety disorders predict? *Journal of Child Psychology and Psychiatry*, *48*, 1174-1183.
- Borsboom, D. (2006). When does measurement invariance matter? *Medical Care*, *44*, 176-181.
- Boyd, R., Ginsburg, G., Lambert, S., Cooley, M., & Campbell, K. (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*, 1188-1196.
- Brown, T. (2006). *Confirmatory factor analysis for applied research*. New York, NY: Guilford Press.
- Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen, J. S. Long (Eds.), *Testing structural equation models* (pp. 136-162). Newbury Park, CA: Sage.
- Bryant, F. B., & Yarnold, P. R. (1995). Principal-components analysis and exploratory and confirmatory factor analysis. In L. G. Grimm & P. R. Yarnold (Eds.), *Reading and understanding multivariate statistics* (pp. 99-136). Washington, DC: American Psychological Association.
- Byrne, B. M., Shavelson, R. J., & Muthén, B. (1989). Testing for the equivalence of factor covariance and mean structures: The issue of partial measurement invariance. *Psychological Bulletin*, *105*, 456-466.
- Carter, M., Miller, O., Sbrocco, T., Suchday, S., Lewis, E. (1999). Factor structure of the Anxiety Sensitivity Index among African-American college students. *Psychological Assessment*, *11*, 525-533.
- Cheung, G. W., & Rensvold, R. B. (1999). Testing factorial invariance across groups: A reconceptualization and proposed new method. *Journal of Management*, *25*, 1-27.
- Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling*, *9*, 233-255.
- Chorpita, B. F., Moffitt, C. E., & Gray, J. (2005). Psychometric properties of the Revised Child Anxiety and Depression Scale in a clinical sample. *Behaviour Research and Therapy*, *43*, 309-322.
- Chorpita, B. F., Yim, L. M., Moffitt, C. E., Umemoto, L., & Francis, S. (2000). Assessment of symptoms of *DSM-IV* anxiety and depression in children: A Revised Child Anxiety and Depression Scale. *Behaviour Research and Therapy*, *38*, 835-855.
- Clark, L. A., & Watson, D. (1991). Tripartite model of anxiety and depression: Psychometric evidence and taxonomic implications. *Journal of Abnormal Psychology*, *100*, 316-336.
- Costello, E. J., Compton, S. N., Keeler, G. P., & Angold, A. (2003). Relationships between poverty and psychopathology: A natural experiment. *Journal of the American Medical Association*, *290*, 2023-2029.
- 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*, 837-844.
- de Ross, R., Gullone, E., & Chorpita, B. F. (2002). The Revised Child Anxiety and Depression Scale: A psychometric investigation with Australian youth. *Behaviour Change*, *19*, 90-101.
- Dunlop, D. D., Song, J., Lyons, J. S., Manheim, L. M., & Chang, R. W. (2003). Racial/ethnic differences in rates of depression among preretirement adults. *American Journal of Public Health*, *93*, 1945-1952.
- Egger, H., Costello, E., & Angold, A. (2003). School refusal and psychiatric disorders: A community study. *Journal of the American Academy of Child and Adolescent Psychiatry*, *42*, 797-807.
- Ginsburg, G. S., Riddle, M. A., & Davies, M. (2006). Somatic symptoms in children and adolescents with anxiety disorders. *Journal of the American Academy of Child and Adolescent Psychiatry*, *45*, 1179-1187.
- Grover, R. L., Ginsburg, G. S., & Ialongo, N. (2007). Psychosocial outcomes of anxious first graders: A seven-year follow-up. *Depression and Anxiety*, *24*, 410-420.
- Hodges, K. (1990). Depression and anxiety in children: A comparison of self-report questionnaires to clinical interview. *Psychological Assessment*, *2*, 376-381.
- Hofmans, J., Pepermans, R., & Loix, E. (2009). Measurement invariance matters: A case made for the ORTOFIN. *Journal of Economic Psychology*, *30*, 667-674.
- Iwata, N., Turner, R. J., & Lloyd, D. A. (2002). Race/ethnicity and depressive symptoms in community-dwelling young adults: a differential item functioning analysis. *Psychiatry Research*, *110*, 281-289.
- Jones-Webb, R., & Snowden, L. (1993). Symptoms of depression among blacks and whites. *American Journal of Public Health*, *83*, 240-244.
- Lahey, B. B., Rathouz, P. J., Van Hulle, C., Urbano, R. C., Krueger, R. F., Applegate, B., . . . Waldman, I. D. (2008). Testing structural models of *DSM-IV* symptoms of common forms of child and adolescent psychopathology. *Journal of Abnormal Child Psychology*, *36*, 187-206.
- Lambert, S., Cooley-Quille, M., Campbell, K., Benoit, M., & Stansbury, R. (2004). Assessing anxiety sensitivity in inner-city African-American children: Psychometric properties of the Childhood Anxiety Sensitivity Index. *Journal of Clinical Child and Adolescent Psychology*, *33*, 248-259.
- Little, R. J. A., & Rubin, D. B. (1987). *Statistical analysis with missing data*. New York, NY: Wiley.
- McDonald, R. (1999). *Test theory: A unified treatment*. Mahwah, NJ: Lawrence Erlbaum.
- McLaughlin, K. A., Hilt, L. M., & Nolen-Hoeksema, S. (2007). Racial/ethnic differences in internalizing and externalizing symptoms in adolescents. *Journal of Abnormal Child Psychology*, *35*(5), 801-816.
- Medsker, G. J., Williams, L. J., & Holahan, P. J. (1994). A review of current practices for evaluating causal models in organizational

- behavior and human resources management research. *Journal of Management*, 20, 439-464.
- Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika*, 58, 525-543.
- Meredith, W., & Teresi, J. A. (2006). An essay on factorial invariance. *Medical Care*, 44, S69-S77.
- Meyer, T. J., Miller, M. L., Metzger, R. L., & Borkovec, T. D. (1990). Development and validation of the Penn State Worry Questionnaire. *Behaviour Research and Therapy*, 28, 487-495.
- Milsap, R. E., & Kwok, O. M. (2004). Evaluating the impact of partial factorial invariance on selection in two populations. *Psychological Methods*, 9, 93-115.
- Muris, P., & Meesters, C. (2002). Symptoms of anxiety disorders and teacher-reported school functioning of normal children. *Psychological Reports*, 91, 588-590.
- Neal, A. M., Lilly, R. S., & Zakis, S. (1993). What are African-American children afraid of? *Journal of Anxiety Disorders*, 7, 129-139.
- Nguyen, H. T., Kitner-Triolo, M., Evans, M. K., & Zonderman, A. B. (2004). Factorial invariance of the CES-D in low socioeconomic status African-Americans compared with a nationally representative sample. *Psychiatry Research*, 126, 177-187.
- Ritsher, J. B., Stuenkel, E. L., Hellman, F., & Guardino, M. (2002). Internal validity of an anxiety disorder screening instrument across five ethnic groups. *Psychiatry Research*, 111, 199-213.
- Samaan, R. A. (2000). The influences of race, ethnicity, and poverty on the mental health of children. *Journal of Health Care for the Poor and Underserved*, 11, 100-110.
- Scott, E. L., Eng, W., & Heimberg, R. G. (2002). Ethnic differences in worry in a nonclinical population. *Depression and Anxiety*, 15, 79-82.
- Seligman, L. D., & Ollendick, T. H. (1998). Comorbidity of anxiety and depression in children and adolescents: An integrative review. *Clinical Child and Family Psychology Review*, 1, 125-144.
- Seligman, L. D., Ollendick, T. H., Langley, A. K., & Baldacci, H. B. (2004). The utility of measures of child and adolescent anxiety: A meta-analytic review of the RCMAS, STAIC, and CBCL. *Journal of Clinical Child and Adolescent Psychology*, 33, 557-565.
- Shahar, G., Henrich, C. C., Winokur, A., Blatt, S. J., Kuperminc, G. P., & Leadbeater, B. J. (2006). Self-criticism and depressive symptoms interact to predict middle school academic achievement. *Journal of Clinical Psychology*, 62, 147-155.
- Silverman, W. K., Ginsburg, G. S., & Goedhart, A. W. (1999). Factor structure of the Childhood Anxiety Sensitivity Index. *Behavior Research and Therapy*, 37, 903-917.
- Silverman, W. K., & Ollendick, T. H. (2005). Evidence-based assessment of anxiety and its disorders in children and adolescents. *Journal of Clinical Child and Adolescent Psychology*, 34, 380-411.
- Smith, L. C., Friedman, S., & Nevid, J. (1999). Clinical and socio-cultural differences in African-American and European American patients with panic disorder and agoraphobia. *Journal of Nervous and Mental Disease*, 187, 549-560.
- Stark, K., & Laurent, J. (2001). Joint factor analysis of the Children's Depression Inventory and the Revised Children's Manifest Anxiety Scale. *Journal of Clinical Child Psychology*, 30, 552-567.
- Steenkamp, J., & Baumgartner, H. (1998). Assessing measurement invariance in cross-national consumer research. *Journal of Consumer Research*, 25, 78-90.
- Steiger, J. H. (1990). Structural model evaluation and modification: an interval estimation approach. *Multivariate Behavioral Research*, 25, 173-180.
- Tallis, F., Eysenck, M., & Mathews, A. (1992). A questionnaire for the measurement of nonpathological worry. *Personality and Individual Differences*, 13, 161-168.
- Thomas, J., Turkheimer, E., & Oltmanns, T. F. (2000). Psychometric analysis of racial differences on the Maudsley Obsessional Compulsive Inventory. *Assessment*, 7(3), 247-258.
- Tucker, L. R., & Lewis, C. (1973). A reliability coefficient for maximum likelihood factor analysis. *Psychometrika*, 38, 1-10.
- Tylee, A., & Gandhi, P. (2005). *The importance of somatic symptoms in depression in primary care. Primary Care Companion to the Journal of Clinical Psychiatry*, 7, 167-176.
- U.S. Census Bureau. (2010, August 16). *State & county QuickFacts: Mississippi*. Retrieved from <http://quickfacts.census.gov/qfd/states/28000.html>
- 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 Method*, 3, 4-70.
- Watson, D. (2005). Rethinking the mood and anxiety disorders: a quantitative hierarchical model for DSM-V. *Journal of Abnormal Psychology*, 114, 522-536.
- Weems, C. F., Cost, N. M., Watts, S. E., Taylor, L. K., & Cannon, M. F. (2007). Cognitive errors, anxiety sensitivity, and anxiety control beliefs: their unique and specific associations with childhood anxiety symptoms. *Behavior Modification*, 31, 174-201.
- Williams, D. R., Gonzalez, H. M., Neighbors, H., Nesse, R., Abelson, J. M., Sweetman, J., & Jackson, J. S. (2007). Prevalence and distribution of major depressive disorder in African-Americans, Caribbean blacks, and non-Hispanic whites: Results from the National Survey of American Life. *Archives of General Psychiatry*, 64, 305-315.
- Williams, K. E., Chambless, D. L., & Steketee, G. (1998). Behavioral treatment of obsessive-compulsive disorder in African-Americans: Clinical issues. *Journal of Behavior Therapy and Experimental Psychiatry*, 29, 163-170.
- Williams, M. T., Turkheimer, E., Schmidt, K. M., & Oltmanns, T. F. (2005). Ethnic identification biases responses to the Padua Inventory for Obsessive-Compulsive Disorder. *Assessment*, 12, 174-185.