Journal of Counseling Psychology
2010, Vol. 57, No. 3, 274–289

0022-0167/10/$12.00 DOI: 10.1037/a0019357

A Multicultural Assessment of Adolescent Connectedness: Testing Measurement Invariance Across Gender and Ethnicity

Michael J. Karcher and Daniel Sass
University of Texas at San Antonio

Counselors, psychologists, and evaluators of intervention programs for youth increasingly view the promotion of connectedness as an important intervention outcome. When evaluating these programs, researchers frequently test whether the treatment effects differ across gender and ethnic or racial groups. Doing so necessitates the availability of culturally and gender-invariant measures. We used the Hemingway: Measure of Adolescent Connectedness to estimate the factor structure invariance and equality of means across gender and 3 racial/ethnic groups with a large sample of middle school adolescents. From a practical perspective, the 10-scale model suggested factor structure invariance across gender and racial or ethnic (i.e., African American, Caucasian, and Latina/o) groups of adolescents. However, tests for partial invariance revealed some group difference on the factor loadings and intercepts between gender and ethnic/racial groups. When testing for mean equivalence, girls reported higher connectedness to friends, siblings, school, peers, teachers, and reading but lower connectedness to their neighborhoods. Caucasians reported higher connectedness to their neighborhoods and friends but lower connectedness to siblings than African Americans and Latinos. African Americans reported the highest connectedness to self (present and future) but lowest connectedness to teachers. Latinos reported the lowest connectedness to reading, self-in-the-present, and self-in-the-future. Overall, this study reveals racial/ethnic and gender mean differences on several connectedness subscales and suggests the Hemingway subscales are, from a practical perspective, invariant across gender and ethnicity and therefore appropriate for most assessment and evaluation purposes.

Keywords: factor analysis, invariance, school connectedness, adolescence, cross-cultural

Supplemental material: http://dx.doi.org/10.1037/a0019357.supp

> As important as it is to reduce or eliminate problems among children and adolescents, it is just as important to help them thrive and form positive connections to the larger world. (Evans et al., 2005, p. 498)

Promoting adolescents' connectedness to school, their families, and the future has become the goal of many school-based prevention and positive youth development programs (Garringer, 2009; Grossman & Bulle, 2006; Roth & Brooks-Gunn, 2003). Yet program developers and evaluators have lacked the measures of connectedness with evidence of racial/ethnic or gender equivalence required to justify the use of these measures for statistical comparisons between these groups for the purpose of research, evaluation, or diagnosis. To assist them, in the present study, we examine the factor structure equivalence of a 10-scale measure of adolescent connectedness and compare means across racial/ethnic groups and gender.

Estimating factor structure equivalence for readily available measures of adolescent connectedness is important because mean differences in levels of family, school, and social connectedness between adolescent boys and girls and across racial and ethnic groups are frequently reported in the literature (Bonny, Britto, Klostermann, Hornung, & Slap, 2000; Lee & Robbins, 2000; McNeely, Nonnemaker, & Blum, 2002; Resnick et al., 1997; Whitlock, 2006). If we assume that the subscales of connectedness used in these studies were invariant across gender and ethnicity/race, then the reported mean differences simply indicate that one gender or ethnic/racial group reported a higher level of connectedness on one or more connectedness subscales. However, there is some evidence that the meaning (or operational definition) of family, school, and social connectedness also may differ across gender and ethnicity at the construct level (e.g., Jacobson & Rowe, 1999). In this case, the underlying construct being measured may vary considerably across groups, thereby rendering these mean comparisons invalid. Determining the validity of such comparisons requires research on the meaning of connectedness and tests of scale construct validity and measurement invariance (Barber & Schluterman, 2008).

What is connectedness? Townsend and McWhirter (2005) reviewed the counseling literature on connectedness and concluded that it reflects "when a person is actively involved with another

Michael J. Karcher, Department of Counseling, University of Texas at San Antonio; Daniel Sass, Department of Educational Psychology, University of Texas at San Antonio.

We thank Michael J. Nakkula and Hardin H. L. Coleman for their helpful comments on a draft of this article. We acknowledge Blane McCann's role in making this study possible; Doug Atkinson's help with data management; and the work of Claytie Davis, III; John Harris; and Michael J. Nakkula, who assisted with data collection, coding, and concept feedback on the first version of the measure. We ascribe the name Hemingway to Don Carli, a Michigan teacher and natural mentor, who altered one student's career path when he said, "You write like Hemingway."

Correspondence concerning this article should be addressed to Michael J. Karcher, Department of Counseling, University of Texas at San Antonio, 501 West Durango Boulevard, Suite 4.316, San Antonio, TX 78207. E-mail: michael.karcher@utsa.edu

person, object, group, or environment, and that involvement promotes a sense of comfort, well-being, and anxiety reduction" (p. 193). This definition is consistent with research literature on both adults and adolescents that characterizes connectedness using indicators of behaviors and affect in (e.g., "being close to people" [Resnick et al., 1997, p. 825] and "feeling a part of" [Barber & Schluterman, 2008, p. 210]) specific contexts and relationships. Yet, only in the adolescent literature has connectedness regularly been differentiated into the domains of school, familial, and social connectedness. There has been little work done to determine whether these ecological domains should be measured separately—that is, whether scales of separate ecological domains demonstrate discriminant validity. Nor has research considered whether the behavioral and affective components of adolescent connectedness manifest similarly for boys and girls and across different cultural groups.

Hemingway: Measure of Adolescent Connectedness

One measure of adolescent connectedness, the Hemingway: Measure of Adolescent Connectedness (Karcher, 2005), may prove useful to researchers, counselors, and evaluators who seek to compare mean differences in adolescent connectedness between genders and across ethnic/racial groups. First, it has utility for program evaluation, because its subscales for connectedness to school, parents, and friends reflect outcomes commonly targeted by youth development programs (Roth, Brooks-Gunn, Murray, & Foster, 1998). Second, studies using earlier versions of the Hemingway have revealed evidence of predictive validity. For example, studies reveal negative relationships between several connectedness subscales and forms of risk taking (both violence and substance use) that are commonly targeted by intervention programs (e.g., Karcher, 2002; Karcher & Finn, 2005). Third, the definition of adolescent connectedness reflected in the Hemingway subscales (as affect and action in specific relationships, contexts, and activities) is consistent with the general view in the field (Barber & Schluterman, 2008; Townsend & McWhirter, 2005).

Gender Differences

Consistent with prior research using other scales, studies using the Hemingway subscales have reported mean differences between boys and girls (with girls scoring higher than boys on most subscales; see Karcher, 2001), but they also reveal different program impacts on several domain-specific connectedness subscales across gender (Herrera, Grossman, Kauh, Feldman, & McMaken, 2007). For example, Karcher (2008) found that the effects of school-based mentoring on connectedness were different for boys and girls. Boys increased in connectedness to school, whereas girls improved primarily in connectedness to friends and peers.

Ethnic and Racial Group Invariance

Programs and related counseling services provided to ethnic minority youth often target problems or promote developmental competencies derived from research on ethnic majority youth (García Coll et al., 1996). Yet theory and research suggest that minority and majority youth may experience the key intervention processes of such programs differently. For example, one study that used the Hemingway with African American, Caucasian, and Latino youth reported that the effect of a cross-age peer mentoring program on academic achievement was mediated by improvements in connectedness to parents (Karcher, Davis, & Powell, 2002). However, if perceptions of connectedness to parents are culturally specific, as suggested by Cooper (1999), it is unclear whether improved parental connectedness mediates program impacts similarly across these groups.

It is possible that some aspects of a construct differ across gender and cultural groups. Therefore, a measure should not include items that are more reflective of one group's experience than another's. Including such items may result in dissimilar item weighting across groups, making the comparison of these scores impossible. Given the importance of valid mean comparisons, the use of measures that are invariant across comparison groups is critical.

Factor Structure Invariance

Regardless of the method used to estimate construct scores, it is assumed that items function similarly across comparison groups and that items are invariant. When a multi-item scale is used in practice, the observed scores (e.g., scale means) provide equal weight to each item, with each item assumed to make an equal and important contribution to that construct. With factor analysis, a factor score is created on the basis of the unstandardized factor loadings and intercepts; thus, items may contribute unequally to the factor. For noninvariant scales, either the unstandardized factor loadings and/or the intercepts contribute differently to the factor score across groups (see Chen, 2008; Little, 1997; Meredith, 1993; Vandenberg & Lance, 2000). Under these conditions, the equation used to create each group's overall score differs, making mean comparisons ill-advised. Therefore, *factor loading invariance* indicates the relationships (i.e., slopes) between the subscale items and the factor are parallel across groups—that is, the unstandardized factor loadings are equal. This means that for each group, a one-unit change in the item response results in the same increase for both groups on the underlying factor. *Intercept invariance* occurs when the groups have the identical item mean when the factor score is zero (i.e., at the average latent trait score). When factors have invariant factor loadings and intercepts, the regression equations are identical, such that the regression lines completely overlap and item contributions to each factor are equal between the groups. Under these circumstances, the factor scores are created in an identical fashion and comparing means is justifiable.

These equations could differ for a number of reasons. For example, *factor loading noninvariance* could occur when (a) items are translated from one language to another but the definitions and meanings of the concepts differ between groups, (b) results differ simply because of an improper translation, (c) subjects interpret the item(s) differently for various reasons within the same culture (e.g., boys and girls within a culture), and/or (d) certain groups may avoid or use more extreme responses (Chen, 2008). *Intercept noninvariance* could occur because of (a) social desirability reasons or social norms, (b) certain groups displaying a propensity to respond more strongly to an item despite having a comparable latent trait value, and/or (c) certain groups having a different reference point when making statements about themselves (Chen, 2008).

The present study addressed the following research questions. First, does the Hemingway measure provide evidence of factorial validity? Second, is the 10-factor model invariant across different gender and racial/ethnic groups of early adolescents? Third, assuming measurement invariance, do the 10 observed means differ across the gender and racial/ethnic groups, as has been reported previously in the literature?

Method

Sample

Data were collected from 4,263 students attending the six middle schools in a Midwestern city of approximately 90,000 people. The U.S. Census data for the year 2000 reveals that 79% of the citizens of this city were Caucasian, 8% were African American, 10% were Latina/o, and 3% self-identified as "other race," and the city's median family income was $41,900. The study data were collected by a school district that regularly uses this survey to track school climate. Nearly 90% of the 4,741 students enrolled in the six middle schools in this city participated. The data were collected by the director of the Office of K–12 Instruction in February of 2003 using a university-approved passive consent approach, which explained that children could choose not to participate, that the data would be anonymous, and that the data would subsequently be provided to us for the purpose of conducting these data analyses.

Of the 4,263 youth sampled, singletons (subjects without siblings) were excluded because of their appropriately missing data on Connectedness to Siblings subscale items.[1] With no responses to Connectedness to Siblings subscale items, those youth would have been dropped from the invariance tests, and we opted not to impute or estimate this data. Unfortunately, we did not explicitly ask respondents whether they had any siblings. To infer singleton status, we took a liberal approach and designated singletons as anyone who was missing two or more Connectedness to Siblings subscale items (n = 294; 7.5%). The responses of these individuals were then removed from the data—that is, we did not attempt to impute what the singletons' sibling connectedness might have been had they not been singletons. Although we have conducted tests of invariance between those designated as singletons and siblings, we consider these analyses more dubious given the uncertain nature of each individual's actual sibling status. Thus, these results are available in the online supplement but are not reported here. In addition, the responses of another 336 subjects were removed because of missing gender and race/ethnicity data.

Table 1 provides the usable sample demographics by grade, gender, and ethnicity for the remaining 3,633 subjects of interest. Tests of gender invariance used this entire sample, whereas invariance tests across ethnicity/race included only African American, Caucasian, or Latino youth. Therefore, 305 subjects were excluded because the sample sizes for their ethnic/racial (i.e., Asian, biracial, and other) groups were inadequate for the analyses, which results in a total useable sample size of 3,328 for these analyses.

Of the 3,633 subjects of interest, the majority lived with both parents (n = 2,225, 61.2%); the remainder of the sample lived with their mother only (n = 910, 25.0%), with their father only (n = 171, 4.7%), or in an alternative living situation (e.g., foster care, grandparents; n = 184, 5.1%). The remaining subjects (n = 143, 3.9%) did not report their living arrangements. The sample appears comparable to the Census 2000 population data for this city.

Missing Data

Missing data at the item level were treated using multiple imputations (MI) via the expected maximization algorithm and the Markov chain Monte Carlo algorithm within LISREL (see du Toit & du Toit, 2001, pp. 387–388). This procedure essentially used random draws or data sets from a multivariate normal probability distribution via Markov chains, with the new parameter estimates recalculated after every draw using the expected maximization algorithm. After an initial burn-in period, the final solution is a data set that represents the average simulated values over the 500 draws. Default values were used, with the exception of increasing the number of draws from 200 to 500 to ensure stable and accurate results.

MI, rather than full information maximum likelihood (FIML) estimation, was used because commonly reported fit indices (e.g., normed fit index, nonnormed fit index, comparative fit index [CFI], goodness-of-fit index) are unavailable when executing FIML given that the chi-square test statistic for the independence (or null) model is unavailable in closed form within LISREL. However, to evaluate the consistency between MI and FIML estimates, we compared MI parameter estimates (i.e., factor loadings, interfactor correlations, available model fit statistics, etc.) with the available FIML estimates. These estimates were nearly identical across both missing data treatment methods. This might be anticipated given that only 2.34% of total observations were missing. Given the percentage of missing data, the model fit statistics should be relatively unbiased (Davey, Savla, & Luo, 2005).

Measures

The Hemingway: Measure of Adolescent Connectedness (final version, Karcher, 2005; see Appendix) self-report survey consists of 57 items designed to measure adolescents' degree of caring for and involvement in specific relationships, contexts, and activities. There are 10 four- to six-item subscales (see Figure 1). Eight of the 10 subscales include a reverse-coded item. All use a response range from 1 = *not at all true* to 5 = *very true*. Once the negatively worded item responses (Items 2, 7, 13, 18, 26, 30, 34, 45, and 51 in the Appendix) are reverse coded, raw scores or factor scores can be created by taking the average of the items used to measure the 10 subscales: Connectedness to Neighborhood, Connectedness to Friends, Connectedness to Self-in-the-Present, Connectedness to Parents, Connectedness to Siblings, Connectedness to School, Connectedness to Peers, Connectedness to Teachers,

[1] Although missing singleton data are arguably neither missing completely at random nor missing at random (see Rubin, 1976), one could view nonresponse by singletons on the Connectedness to Siblings items as appropriately missing (Marshall et al., 2001), such that if singletons had siblings they would respond in a similar fashion as subjects with siblings. It is because some readers may find this conceptually disconcerting that we omitted all singletons from the analyses reported here. However, tests of invariance between singletons and subjects with siblings are available in the online supplement.

Table 1
Observed Frequencies of Participants by Grade, Gender, and Ethnic/Racial Group

Grade and sex	Caucasian	African American	Latina/o	Asian	Biracial	Other
6th						
Boys	444	84	65	8	27	14
Girls	378	66	60	5	19	13
7th						
Boys	420	58	65	15	28	23
Girls	447	63	59	12	47	9
8th						
Boys	437	55	77	6	22	9
Girls	412	44	57	10	18	12
Missing grade data						
Boys	17	6	2	1	1	2
Girls	9	2	1	1	2	1
Total ethnicity	2,564	378	386	58	164	83

Note. The total sample size used for the gender invariance analyses was 3,633, but for the racial/ethnic group invariance analyses the sample was 3,328 because these analyses excluded Asians, biracial, and other subjects as a result of inadequate sample sizes.

Connectedness to Self-in-the-Future (Item 55 is excluded due to prior evidence of poor construct and discriminate validity; see Karcher, 2001), and Connectedness to Reading.

Initially, this survey instrument was developed by asking adolescents in two ethnically diverse focus groups to explain what they thought it meant to be "connected" (Karcher, 2001). The proportion of African American and Latino youth in both focus groups was equal to or greater than their representation in the present study, as both were conducted in more ethnically diverse locations (one was in Texas, the other Massachusetts). The focus groups resulted in the identification of multiple domains of connectedness, and youth helped generate subscale items that reflect caring for and involvement in these domains. For example, the Connectedness to School subscale focuses on the importance youth place on school and how actively they try to be successful in school. The Connectedness to Teachers subscale assesses effort made to get along with teachers and concerns about earning teachers' respect and trust. The Connectedness to Peers subscale assesses feelings about peers and about working with peers. The two self subscales assess present and future-oriented self-esteem. The Connectedness to Self-in-the-Present subscale assesses feelings about current relationships, continuity in behavior across contexts, and an awareness of skills and interests that make them liked by others. The Connectedness to Self-in-the-Future subscale asks about behaviors and qualities that will help them in the future.

Studies using prior versions of the subscales have demonstrated a distinct factor structure, evidence of convergent and discriminant validity, and good one-month test–retest reliability (Karcher, 2001). Karcher (2001) reported a series of five studies that describe construct, item, and subscale development. These studies used the fourth version of the measure, which included eight items that are worded differently than they are in this final version. In those studies, Karcher used exploratory factor analyses and confirmatory factor analyses (CFA) across separate samples to assess factorial validity evidence and compare mean differences across several groups (i.e., genders, teens vs. preteens, delinquent vs. nondelinquent youth). These studies reported the strongest evidence of convergent validity with measures of family connectedness, school connectedness, self-esteem, and future orientation. One-month test–retest reliabilities ranged from $r = .68$ (Connectedness to Self-in-the-Future) to $r = .91$ (Connectedness to Siblings). Cronbach's alpha ranged from weak (αs = .60 and .68 for Connectedness to Peers and Connectedness to Self-in-the-Future, respectively) to strong (αs = .91 and .94 for Connectedness to Reading and Connectedness to Siblings, respectively).

Statistical Analysis Procedures

Invariance analyses. To test factor structure invariance, this study assessed the following model components: (a) factor loadings, (b) intercepts, (c) factor loading residuals, and (d) the variance–covariance matrix of the latent trait factors. To date, a mandatory sequential order to test for first-order factor structure invariance does not exist (see Ployhart & Oswald, 2004; Vandenberg & Lance, 2000). Agreement does exist with regard to testing latent mean equivalence across different groups. Tests of latent or observed score mean equivalence should only be conducted if the unit of measurement (i.e., unstandardized factor loadings) and scale origin (i.e., intercepts) are invariant between groups (Little, 1997; Ployhart & Oswald, 2004).

Testing for factorial invariance. Our tests of first-order factor model invariance started with an examination of model fit for each group (e.g., boys and girls) separately. If adequate model fit was obtained for each group, a test of configural invariance (weighted combination of both samples) was acquired to provide the baseline model to subsequently compare the more restrictive invariance models. The next two models, which test for unstandardized factor loading and intercept invariance, determined whether the preconditions were met to allow for mean comparisons. These invariance tests were critical to assess whether the latent trait scores (i.e., factor score estimates) were created in an identical fashion (see Jöreskog & Sörbom, 1996, pp. 171–173). The final two analyses evaluated whether the measured variable's (i.e., item's) residuals and the factor's variance–covariance matrix were equal. These comparisons are considered optional and of less theoretical interest, and they are not required to compare means.

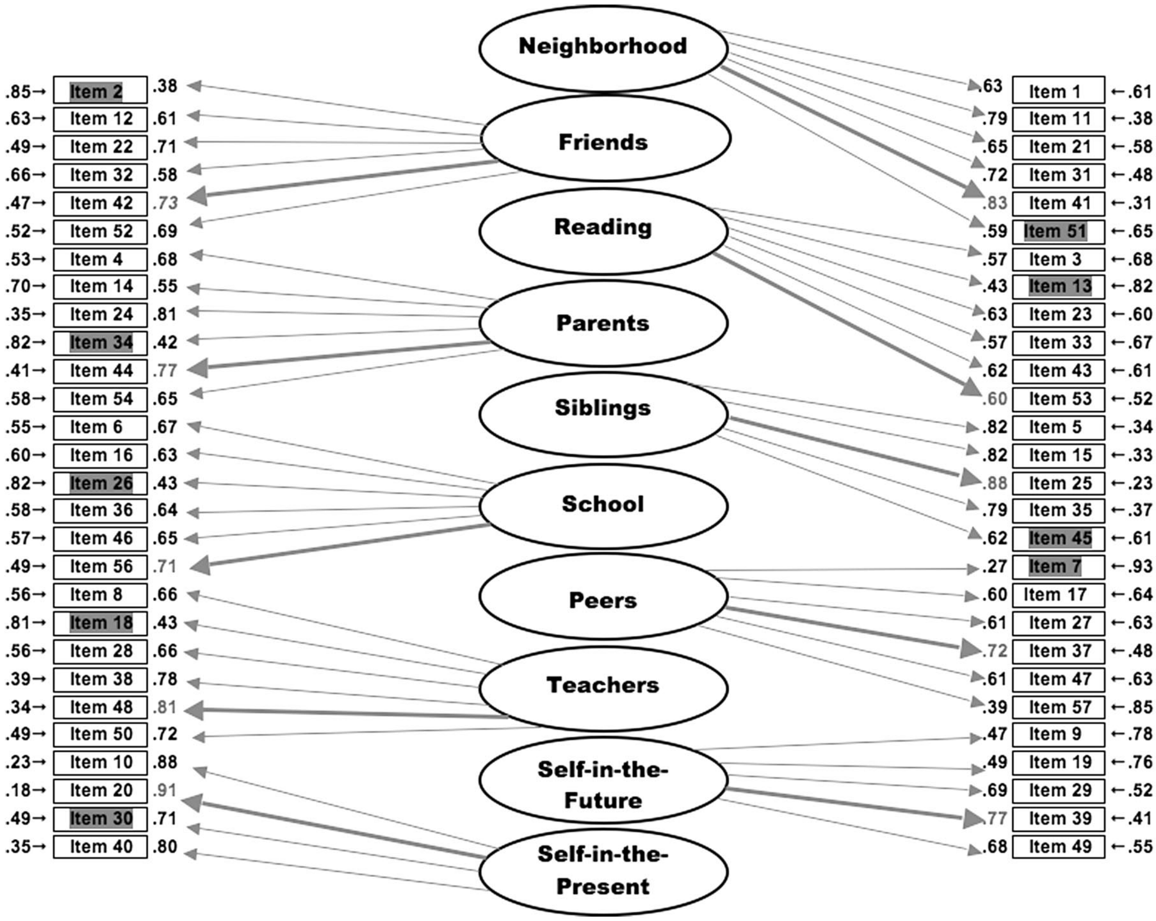

Figure 1. Displays parameter estimates for the completely standardized item factor loadings and residuals for the 10 adolescent connectedness subscales. These parameter estimates complement the interfactor correlations (ϕs) in Table 2. Grayed items are reverse-scored, negatively worded items. Grayed factor loadings indicate parameters fixed to 1 for scale identification.

These models simply test whether the other measurement model components are equal. Note that invariance was tested cumulatively, meaning that the higher order (e.g., intercept) invariance was only tested if the lower order (e.g., unstandardized factor loadings) invariance was met.

Model estimation. Data analysis was conducted with LISREL 8.80 (Jöreskog & Sörbom, 1996) using the covariance matrix and a maximum likelihood estimation procedure. When evaluating the models, we fixed several parameter estimates (i.e., reference indicators) at 1.0 (denoted in figures by bolded arrows and grayed coefficients) to identify the model and set the metric of the factor. The reference indicator item was not selected arbitrarily as the selection of these items can have a considerable influence on the invariance results (French & Finch, 2008). Instead, numerous invariance models were tested to select the item that was most invariant across gender and ethnic/race groups. Each estimated (i.e., freed) standardized factor loading and corresponding residual is provided in Figure 1; the other factor loadings and residuals were fixed at zero (i.e., not estimated). All interfactor correlations (ϕ) were also estimated (see Table 2).

Model identification. To help ensure model identification, we made sure the following conditions were met: (a) A single unstandardized factor loading per factor was set at 1.0, (b) at least three indicators (i.e., items) existed per factor with uncorrelated error terms, and (c) no error terms were correlated. The t-rule was also applied to ensure that each model resulted in an overidentified model that could be estimated (Bollen, 1989). Thus, there were always more known than unknown pieces of information.

Overall model fit criteria. The statistics used to evaluate model fit for each gender and ethnicity/race sample were the minimum fit function chi-square, CFI, root mean square error of approximation (RMSEA), and standardized root mean square residual (SRMR). Description of these model fit statistics can be obtained from Hoyle (1995) and Hu and Bentler (1999). Hu and Bentler described CFI statistics greater than .90 as an "adequate" fit and values greater than .95 (which are preferable to minimize Type I and Type II errors) as a "good" fit. They denoted fit indexes for RMSEA and SRMR values less than .06 and .08, respectively, as "good" and values between .08 and .10 as "mediocre."

Invariance model fit criteria. The problems associated with evaluating model fit when testing for model invariance are well documented (see Chen, Sousa, & West, 2005). Although a chi-square difference test (i.e., likelihood ratio test) allows a statistical

Table 2
Interfactor Correlations (ϕ) Using the Complete Sibling Sample (n = 3,633)

Connectedness subscale	1	2	3	4	5	6	7	8	9	10
1. Neighborhood	.85									
2. Friends	.34	.78								
3. Self-in-the-Present	.36	.47	.76							
4. Parents	.30	.21	.54	.80						
5. Siblings	.26	.20	.38	.61	.89					
6. School	.24	.28	.61	.63	.39	.79				
7. Peers	.43	.48	.66	.51	.39	.67	.68			
8. Teachers	.25	.30	.50	.60	.39	.85	.61	.82		
9. Self-in-the-Future	.31	.31	.71	.58	.43	.73	.56	.68	.75	
10. Reading	.06	.12	.22	.24	.20	.44	.19	.37	.36	.89

Note. These interfactor correlations (off diagonal) complement the standardized factor loadings in Figure 1, with the internal consistency estimates provided on the diagonal.

comparison between nested models, this test presents several statistical problems (Chen, 2007; Marsh & Hocevar, 1985): (a) The chi-square statistic is sensitive to departures from multivariate normality and (b) with complex models and/or large samples, the chi-square (or $\Delta\chi^2$) statistic is nearly always large and statistically significant. For these reasons, the results were interpreted from *practical* (ΔCFI, ΔRMSEA, and ΔSRMR) and *statistical* ($\Delta\chi^2$) model fit perspectives. Three practical model fit statistics (CFI, RMSEA, and SRMR) less sensitive to model complexity and sample size were emphasized. Following Chen's (2007) recommendations based on simulation research, we used the following criteria to determine acceptable model fit: ΔCFI $\leq$ .01, ΔRMSEA $\leq$ .015, and ΔSRMR $\leq$ .03 for tests of factor loading invariance, and ΔCFI $\leq$ .01, ΔRMSEA $\leq$ .015, and ΔSRMR $\leq$ .01 for tests of intercept invariance and residual invariance.

Results

Factorial Validity

Prior to assessing the invariance models, we evaluated evidence of the measure's factorial validity using CFA for the entire sample. The model estimation procedures carried out were identical to the invariance tests, with the exception that the completely standardized solutions, rather than unstandardized solutions, were evaluated. The completely standardized parameter estimates for the entire sample (without singletons, n = 3,633) are provided in Figure 1, which displayed an overall good model fit, $\chi^2(1439)$ = 12,555.58, $p < .0001$, CFI = .964, RMSEA = .051, SRMR = .048. The model fit and modification indices, which indicate minimal cross-loadings, provide strong evidence of factorial validity, as all the items had relatively large estimated standardized factor loadings on their corresponding factors. Except for Item 7, all the standardized factor loadings were greater than .30. One trend worth noting is that reverse-scored items often had smaller estimated standardized factor loadings than did other items. The interfactor correlations are provided in Table 2 to complement the estimated standardized factor loadings in Figure 1. Internal consistency estimates for all students also are reported in the diagonals of Table 2. Gender-specific and ethnic subgroup internal consistency estimates varied slightly between the groups; these results are available in the online supplement.

Tests for Gender Invariance Based on the Practical Fit Indices

Prior to testing for measurement invariance, we estimated the factor models for girls and boys separately. Results revealed a good model fit for both genders, with relatively equal model fit statistics (see Table 3). The configural invariance model also provided a good model fit based on the RMSEA, SRMR, and CFI. The next two models tested whether the unstandardized factor loadings and intercepts were invariant across gender. Model fit results based on the practical indices revealed the ΔRMSEA,

Table 3
Model Fit Statistics Across Gender

Model	χ^2	*df*	$\Delta\chi^2$	Δdf	RMSEA	ΔRMSEA	SRMR	ΔSRMR	CFI	ΔCFI
Boys	7,195.20	1439			.050		.050		.963	
Girls	7,131.94	1439			.052		.051		.962	
Configural	14,327.13	2878			.051		.051		.963	
Factor loadings	14,554.85	2924	227.72	46	.051	.000	.051	.000	.962	−.001
Item intercepts	15,945.82	2980	1,390.97	56	.054	.002	.051	.001	.958	−.004
Item residuals	16,676.92	3036	731.10	56	.054	.001	.051	.000	.955	−.002
Variance/covariance	16,805.40	3091	128.48	55	.054	.000	.053	.002	.955	.000

Note. Sample sizes for boys, girls, and combined were 1,886, 1,747, and 3,633, respectively. All chi-square and change in chi-square values were statistically significant at α = .001. RMSEA = root mean square error of approximation; SRMR = standardized root mean square residual; CFI = comparative fit index.

ΔSRMR, and ΔCFI were always less than .01, thus suggesting that the factor score estimates were created similarly across genders (see Table 3). More specifically, this implies that the linear equations used to create factor score estimates were nearly equivalent for girls and boys. Tests for gender invariance of the item residuals (θ_δ) and the variance–covariance matrix (ϕ) of the latent variables revealed that ΔRMSEA, ΔSRMR, and ΔCFI were consistently small, suggesting that item residuals and the variance–covariance matrix were largely invariant. In short, the practical fit indices, which adjust for model complexity and sample size, suggest factorial invariance was obtained, and therefore gender mean differences can be examined.

Despite the indication of factor structure invariance from a practical standpoint, the statistical model fit index (i.e., $\Delta\chi^2$) was relatively large for some models, suggesting some differences exist. Pursuant to Byrne and Stewart (2006), we conducted additional item-level analyses to better understand the statistically significant change in chi-square and identify whether some items were more equivalent indicators of connectedness across gender than others.

Tests for Gender Invariance From a Traditional (or Statistical) Perspective: Item-Level Tests

To ensure our data met the conditions necessary to compare the latent variables (ξ), we conducted post hoc analyses on the unstandardized factor loadings and intercepts to ascertain the degree of noninvariance from a statistical perspective. As shown in Table 4, several item parameters were statistically significant (i.e., noninvariant) even after applying a Bonferroni adjustment to control for Type I error.

Half of the Connectedness to Friends, Connectedness to Parents, Connectedness to School, and Connectedness to Self-in-the-Present subscale items ($n_i = 3$) were not invariant across gender, typically because of statistically noninvariant intercepts. When viewing Table 4, it may be useful to recall that items on the same factor end with the same digit. For example, items ending in 4, 6, 8, and 10 correspond to the Connectedness to Parents, Connectedness to School, Connectedness to Teachers, and Connectedness to Self-in-the-Present subscales, respectively. Most of these differences appeared relatively small on the basis of the Diff estimates and change in chi-square (see Table 4). Steenkamp and Baumgartner (1998, p. 81) suggested partial invariance tests are only used when modification indices are highly significant (both in absolute magnitude and in comparison with the other items) and the expected parameter changes are substantial. They also encourage researchers to focus on the relative weight between change in chi-square statistics; thus, one could argue that only the intercept differences on Items 32 and 52 are of considerable concern given their change in chi-square values compared with the other change in chi-square values. Unfortunately, no standards exist for estimating the "practical significance" for a specific magnitude of difference in item intercepts and factor loadings given that unstandardized coefficients are being compared. For this reason, we later compare the mean difference effect sizes for the full and partial invariance models to assess the overall impact of these noninvariant items.

Regarding differences in factor loadings, only five of 56 (less than 10%) unstandardized factor loadings were noninvariant between boys and girls. This suggests that, in general, most relationships between items and the overall factor scores do not differ between genders—that is, the factor loadings do not differ for boys and girls. In regard to the five differences, boys displayed larger estimated factor loadings on Items 12 and 14, where both items dealt with trust related to friends (Item 12) or parents (Item 14). Boys also had significantly lower intercepts than girls on these items. As seen in Figure 2, although boys' connectedness to friends (Item 12) increased at a greater rate on this trust-specific item, meaning the unstandardized factor loading was larger, they also had a slightly lower intercept (i.e., average item response when the factor score is zero). Collectively, these results suggest that trust may function differently for boys and girls: Boys report lower levels of trust when their connectedness to friends and parents is low, but the gap reduced as their connectedness increased. Had the item been invariant, the two lines would have overlapped because they would have had the same slope (i.e., equal factor loadings) and intercept. Instead, the greater slope of the regression line for boys seen in Figure 2 represents a stronger relation between the observed variable and the underlying latent construct for boys than for girls.

The three other noninvariant factor loadings were reverse-scored items (Items 34, 18, and 26), with girls having larger unstandardized factor loadings than boys. This implies that for these items, every increase in item response increased the factor

Table 4
Noninvariant Items Based on the $\Delta\chi^2$ Test Across Gender

	Unstandardized estimated values			
Item	Boys	Girls	Diff	$\Delta\chi^2$
		Factor loading		
14	0.65	0.44	0.21	37.94
12	0.88	0.69	0.19	23.81
34	0.53	0.74	−0.21	21.07
18	0.55	0.72	−0.17	16.99
26	0.60	0.94	−0.34	16.31
		Intercept		
32	3.10	3.95	−0.85	156.14
52	3.54	4.20	−0.66	124.81
33	4.03	3.84	0.19	58.63
14	4.46	4.57	−0.11	50.67
21	3.54	3.63	−0.08	42.11
46	3.31	3.41	−0.11	39.91
43	4.19	4.11	0.08	36.97
54	4.46	4.54	−0.08	33.62
12	3.99	4.37	−0.38	31.95
34	3.10	2.97	0.13	27.87
19	4.41	4.55	−0.14	25.10
31	3.42	3.13	0.29	22.78
18	3.34	3.52	−0.18	17.46
53	2.49	3.06	−0.57	16.88
16	2.83	3.25	−0.42	16.43
27	3.63	3.66	−0.04	14.49
26	2.50	2.77	−0.27	13.96
49	3.72	3.84	−0.12	13.10

Note. Diff represents the difference between unstandardized parameter estimates (i.e., factor loadings or intercepts). All parameter estimates were significantly different from each other after a Bonferroni adjustment (BA; $\alpha_{BA} = .05/102 = .00049$) based on the change in chi-square.

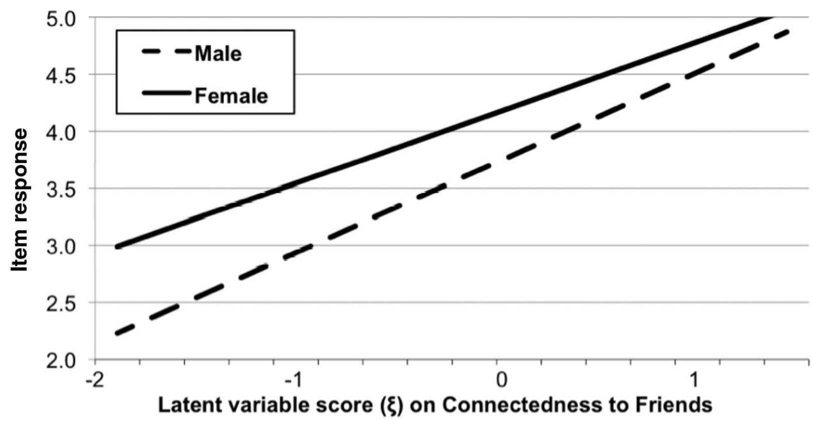

Figure 2. Illustration of item slope and intercept noninvariance. Graph of predicted scores on Item 12 (“I have friends I’m really close to and trust completely”) as a function of scores on the Connectedness to Friends factor. This illustrates the dual noninvariance on this item, which has a significantly different slope (factor loading) and intercept (predicted item score when the factor score is 0) for boys and girls at the same overall levels of connectedness to friends (ξ).

score at a greater rate for girls than for boys. These same items also had significantly different intercepts, with girls having higher intercepts on Items 18 and 26 and boys displaying higher intercepts on Item 34. The girls’ higher intercepts on Items 18 and 26 meant that at the factor score mean (i.e., factor score equal to zero), girls had higher item responses than boys.

In terms of item content, two other items are worth discussing. The intercepts for Items 32 and 52 on the Connectedness to Friends factor displayed a much larger change in chi-square and Diff estimate than did the other noninvariant intercepts in Table 4. Both items measure the importance of talking with friends as indicators of connectedness to friends. On average, girls talked more with their friends than did boys, despite having identical levels of connectedness to friends (and regardless of their level of connectedness to friends; i.e., intercepts were higher but the slopes were parallel).

Tests for Ethnicity Invariance Based on the Practical Fit Indices

Ethnicity analyses revealed that the model fit very well for the Caucasian group. Although the model fit statistics were not as good for the African American and Latina/o groups (see Table 5), the practical fit indices met the standards for adequate model fit. Given this, the configural model was evaluated and demonstrated a good baseline model fit. Subsequent analysis of the unstandardized factor loading and intercept invariance models revealed that the model fit did not differ between the three ethnic/racial groups on the basis of the ΔRMSEA, ΔSRMR, and ΔCFI, thereby justifying a comparison of means. However, once again, the change in chi-square was statistically significant for both factor loading and intercept invariance models, which suggests that in the population differences probably exist; therefore, item analyses were conducted.

Tests for Ethnicity Invariance From a Traditional (or Statistical) Perspective: Item-Level Tests

Tests of individual item differences from a statistical standpoint revealed several differences after we controlled for Type I error using a Bonferroni adjustment. Similar to the gender comparisons, most differences occurred at the intercept level (see Table 6). The only ethnic group differences in factor loadings were on reverse-scored items (i.e., Items 34, 18, and 07).

Table 5
Model Fit Statistics Across the Different Ethnic/Racial Groups

Model	χ^2	*df*	$\Delta\chi^2$	Δdf	RMSEA	ΔRMSEA	SRMR	ΔSRMR	CFI	ΔCFI
Caucasian	9,619.10	1439			.052		.051		.966	
African American	2,876.43	1439			.052		.060		.941	
Latina/o	2,884.47	1439			.053		.062		.925	
Configural	15,380.00	4317			.052		.062		.961	
Factor loadings	15,627.25	4409	247.25	92	.052	.000	.064	.003	.961	−.001
Item intercepts	16,589.82	4521	962.58	112	.053	.001	.064	.000	.959	−.001
Item residuals	18,120.36	4633	1530.53	112	.059	.006	.066	.002	.954	−.005
Variance/covariance	18,401.74	4743	281.38	110	.058	.000	.098	.032	.954	−.001

Note. Sample sizes for Caucasians, African Americans, Latina/os, and combined were 2,564, 378, 386, and 3,328, respectively. All chi-square and change in chi-square values were statistically significant at $\alpha = .001$. RMSEA = root mean square error of approximation; SRMR = standardized root mean square residual; CFI = comparative fit index.

Table 6
Noninvariant Items Based on the $\Delta\chi^2$ Test Across the Ethnic Groups

	Unstandardized estimated values				
Item	Caucasian	African American	Latino	Diff	$\Delta\chi^2$
		Factor loading			
34	0.76	0.24		0.52	28.39
18		0.37	0.65	−0.28	25.07
07		0.24	0.46	−0.22	15.06
		Intercept			
06	3.90		3.63	0.27	48.53
22	4.05	3.54		0.51	37.52
18	3.50	2.97		0.53	37.44
06	3.90		3.63	0.27	35.59
07	3.42	3.31		0.11	29.74
12	4.27	3.89		0.38	29.43
01	3.57	3.12		0.45	29.28
19	4.51	4.36		0.15	25.47
17	3.04		3.29	−0.26	24.33
49	3.73	4.14		−0.41	24.09
42	4.07	3.81		0.26	23.76
19		4.36	4.46	−0.10	22.29
26	2.60		2.89	−0.29	21.87
18		2.97	3.48	−0.51	20.52
36	3.75	3.53		0.22	19.88
32	3.50		3.54	−0.05	18.56
57	3.46	3.11		0.35	17.98
57	3.46		3.17	0.29	17.23
37	3.71	3.54		0.17	16.44
33	3.90	4.25		−0.35	16.25
34	3.11	2.82		0.29	15.96
33	3.90		4.02	−0.12	15.19
45	3.75	3.72		0.02	14.24
29	3.48	3.78		−0.29	14.21
39	3.41	3.81		−0.40	14.14

Note. Diff represents the difference between unstandardized parameter estimates (i.e., factor loadings or intercepts). All parameter estimates were significantly different from each other after a Bonferroni adjustment (BA; α_{BA} = .05/204 = .00025) based on the change in chi-square.

Evaluating the item intercepts suggested that most (16 of 25) of the group differences emerged between Caucasian and African American subjects, with many of these differences on the Connectedness to Peers (3 of 16) and Connectedness to Self-in-the-Future (3 of 16) subscales. Fewer differences were obtained between Caucasians and Latinos (7 of 25) or between African Americans and Latinos (2 of 25). Other than the noninvariant unstandardized factor loadings on the three negatively worded items, the intercept differences between Caucasians and African Americans were not accompanied by unstandardized factor loading differences. That is, these item differences remained constant across levels of connectedness for a given factor.

Tests of Mean Differences Across Gender and Ethnicity

On the basis of the practical model fit indices, the conditions were met to allow for accurate mean comparisons across gender and ethnic/racial groups using the full invariance model. This is important because when observed scores (average item responses within a subscale) are compared across groups, the implicit assumptions are that full measurement invariance is met and the factor structure possesses tau equivalence (i.e., factor loadings measure the factor with the same degree of precision). Although some would suggest that we should report and discuss latent factor means, there is not a consensus on this issue. To facilitate the interpretation of group comparisons and to aid in the understanding of scale means by future subscale users (particularly those whose sample sizes do not support the comparison of latent means), we report (see Table 7) and discuss observed group mean differences across the 10 factors. The factor mean difference scores for the full and partial invariance models are provided in Table 7. Effect size interpretations are based on standards suggested by Cohen (1988), which are as follows: small ($|d| = 0.20$), medium ($|d| = 0.50$), and large ($|d| = 0.80$). Absolute z values greater than 3.30 ($p < .001$) were considered statistically significant for these analyses.[2] The mean differences (M_{Diff}) always favored the reference group, which is the first group listed in the table. Therefore, the mean difference of 0.09 between boys and girls on the Connectedness to Neighborhood factor indicates the mean score was 0.09 units higher for boys than for girls.

Of the 10 comparisons, only three observed score mean differences were not statistically different between boys and girls: Connectedness to Parents, Connectedness to Self-in-the-Present, and Connectedness to Self-in-the-Future. On all but the Connectedness to Neighborhood subscale, girls scored significantly higher than did boys. Most differences were, for practical purposes, relatively small according to Cohen's (1988) standards. Only the differences on the Connectedness to Friends and Connectedness to Reading subscales, favoring girls, reflected a medium effect size.

The ethnic group comparisons of observed score means revealed that Caucasians and African Americans differed on six of 10 means, and Caucasians and Latino/as differed on six of 10 means (see Table 7). Caucasians scored higher than African Americans on means for Connectedness to Neighborhood, Connectedness to Friends, and Connectedness to Teachers, whereas African Americans scored higher than Caucasians on the Connectedness to Self-in-the-Present, Connectedness to Self-in-the-Future, and Connectedness to Siblings subscales. However, the differences between Caucasians and African Americans on the Connectedness to Friends and Connectedness to Self-in-the-Future subscales were not found to be statistically significant with the partial invariance models, $z = -0.38$, $p > .05$, and $z = -1.22$, $p > .05$, respectively, which makes sense given the number of noninvariant Connectedness to Friends ($n_i = 4$) and Connectedness to Self-in-the-Future ($n_i = 5$) items.

Caucasians were higher than Latino/as on Connectedness to Neighborhood, Connectedness to Friends, Connectedness to Self-in-the-Present, Connectedness to Self-in-the-Future, and Connectedness to Reading subscales, but Latino/as were higher on the Connectedness to Sibling subscale. African Americans and Latino/as differed on the Connectedness to Self-in-the-Present,

[2] This study compared latent variable means following the procedure of Byrne (1998, pp. 303–325). Therefore, the limitation of using the change in chi-square to statistically compare latent means with complex models was not encountered (see Fan & Sivo, 2009).

Table 7
Observed Mean Differences Between Gender and Ethnic/Racial Across Groups of Subjects With Siblings

	Connectedness subscale									
Statistic	Neighborhood	Friends	Self-in-the-Present	Parents	Siblings	School	Peers	Teachers	Self-in-the-Future	Reading
					Boys vs. girls					
M_{Diff}	0.09	−0.41	−0.03	0.04	−0.13	−0.27	−0.13	−0.28	−0.05	−0.54
t	2.55[a]	−17.33[a]	−1.20	1.68	−3.55[b]	−10.24[b]	−5.47[b]	−9.35[b]	−1.92	−13.06[a]
$d_{\text{total score}}$	0.09	−0.60	−0.04	0.06	−0.12	−0.36	−0.19	−0.32	−0.07	−0.45
d_{full}	0.13[a]	**−0.53**[b]	**−0.04**	0.05	−0.11[b]	−0.33[b]	−0.18[b]	−0.31[b]	−0.05	−0.43[b]
d_{partial}	0.12[a]	**−0.26**[b]	**−0.18**[a]	0.07	−0.11[b]	−0.29[b]	−0.22[b]	−0.29[b]	0.01	−0.43[b]
					Caucasian vs. African American					
M_{Diff}	0.26	0.18	−0.18	0.00	−0.28	−0.01	0.07	0.17	−0.21	0.00
t	4.65[b]	4.74[b]	−4.32[b]	−0.08	−4.97[b]	−0.20	1.72	3.59[b]	−4.79[b]	−0.05
$d_{\text{total score}}$	0.17	0.17	−0.16	0.00	−0.18	−0.01	0.06	0.13	−0.18	0.00
d_{full}	0.16[b]	**0.18**[b]	−0.18[b]	−0.05	−0.21[b]	0.00	0.02	0.09[a]	**−0.23**[b]	−0.01
d_{partial}	0.13[b]	**−0.01**	−0.12[a]	−0.06	−0.23[b]	−0.04	−0.05	0.07[a]	**−0.05**	−0.01
					Caucasian vs. Latina/o					
M_{Diff}	0.36	0.16	0.09	−0.04	−0.23	0.02	−0.05	0.01	0.09	0.21
t	6.68[b]	4.07[b]	2.12[a]	−0.94	−4.16[b]	0.45	−1.20	0.29	2.03[a]	3.09[a]
$d_{\text{total score}}$	0.25	0.15	0.08	−0.03	−0.15	0.02	−0.04	0.01	0.07	0.11
d_{full}	0.26	0.16[b]	0.08[a]	−0.07	−0.17[b]	0.04	−0.05	0.01	0.08[a]	0.14[b]
d_{partial}	0.27	0.18[b]	0.13[b]	−0.06	−0.17[b]	−0.04	−0.07	0.02	0.08[a]	0.11[a]
					African American vs. Latina/o					
M_{Diff}	0.11	−0.03	0.27	−0.04	0.05	0.03	−0.11	−0.16	0.29	0.21
t	1.52	−0.50	5.33[b]	−0.70	0.72	0.53	−2.33[a]	−2.72[a]	5.49[b]	2.69[a]
$d_{\text{total score}}$	0.11	−0.04	0.39	−0.05	0.05	0.04	−0.17	−0.20	0.40	0.19
d_{full}	0.13	−0.05	0.39[b]	−0.01	0.09	0.09	−0.13	−0.13	0.44[b]	0.21[a]
d_{partial}	0.10	−0.07	0.36[b]	−0.01	0.09	0.07	−0.09	−0.14	0.38[b]	0.18[a]

Note. The degree of freedom for the Boys vs. Girls, Caucasian vs. African American, Caucasian vs. Latina/o, and African American vs. Latina/o were 3326, 2940, 2948, and 762, respectfully. The first group listed served as the reference group: boys, Caucasians, Caucasians, and African Americans, respectively. Bolded effect sizes had $\Delta d > |.10|$ between the full and partial invariance models.
[a] $\alpha = .05$. [b] $\alpha = .001$.

Connectedness to Self-in-the-Future, Connectedness to Reading, Connectedness to Peers, and Connectedness to Teachers subscales, with African Americans scoring higher on the Connectedness to Self-in-the-Present, Connectedness to Self-in-the-Future, and Connectedness to Reading subscales. It is important to note that no differences emerged when using the factor scores with the full or partial invariance models on the Connectedness to Peers and Connectedness to Teachers subscales, despite having several non-invariant items on these subscales.

Practically speaking (on the basis of Cohen's *d*), differences between the three ethnic groups were relatively small. The largest effect sizes were between African Americans and Latino/as on the Connectedness to Self-in-the-Present and Connectedness to Self-in-the-Future subscales, with effect sizes of 0.39 and 0.40, respectively.

Power Considerations

From a statistical perspective, it is worth recognizing that the power to detect very small differences between African Americans and Latinos was considerably less than with Caucasians. That is, more statistically significant differences between these groups may have emerged if the overall sample size was comparable to those used in the comparisons with Caucasians. Assuming a sample size approximate to that used with the Caucasian analyses ($n = 2,946$) and the conventional power level (i.e., .80) at $\alpha = .05$, effect sizes of .1033 would be needed to reject the null hypothesis of no group differences 80% of the time. Therefore, only the total score difference on Connectedness to Neighborhood would likely become statistically significant between African Americans and Latinos at $\alpha = .05$. Regardless, the effect sizes would remain small.

Discussion

The primary function of this study was to assess measurement invariance across gender and racial/ethnic groups for a measure of adolescent connectedness, thereby allowing the assessment of group differences with greater confidence. A second goal was to estimate mean differences on the 10 connectedness subscales. Results revealed that from a practical standpoint, measurement invariance was upheld. This implies that subjects with equivalent latent construct scores respond similarly to items across ethnic/racial groups and gender, which satisfies the statistical assumption when comparing latent or observed means (Byrne, 1998). Given that invariance was met from a practical perspective, a discussion of mean differences is presented first. Later, we discuss the implications of and benefits to testing both full and partial invariance.

Mean Group Comparisons

Ultimately, an assessment's utility lies in its usefulness for client diagnosis (identification and referral), program evaluation, and research. This study provides a foundation for such work by revealing evidence of mean differences across genders and ethnic/racial groups on several subscales of adolescent connectedness, with the largest differences between genders.

Between girls and boys, statistically significant gender differences in observed means were found on Connectedness to Friends, Connectedness to Siblings, Connectedness to School, Connectedness to Peers, Connectedness to Teachers, and Connectedness to Reading subscales. These differences favored the girls and typically reflected medium effect size differences (Cohen's *d*s between 0.12 and 0.63). A significant but very small difference on Connectedness to Neighborhood was the only subscale on which the boys scored higher than girls. There were no gender differences on the Connectedness to Parents, Connectedness to Self-in-the-Present, or Connectedness to Self-in-the-Future subscales.

Statistically significant ethnic/racial group mean differences were found between Caucasian and ethnic minority (i.e., African Americans and Latinos) youth but represented relatively small effect sizes (Cohen's *d*s between 0.08 and 0.25). Caucasian youth scored higher than ethnic minority youth on the Connectedness to Neighborhood and Connectedness to Friends subscales but lower on the Connectedness to Siblings subscale. All groups differed on the Connectedness to Self-in-the-Present and Connectedness to Self-in-the-Future subscales, with African Americans reporting the highest and Latinos the lowest mean scores. Latinos also reported the lowest levels of connectedness to reading. These differences are consistent with the current literature on adolescent connectedness.

What was contrary to the literature was the absence of between–ethnic group differences in connectedness to school (and, to a lesser degree, to parents), and this may reveal one of the advantages of using the Hemingway subscales over other measures. The absence of differences in school connectedness reported here likely reflects the fact that most scales of school connectedness merge connectedness to teachers, peers, and school. Yet we found Caucasian youth were more connected to their teachers than were African American youth and that only African American and Latino youth differed on connectedness to peers. This suggests that an assessment of each domain of school connectedness is more accurate or revealing than using a global scale. Prior research may have suggested group differences on school connectedness when, in fact, the true differences were actually on connectedness to teachers or peers.

Unlike research on social connectedness among adults (e.g., Lee & Robbins, 1998), research on adolescent connectedness has consistently described connectedness as ecologically and relationally specific. Hoyt, Warbasse, and Chu (2006) suggested that studies such as this one can provide the evidence of construct multidimensionality that is necessary to justify the use of separate subscales. Our findings suggest the Hemingway's connectedness subscales (e.g., to peers, teachers, and school) are conceptually and statistically distinct—only one of these interfactor correlations was greater than .70—and that comparisons using specific subscales within a given context can yield surprisingly different findings. This, along with the evidence of discriminant validity, bodes well for using the Hemingway's separate subscales in research and in the field.

Lessons Learned With Tests of Partial Invariance

The results of this study highlight the consequences of not testing measures using both full and partial invariance models, as conclusions related to latent mean differences sometimes varied on the basis of the model estimated. Within this study, the individual CFA models for each gender and ethnic group sample suggested that the same factor structure (e.g., number of items per factors, pattern of fixed and freed parameters) existed across each group and, from a practical perspective, the criteria for measurement invariance were met. However, from a statistical perspective—that is, estimating the change in chi-square when a given factor loading and/or intercept was allowed to be freely estimated rather than being held constant across groups—a few items did not display factor loading or intercept invariance. This evidence of factor noninvariance suggests that youth of different genders or ethnic/racial groups interpreted, conceptualized, and/or simply responded to some of these items differently. Although these differences should be considered exploratory and preliminary, they may contribute to theory.

Gender differences in the role of communication with friends. Two of the largest item differences were on the Connectedness to Friends factor, on which two items (Item 32 and Item 52) measuring time spent talking with friends had significantly higher intercepts for girls than for boys. This suggests that the item responses or amount of communication for girls with their friends was significantly higher than for boys having the same overall latent trait score on connectedness to friends. Because item intercepts contribute to the factor score, these items would produce higher factor score means for girls than boys on Connectedness to Friends. Stated differently, at the average factor score ($\xi = 0$), girls have higher item responses than boys. Of course, as Byrne and Stewart (2006) explained, intercept noninvariance is generally less serious than factor loading noninvariance, such that the primary utility of this finding may be in how it helps researchers understand gender-specific elements of connectedness. Nevertheless, the lack of invariance at the unstandardized factor loading or intercept level renders between-gender mean comparisons using the Connectedness to Friends subscale dubious.

Gender differences in the role of trust in adolescent connectedness. There appeared to be only a few circumstances under which unstandardized factor loadings differed between genders. For example, Items 12 and 14 both measure trust as an indicator of connection. These items had higher unstandardized factor loadings for boys (.88 and .65, respectively) than for girls (.69 and .44, respectively), suggesting that the relationship between their item responses and overall Connectedness to Friends and Connectedness to Parents scores changed at different rates for boys and girls. Moreover, girls had significantly higher intercepts on these items than did boys. As shown in Figure 2, girls had higher scores than boys on the trust item, but more so for youth who were least connected to their friends. This interaction makes it difficult to compare genders in connectedness to friends as Item 12 functions differently depending on the respondent's degree of overall connectedness to friends.

Using Full Versus Partial Invariance Models

Although there were only a few subscales demonstrating evidence of partial rather than full invariance, the latent mean differences resulting from these models render the use of the Connectedness to Friends and the Connectedness to Self subscales suspect when conducting cross-cultural and gender comparisons. Compared with latent means estimated using a full invariance model, our follow-up analyses that relaxed the constraints on noninvariant items to create a partial invariance factor model resulted in smaller mean difference effect sizes between Caucasians and African Americans on both the Connectedness to Friends ($d_{\text{partial}} = -0.01$ vs. $d_{\text{full}} = 0.18$) and Connectedness to Self-in-the-Future ($d_{\text{partial}} = -0.05$ vs. $d_{\text{full}} = -0.23$) factors. Evaluating the gender comparisons, we found that the Connectedness to Friends factor mean difference effect size was larger when using the full invariance model ($d_{\text{partial}} = -0.26$ vs. $d_{\text{full}} = 0.53$). Yet the partial invariance model yielded larger effect sizes for gender differences on the Connectedness to Self-in-the-Present factor ($d_{\text{partial}} = -0.18$ vs. $d_{\text{full}} = 0.04$). All of these differences are greater than .10 and are bolded in Table 7.

When considering these effect size differences between invariance models, recognize that the amount of bias is unknown and, for practical purposes, neither effect size is necessarily correct. The full invariance model incorrectly assumes that each item contributes the same amount of weight to the factor, whereas the partial invariance model created factor scores using a different weighting schema. Related to the latter situation, if one creates factor scores using a different equation (i.e., different set of unstandardized factor loadings and/or intercepts), the researcher is not necessarily comparing the same factors or constructs.

Collectively, these results suggest that when an assumption of scale invariance is made by program evaluators or researchers (as is done implicitly when observed scores reflect item means) but the subscale is only partially invariant, researchers are likely to produce biased and invalid effect sizes. These errors would mischaracterize group differences, promulgating incorrect information. Most unfortunate for the field is that where noninvariance occurs, this should serve as a harbinger for researchers, signaling them to further explore the meaning of a given construct and the reasons for between-group variability on items or scales. But, where invariance has not been tested, no such signal will be heard.

Implications for Noninvariant Items

Given that an assumption of partial invariance at times may be more appropriate (Schmitt & Kuljanin, 2008), it is important for scale users to consider early on how to deal with items not found to be invariant, such as the items on trust and talkativeness. Research (Cheung & Rensvold, 1999; Millsap & Kwok, 2004) indicates that several procedures can be used if the factor model is not invariant: (a) Delete the noninvariant items, (b) use all the items assuming that differences are small in the population and will not adversely influence the mean differences, (c) avoid using the scale altogether or use it but interpret the scores independently (avoiding group comparisons), and/or (d) use the partial invariance model. The fourth option, however, requires large samples. For users of the Hemingway whose samples are small (e.g., $n < 300$), we believe Option a (delete invariant items) is unwise because doing so creates new versions of the subscale that will not benefit from existing evidence of subscale construct validity (e.g., Karcher, 2001; Karcher et al., 2008). When the third option (avoid group comparisons) is not tenable, the second option will work for gender and ethnic group comparisons on most subscales, specifically when between-group differences on the underlying factor structures are small. As shown in Table 7 and described above, on three factors, there were larger than acceptable between-group differences in the estimated effect sizes when tests were conducted using partial and full invariance models. These are the Connectedness to Friends and Connectedness to Self-in-the-Present subscales for gender comparisons and the Connectedness to Friends and Connectedness to Self-in-the-Future subscales for mean comparisons between Caucasian and African American youth.

Implications for Theories of Attachment During Adolescence

The present study converges with several aspects of attachment theory. First, the connectedness items about affect and action consistently loaded together. Like the two main dimensions of the attachment behavioral system in childhood (viz., proximity seeking and experiencing pleasure and security in specific relationships and contexts), affect and action also appear to be essential elements of connectedness among adolescents. Second, evidence of subscale discriminant validity affirm the person- and place-specific nature of the Hemingway subscales, which is consistent with Ainsworth's (1989) proposition that attachment tendencies differentiate into more distinct forms of "affectional bonds" (p. 709) in adolescence. Third, in addition to the interpersonal and context-specific "worlds" of connectedness (Cooper, 1999), the Hemingway includes two intrapersonal connectedness-to-self subscales, which demonstrated evidence of construct and discriminant validity. These constructs may provide a new way to examine Bowlby's (1969) description about the importance of working models of the self (pp. 710–713).

Yet, between-group mean differences, as well as item-specific noninvariance, point toward new questions. A better understanding is needed of the vicissitudes of this "connectedness-to-self" phenomenon and particularly why it may differ between Black and White adolescents. Item-level gender differences in the role of trust in connectedness to friends and to parents and in rates of talkativeness as indicators of connectedness to friends also deserve further study.

Special Considerations When Using the Connectedness to Siblings Subscale

How the Connectedness to Siblings subscale is used in research and in applied settings needs to be given serious attention. For this study, we excluded the singletons from the analyses and conducted analyses that simply compared subjects with siblings. Thus, the factor invariance results and mean comparisons should only be interpreted as relevant to individuals with siblings. The exclusion of singletons, of course, poses a limitation to the external validity for singletons. However, to address this limitation, invariance analyses were conducted between singletons and subjects with siblings. The results, available in the online supplement, indicated that the models are invariant across these groups and therefore the results should generalize to singletons.

There are a few other points all scale users should consider when addressing this sibling problem. The first is to explicitly request that respondents identify their singleton status. The second is that if singletons' data is treated as missing data on the Connectedness to Siblings subscale, users should verify the validity of the missing data method by testing for factor structure invariance. (Again, we conducted such analyses, but because our designation of singletons is questionable, we do not report them here.) Of course, this approach is controversial and therefore users may elect to omit these subjects, as was done within this study. A third option is to test whether sibling status is a moderator of the associations between the other variables of interest in one's statistical models. Clearly, how to handle inapplicable data or appropriately missing data is an area that deserves attention (Marshall, Morales, Elliot, Spritzer, & Hays, 2001).

Unexamined Sources of Potential Invariance and Questions Raised by This Study

Another limitation of this study was the lack of information on socioeconomic status, older adolescents, and other ethnic groups. For example, the small ethnic/racial group mean differences might have been absent altogether had socioeconomic status been accounted for. Alternatively, it may be that the factor structures are not invariant across other ethnic groups or among older adolescent respondents, such as those in high school or college.

This study may raise as many questions as it answers about measuring adolescent connectedness. Future researchers should examine (a) the role of appropriately missing data (such as for siblings and singletons), (b) how to deal with noninvariant items (e.g., negatively worded items, trust items) or scales, and (c) whether to assess factor structure invariance across additional groups of subjects (e.g., age or grade differences, different socioeconomic status groups).

Overall, however, there is strong evidence that, when assessed using the Hemingway: Measure of Adolescent Connectedness, adolescent connectedness can be characterized by an ecology of adolescent connectedness. In terms of factor structure invariance, most factors (except arguably Connectedness to Friends and Connectedness to Self) appeared invariant across the groups tested, making these connectedness subscales promising for assessment and evaluation purposes across gender and with African American, Caucasian, and Latino early adolescents.

References

Ainsworth, M. D. S. (1989). Attachments beyond infancy. *American Psychologist, 44,* 709–716.

Barber, B. K., & Schluterman, J. M. (2008). *Connectedness* in the lives of children and adolescents: A call for greater conceptual clarity. *Journal of Adolescent Health, 43,* 209–216.

Bollen, K. (1989). *Structural equations with latent variable*. New York, NY: Wiley.

Bonny, A. E., Britto, M. T., Klostermann, B. K., Hornung, R. W., & Slap, G. B. (2000). School disconnectedness: Identifying adolescents at risk. *Pediatrics, 106,* 1017–1021.

Bowlby, J. (1969). *Attachment and loss: Vol. 1. Attachment.* New York, NY: Basic Books.

Byrne, B. M. (1998). *Structural equation modeling with LISREL, PRELIS, and SIMPLIS.* Hillsdale, NJ: Erlbaum.

Byrne, B. M., & Stewart, S. M. (2006). The MACS approach to testing for multigroup invariance of a second-order factor structure: A walk through the process. *Structural Equation Modeling, 13,* 287–321.

Chen, F. F. (2007). Sensitivity of goodness of fit indices to lack of measurement invariance. *Structural Equation Modeling, 14,* 464–504.

Chen, F. F. (2008). What happens if we compare chopsticks with forks? The impact of making inappropriate comparisons in cross-cultural research. *Journal of Personality and Social Psychology, 95,* 1005–1018.

Chen, F. F., Sousa, K. H., & West, S. G. (2005). Testing measurement invariance of second-order factor models. *Structural Equation Modeling, 12,* 471–492.

Cheung, G. W., & Rensvold, R. B. (1999). Testing factorial invariance across groups: A reconceptualization and proposed new method. *Journal of Management, 25,* 1–27.

Cohen, P. A. (1988). *Statistical power analysis for the behavioral sciences.* Hillsdale, NJ: Erlbaum.

Cooper, C. R. (1999). Multiple selves, multiple worlds: Cultural perspectives on individuality and connectedness in adolescent development. In A. S. Masten (Ed.), *Cultural processes in child development* (pp. 25–57). Mahwah, NJ: Erlbaum.

Davey, A., Savla, J. S., & Luo, Z. (2005). Issues in evaluating model fit with missing data. *Structural Equation Modeling, 12,* 578–597.

du Toit, M., & du Toit, S. H. C. (2001). *Interactive LISREL: User's guide.* Lincolnwood, IL: Scientific Software International.

Evans, D. L., Foa, E. B., Gur, R. E., Hendin, H., O'Brien, C. P., Seligman, M. E. P., & Walsh, B. T. (Eds.). (2005). *Treating and preventing adolescent mental health disorders: What we know and what we don't know.* New York, NY: Oxford University Press.

Fan, X. T., & Sivo, S. A. (2009). Using goodness-of-fit indexes in assessing mean structure invariance. *Structural Equation Modeling, 16,* 54–69.

French, B. F., & Finch, W. H. (2008). Multi-group confirmatory factor analysis: Locating the invariant referent. *Structural Equation Modeling, 15,* 96–113.

García Coll, C., Crnic, K., Lamberty, G., Wasik, B. H., Jenkins, R., García, H. V., & McAdoo, H. P. (1996). An integrative model for the study of developmental competencies in minority children. *Child Development, 67,* 1891–1914.

Garringer, M. (2009). *Redefining your success: The role of adolescent connectedness in demonstrating mentoring impact* (Mentoring Resource Center Fact Sheet No. 28). Retrieved from http://www.edmentoring.org/pubs/factsheet28.pdf

Grossman, J. B., & Bulle, M. J. (2006). Review of what youth programs do to increase the connectedness of youth with adults. *Journal of Adolescent Health, 39,* 788–799.

Herrera, C., Grossman, J. B., Kauh, T. J., Feldman, A. F., & McMaken, J. (2007). *Big Brothers Big Sisters school-based mentoring impact study.* Philadelphia, PA: Public/Private Ventures.

Hoyle, R. H. (1995). *Structural equation modeling: Concepts, issues, and applications.* Thousand Oaks, CA: Sage.

Hoyt, W. T., Warbasse, R. E., & Chu, E. Y. (2006). Construct validation in counseling psychology research. *Counseling Psychologist, 3,* 769–805.

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

Jacobson, K. C., & Rowe, D. C. (1999). Genetic and environmental influences on the relationships between family connectedness, school connectedness, and adolescent depressed mood: Sex differences. *Developmental Psychology, 35,* 926–939.

Jöreskog, K. G., & Sörbom, D. (1996). *LISREL 8: User's reference guide.* Chicago, IL: Scientific Software International.

Karcher, M. J. (2001). *The Hemingway: Measure of Adolescent Connectedness: Validation studies.* Retrieved from ERIC database. (ED477969)

Karcher, M. J. (2002). The cycle of violence and disconnection among rural middle school students: Teacher disconnection as a consequence of violence. *Journal of School Violence, 1,* 35–51.

Karcher, M. J. (2005). *The Hemingway: Measure of Adolescent Connectedness: A manual for scoring and interpretation.* Unpublished manuscript, University of Texas at San Antonio. Retrieved June 12, 2007, from http://adolescentconnectedness.com/media/HemManual2006.pdf

Karcher, M. J. (2008). The Study of Mentoring in the Learning Environment (SMILE): A randomized evaluation of the effectiveness of school-based mentoring. *Prevention Science, 9,* 99–113.

Karcher, M. J., Davis, C., & Powell, B. (2002). The effects of developmental mentoring on connectedness and academic achievement. *The School Community Journal, 12*(2), 36–50.

Karcher, M. J., & Finn, L. (2005). How connectedness contributes to experimental smoking among rural youth: Developmental and ecological analyses. *Journal of Primary Prevention, 26,* 25–36.

Karcher, M. J., Holcomb, M., & Zambrano, E. (2008). Measuring adolescent connectedness: A guide for school-based assessment and program evaluation. In H. L. K. Coleman & C. Yeh (Eds.), *Handbook of school counseling* (pp. 649–669). Mahwah, NJ: Erlbaum.

Lee, R. M., & Robbins, S. B. (1998). The relationship between social connectedness and anxiety, self-esteem, and social identity. *Journal of Counseling Psychology, 45,* 338–345.

Lee, R. M., & Robbins, S. B. (2000). Understanding social connectedness in college women and men. *Journal of Counseling and Development, 78,* 484–491.

Little, T. D. (1997). Mean and covariance structures (MACS) analyses of cross-cultural data: Practical and theoretical issues. *Multivariate Behavioral Research, 32,* 53–76.

Marsh, H. W., & Hocevar, D. (1985). Application of confirmatory factor analysis to the study of self-concept: First- and higher-order factor models and their invariance across groups. *Psychological Bulletin, 97,* 562–582.

Marshall, G. N., Morales, L. S., Elliot, M., Spritzer, K., & Hays, R. D. (2001). Confirmatory factor analysis of the Consumer Assessment of Health Plans Study (CAHPS) 1.0 Core Survey. *Psychological Assessment, 13,* 216–229.

McNeely, C. A., Nonnemaker, J. M., & Blum, R. W. (2002). Promoting school connectedness: Evidence from the National Longitudinal Study of Adolescent Health. *Journal of School Health, 72,* 138–146.

Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika, 58,* 525–543.

Millsap, R. E., & Kwok, O. (2004). Evaluating the impact of partial factorial invariance on selection in two populations. *Psychological Methods, 9,* 93–115.

Ployhart, R. E., & Oswald, F. L. (2004). Applications of mean and covariance structure analysis: Integrating correlational and experimental approaches. *Organizational Research Methods, 7,* 27–65.

Resnick, M. D., Bearman, P. S., Blum, R. W., Bauman, K. E., Harris, K. M., Jones, J., . . . Udry, J. R. (1997). Protecting adolescents from harm: Findings from the National Longitudinal Study on Adolescent Health. *JAMA, 278,* 823–832.

Roth, J. L., & Brooks-Gunn, J. (2003). What exactly is a youth development program? Answers from research and practice. *Applied Developmental Science, 7,* 94–111.

Roth, J., Brooks-Gunn, J., Murray, L., & Foster, W. (1998). Promoting healthy adolescents: Synthesis of youth development program evaluations. *Journal of Research on Adolescence, 8,* 423–459.

Rubin, D. B. (1976). Inference and missing data. *Biometrika, 63,* 581–592.

Schmitt, N., & Kuljanin, G. (2008). Measurement invariance: Review of practice and implications. *Human Resource Management Review, 18,* 210–222.

Steenkamp, J. E. M., & Baumgartner, H. (1998). Assessing measurement invariance in cross national consumer research. *Journal of Consumer Research, 25,* 78–90.

Townsend, K. C., & McWhirter, B. T. (2005). Connectedness: A review of the literature with implications for counseling, assessment, and research. *Journal of Counseling and Development, 83,* 191–201.

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,* 4–70.

Whitlock, J. L. (2006). Youth perceptions of life at school: Contextual correlates of school connectedness in adolescence. *Applied Developmental Science, 10,* 13–29.

(*Appendix follows*)

Appendix

Hemingway—Measure of Adolescent Connectedness

From *The Hemingway: Measure of Adolescent Connectedness: A Manual for Scoring and Interpretation* (pp. 24–26) by M. J. Karcher, 2005, unpublished manuscript, University of Texas at San Antonio.

Instructions: First, tell us, do you have any brothers or sisters? No Yes (circle one).

Next, please use this survey to tell us about yourself. Read each statement. CIRCLE the number that best describes how true that statement is for you or how much you agree with it. If a statement is unclear to you, ask for an explanation. If it still unclear, put a " ?".

How TRUE about you is each sentence? not at all = 1, not really = 2, sort of true = 3, true = 4, very true = 5.

	Not at All	Not Really	Sort of	True	Very True
1. I like hanging out around where I live (like my neighborhood).	1	2	3	4	5
2. Spending time with friends is not so important to me.	1	2	3	4	5
3. I can name 5 things that others like about me.	1	2	3	4	5
4. My family has fun together.	1	2	3	4	5
5. I have a lot of fun with my brother(s) or sister(s). (Skip if you have none.)	1	2	3	4	5
6. I work hard at school.	1	2	3	4	5
7. My classmates often bother me.	1	2	3	4	5
8. I care what my teachers think of me.	1	2	3	4	5
9. I will have a good future.	1	2	3	4	5
10. I enjoy spending time by myself reading.	1	2	3	4	5
11. I spend a lot of time with kids around where I live.	1	2	3	4	5
12. I have friends I'm really close to and trust completely.	1	2	3	4	5
13. There is not much that is unique or special about me.	1	2	3	4	5
14. It is important that my parents trust me.	1	2	3	4	5
15. I feel close to my brother(s) or sister(s). (Skip if you have none.)	1	2	3	4	5
16. I enjoy being at school.	1	2	3	4	5
17. I like pretty much all of the other kids in my grade.	1	2	3	4	5
18. I do not get along with some of my teachers.	1	2	3	4	5
19. Doing well in school will help me in the future.	1	2	3	4	5
20. I like to read.	1	2	3	4	5
21. I get along with the kids in my neighborhood.	1	2	3	4	5
22. Spending time with my friends is a big part of my life.	1	2	3	4	5
23. I can name 3 things that other kids like about me.	1	2	3	4	5
24. I enjoy spending time with my parents.	1	2	3	4	5
25. I enjoy spending time with my brothers/sisters. (Skip if you have none.)	1	2	3	4	5
26. I get bored in school a lot.	1	2	3	4	5
27. I like working with my classmates.	1	2	3	4	5
28. I want to be respected by my teachers.	1	2	3	4	5
29. I do things outside of school to prepare for my future.	1	2	3	4	5
30. I never read books in my free time.	1	2	3	4	5
31. I often spend time playing or doing things in my neighborhood.	1	2	3	4	5
32. My friends and I talk openly with each other about personal things.	1	2	3	4	5
33. I really like who I am.	1	2	3	4	5

	Not at All	Not Really	Sort of	True	Very True
34. My parents and I disagree about many things.	1	2	3	4	5
35. I try to spend time with my brothers/sisters when I can. (Skip if you have none.)	1	2	3	4	5
36. I do well in school.	1	2	3	4	5
37. I get along well with the other students in my classes.	1	2	3	4	5
38. I try to get along with my teachers.	1	2	3	4	5
39. I do lots of things to prepare for my future.	1	2	3	4	5
40. I often read when I have free time.	1	2	3	4	5
41. I hang out a lot with kids in my neighborhood.	1	2	3	4	5
42. I spend as much time as I can with my friends.	1	2	3	4	5
43. I have special hobbies, skills, or talents.	1	2	3	4	5
44. My parents and I get along well.	1	2	3	4	5
45. I try to avoid being around my brother/sister(s). (Skip if you have none.)	1	2	3	4	5
46. I feel good about myself when I am at school.	1	2	3	4	5
47. I am liked by my classmates.	1	2	3	4	5
48. I always try hard to earn my teachers' trust.	1	2	3	4	5
49. I think about my future often.	1	2	3	4	5
50. I usually like my teachers.	1	2	3	4	5
51. My neighborhood is boring.	1	2	3	4	5
52. My friends and I spend a lot of time talking about things.	1	2	3	4	5
53. I have unique interests or skills that make me interesting.	1	2	3	4	5
54. I care about my parents very much.	1	2	3	4	5
55. What I do now will not affect my future.	1	2	3	4	5
56. Doing well in school is important to me.	1	2	3	4	5
57. I rarely fight or argue with the other kids at school.	1	2	3	4	5

Received May 7, 2008
Revision received February 5, 2010
Accepted February 8, 2010 ■