Psychological Assessment
2016, Vol. 28, No. 3, 331–344

1040-3590/16/$12.00 http://dx.doi.org/10.1037/a0039487

Cross-Informant and Cross-National Equivalence Using Item-Response Theory (IRT) Linking: A Case Study Using the Behavioral Assessment for Children of African Heritage in the United States and Jamaica

Michael Canute Lambert
The University of North Carolina at Chapel Hill

Gail M. Ferguson
University of Illinois at Urbana–Champaign

George T. Rowan
Michigan State University

Cross-national study of adolescents' psychological adjustment requires measures that permit reliable and valid assessment across informants and nations, but such measures are virtually nonexistent. Item-response-theory–based linking is a promising yet underutilized methodological procedure that permits more accurate assessment across informants and nations. To demonstrate this procedure, the Resilience Scale of the Behavioral Assessment for Children of African Heritage (Lambert et al., 2005) was administered to 250 African American and 294 Jamaican nonreferred adolescents and their caregivers. Multiple items without significant differential item functioning emerged, allowing scale linking across informants and nations. Calibrating item parameters via item response theory linking can permit cross-informant cross-national assessment of youth.

Keywords: cross-informant, cross-national, youth assessment, IRT, linking, Caribbean BACAH

Researchers conducting comparative studies are charged with ensuring that their findings are relatively free from measurement artifacts. Ideally, comparing findings across groups should as closely as possible reflect true differences or similarities. This requires establishing language/conceptual equivalence and cross-group psychometric invariance of assessment tools (see Vlachopoulos, Ntoumanis, & Smith, 2010). Attention is often given to translation equivalence (i.e., typically through translation and back translation) when differences in language exist (see King, Khan, Leblanc, & Quan, 2011), and sometimes conceptual equivalence (i.e., how populations conceptualize the constructs being measured; see King et al., 2011; Olsen, Jensen, Tesfaye, & Holm, 2013). It is also essential to ensure that scores across different groups are equivalent (i.e., that ratings are given on an equivalent metric; Kim et al., 2014). Unfortunately, few studies have examined psychometric equivalence across groups.

The burden of ensuring psychometric scalar equivalence for studying youth adjustment is greater, and the task is more complex, because cross-informant assessment (i.e., scales that measure similar constructs across informants) is the gold standard (Carter, Briggs-Gowan, & Davis, 2004; Renk, 2005). When equivalence studies are conducted, they generally address configural invariance of ratings derived from one set of informants (e.g., parents, teachers, youth) within different demographic groups (e.g., Ivanova et al., 2007; Rescorla et al., 2007). That is, invariance studies generally examine whether the existing factor structure (including the same number of factors within it) established on ratings from one group is evident for other groups (Campbell, Barry, Joe, & Finney, 2008). Ignored are metric and scalar equivalence, which, respectively, test whether the relationship between items and factors are identical across groups and whether two groups respond to a scale in identical fashion (Campbell et al., 2008; Schmitt & Kuljanin, 2008). Given the growing emphasis placed on cross-national/ethnic/cultural studies of youth development and functioning (e.g., Berry, Phinney, Sam, & Vedder, 2006; Georgas, Berry, van de Vijver, Kagitçibasi, & Poortinga, 2006; Ivanova et al., 2007;

This article was published Online First July 6, 2015.

Michael Canute Lambert, School of Social Work, The University of North Carolina at Chapel Hill; Gail M. Ferguson, Department of Human and Community Development, University of Illinois at Urbana–Champaign; George T. Rowan, Department of Community, Agriculture, Recreation and Resource Studies, Michigan State University.

Data collection occurred while Michael Canute Lambert was at the Department of Human Development and Family Studies, University of Missouri–Columbia. Gail M. Ferguson was at the Department of Psychology, Knox College.

Supported by the Millsap Professorship of Diversity and Multicultural Studies Endowment fund from the University of Missouri–Columbia; the David Walker Research Institute, College of Human Medicine, Michigan State University; Faculty Development Funds from Knox College; and the Intramural Research Program of the National Institute of Child Health and Human Development. Earlier versions of this paper were presented at the 2011 Caribbean Research Conference in Psychology, Nassau, Bahamas, and at the 2011 Biennial Meeting of the Society for Research in Child Development, Montreal, Canada. Michael Canute Lambert is the primary author of the BACAH and George T. Rowan is the second author. Although the BACAH might have marketing and monetary potential, there has been no financial gain from the BACAH currently or in the past.

Correspondence concerning this article should be addressed to Michael Canute Lambert, The University of North Carolina at Chapel Hill, School of Social Work, CB# 3550, 325 Pittsboro Street, Chapel Hill, NC 27599-3550. E-mail: mclamber@email.unc.edu

Rescorla et al., 2007), methods that permit accurate assessment across informants and cultural groups are needed. To this end, we aimed to demonstrate how item response theory (IRT) can contribute to metric and scalar equivalence via its linking capabilities and thus ensure more accurate cross-informant and cross-national youth assessment. As a case example, we used a sample of Black youth in the United States and a comparison sample in Jamaica, whose behavioral competencies were rated both by themselves and their caregivers, using the Behavioral Assessment for Children of African Heritage (BACAH) Resilience scale.

Cross-Informant Assessment

Thorough assessment of youth depends on the availability of measures that permit the acquisition of information from multiple sources (Achenbach & Rescorla, 2004). Such information sources include youth who are targeted for assessment as well as adults who play significant roles in their lives, including parents and teachers. Youth are typically considered as being capable of informing on their internal psychological states, yet parents and teachers can also provide information on adolescents' functioning in contexts such as home or school (see Renk, 2005, for review). Hence, comparing information derived from different informants permits the professional to pinpoint not only levels at which youth function, but also whether such levels vary according to context. Such cross-contextual assessment requires availability of cross-informant scales.

Cross-informant means that two or more informants are rating a child/adolescent on a similar dimension. For example, to conduct cross-informant assessment of resilience would require that two or more informants (e.g., teachers and their parents) rate youth on items that reflect this construct. Cross-informant scales are often derived through dimension reduction techniques such as factor analyses that are typically conducted on responses given by each specific type of informant considered separately (e.g., Achenbach & Rescorla, 2001). Whereas the item content of cross-informant dimensions might be similar for different informants, they are not usually identical. This phenomenon emerges for two reasons. First, as discussed above, different informants provide information that is not only informant-specific but also context-specific. Hence, the item content on each measure must match youth functioning in each setting, and identical behaviors do not always emerge in each context. For example, "having good relationships with family members" is often observable in the home context but seldom exhibited in the school environment. The same is true for classroom-related behavior such as the child "being helpful to teachers and other students." The second reason item content of cross-informant dimensions is not identical across informants rests on the understanding that many empirically derived dimensions are obtained from factor analyses conducted on responses from each set of informants considered separately. Hence, although some items might be identical for two or more informant groups, all such items might not load on cross-informant scales for other sets of informants. Such differences often pose a challenge for scale developers who must determine how to reliably integrate findings across different informants who respond to different sets of items measuring a specific construct. These challenges are addressed inconsistently across measures of youth adjustment.

Procedures Used to Develop Cross-Informant Dimensions of Youth Assessments

The Achenbach System of Empirically Based Assessment (ASEBA) school age forms (Achenbach & Rescorla, 2004) are arguably some of the most widely used measures of psychological adjustment in youth internationally. To develop cross-informant dimensions, the authors of this set of measures examined item content of similar scales derived from factor analyses for each informant group. Items that loaded on a given factor for two or more informants were considered as content for cross-informant factors. Our earlier studies of the BACAH adopted this approach (see Lambert et al., 2005), although it was not ideal for the following two reasons. First, an item loading on the cross-informant factor for only two informants of three neglects the knowledge that such items are often not really indicators on this construct for at least one informant. This situation violates important measurement tenets: (a) essential dimensionality—all items on a specific dimension should essentially represent a single dimension (Stout, 2001); and (b) conditional independence—respondents' responses should only reflect their standing on the construct being measured and not that of a higher order or different constructs (Zhang & Stout, 1999). The second reason that the ASEBA approach to creating cross-informant scales is not ideal is that requiring all items to load on any given cross-informant dimension for at least two sets of informants might remove items that load on the dimension for a single informant group. Removing such items can result in loss of significant information that could be critical to measuring the construct for a given set of informants.

Besides the ASEBA (Achenbach & Rescorla, 2001), the establishment of cross-informant dimensions in other widely used measures of children's functioning is less clear. In the manual for the third edition of his cross-informant scales, Conners (2008) only provided correlation coefficients for similar scales across parent-, teacher-, and self-reports. The same is true for Brown's (2001) Attention Deficit Scales. The manual for the Behavioral Assessment System for Children–Second Edition (BASC-2) listed correlation coefficients between parent, teacher, and child self-report ratings, and provided information to make qualitative comparisons of BASC-2 item content and scores across two or three types of informants (Reynolds & Kamphaus, 2004). However, it appears that no information was provided for any of these widely used measures regarding metric or scalar equivalence across respondent type.

Psychometric parameters (e.g., reliability, validity) for the ASEBA, Conners, and BASC-2 measures were estimated separately for each type of informant (Achenbach & Rescorla, 2001; Conners, 2008; Reynolds & Kamphaus, 2004). However, parameters estimated separately for different respondent groups are not comparable (see de Ayala, 2009; Embretson & Reise, 2000). This is especially true if such respondents' ratings lack metric and scalar equivalence. Absence of measurement invariance information makes it difficult to be certain that findings from cross-informant study of youth were rated using the same metric or scale.

Cross-National Assessment

Appropriate cross-national assessment exists if one or more sets of informants (e.g., parents and/or adolescents) in one nation rate

youth with identical levels of functioning on the same metric and scale as do similar set(s) of informants in another nation. This can reduce measurement artifacts due to cross-national response style differences (Schmitt & Kuljanin, 2008).

Case Example: BACAH Across Nations

The practice of clinical psychology with children and adolescents has shifted from being solely deficit-focused (biomedical disease model) to recognizing the importance of strengths in predicting youth adaptation (resilience model; Goldstein & Brooks, 2013). Moreover, professionals who develop or execute interventions for youth of varying socioethnic backgrounds are usually aware that assessing and harnessing their strengths is critical to combating clinical concerns (Robinson, 2001). Resilience refers to "positive adaptation in the face of risk or adversity" (Wright, Masten, & Narayan, 2013, p. 17). Because "judging resilience . . . involves decisions about how well a person is doing in life" (Wright et al., 2013, p. 18), resilience can be measured by assessing a young person's competence at meeting expectations for his or her developmental stage and sociocultural context.

Measures of youth competencies are particularly rare for youth of African heritage because problem-based models, which focus on deviant behavior among Black youth, are most often used (Nicolas et al., 2008). One exception is the multi-informant BACAH (Lambert et al., 2005), which takes a strength-based assessment approach and was designed and normed with extensive input from the African American community. This set of cross-informant scales assesses psychological functioning in youth of the African diaspora. Normed on reports given by large samples of Black youth as well as their parents and teachers, BACAH scales can measure multiple dimensions of behavioral and emotional strengths, as well as problems (see Lambert et al., 2005; Lambert, Rowan, Rowan, & Mount, 2014). The BACAH dimensions measure strengths that are common to Black and non-Black youth (e.g., academic and social skills, self-regulation; Spencer & Tinsley, 2008; Goldstein & Brooks, 2013). In addition, BACAH dimensions tap strengths rooted in Black children's heritage such as spirituality and religion, as well as cultural interest and engagement (Chatters, Taylor, Jackson, & Lincoln, 2008; Lambert et al., 2005). These strengths serve as sources of resilience for Black youth in response to challenges presented by the broader society, including racism. Metric and scalar equivalence of BACAH scores across informants has been investigated in African American youth (Lambert et al., 2005; Lambert et al., 2014). Where lack of measurement invariance has been identified, IRT linking has been used to permit equivalence, but invariance considered across nations *and* informants simultaneously has not yet been investigated.

The Current Study

This study aimed to test the measurement invariance of reports on the BACAH Resilience scale, using this as a case study to demonstrate IRT linking of measurement across informants (adolescents vs. their parents) and nations (the majority White nation of the United States vs. the majority Black nation of Jamaica). To achieve this goal, we first identified BACAH Resilience items that showed invariance across all possible pairs of informants and nations, then we used these psychometrically invariant items to link scales across informants and nations. We hypothesized that items would load on a single Resilience factor for Jamaicans based on this finding in the original BACAH validation sample of referred and nonreferred Black youth in the United States (Lambert et al., 2005). Although little work has been conducted on strengths across Jamaican and African American youth, psychopathology research has shown that Jamaican parent- and youth self-report informants underreport adolescent functioning (Lambert, Essau, Schmitt, & Samms-Vaughan, 2007). Thus, we also hypothesized that should the absence of invariance emerge, it would primarily be related to Jamaican informants underreporting on the Resilience scale.

Method

Participants

African American. The African American participants reported on nonreferred youth who were part of the sample (i.e., referred and nonreferred youth) used in estimating the psychometric properties of all BACAH scales as described in detail elsewhere (Lambert et al., 2005). There were 250 nonreferred African American youth ages 11–18 years (M = 14.68 years, SD = 2.55; 52% girls; 100% self-identified as Black). A parent of each adolescent also participated in the study (96% mothers and 4% fathers). All parents in this study identified themselves and their children as African American. All parents were born and raised in the United States; hence, the sample included no recent immigrants. Socioeconomic status (SES) was derived from the occupation of the main income earner in each adolescent's family and coded according to the Hollingshead (1975) Nine-Step Scale, ranging from 1 (*lowest*) to 9 (*highest*; M = 5.05, SD = 1.93). Although income level considered together with occupational level might provide more precise measurement of SES, respondents in our pilot study found questions about income intrusive and offensive, so these items were dropped from our survey.

Jamaican. The Jamaican participants were recruited for a study on the acculturation and adjustment of Jamaican adolescents. There were 294 nonreferred adolescents ages 11–18 years (M = 13.70 years, SD = 2.19; 44% girls; 85% self-identified as Black and 11% as multiracial). Adolescents reported similar middle-class SES (M = 4.81, SD = 1.57, on a 9-point scale; Hollingshead, 1975). The mother of each adolescent also participated.

Data Collection Procedures

African American sample. Youth were recruited from 12 randomly selected schools throughout one midwestern and two northeastern states during the period of 1999–2003. A boy and a girl were randomly sampled from each classroom in the selected schools and written parental consent was sought prior to contacting all youth. Written consent was obtained from each adolescent and his or her parent prior to their completion of the BACAH. Each adolescent and parent was offered U.S. $10 to complete the BACAH form.

Jamaican sample. Youth were recruited from two schools, one afterschool program, and two churches in 2009. The two schools selected were targeted because they are fairly representative of high schools in Jamaica, and together they serve a diverse

array of students, both geographically and socioeconomically (note that students begin Jamaican high schools at age 11/12 and graduates of the highest grades are 18/19 years old). Written adolescent assent and parental/guardian consent were obtained for all families prior to the completion of BACAH forms. Individual incentives (≤U.S. $10 value) or group drawings (≤U.S. $50 value) were used.

Measures

The development of BACAH forms and estimation of their psychometric properties are fully described elsewhere (see Lambert et al., 2005); we only briefly summarize them here. The Parent-, Teacher-, and Self-Report Forms, as well as an Interview Schedule, were developed with extensive input from the African American community using focus groups conducted throughout the State of Michigan. To further ensure content and cultural validity of the BACAH, 30 additional Black youth ages 11–18 who did not participate in the focus groups, along with their parents and their teachers, completed BACAH forms and provided us with feedback on their content and ease of completion. All participants provided positive feedback, stating that the items and instructions were clear and that the content was extremely relevant for Black youth. A second set of 30 Black youth, their parents, and their teachers, as well as 30 clinicians who had not participated in the focus group or pilot studies, were also asked to rate the instructions, items, and rating scales from all three BACAH forms for clarity and relevance on two separate 5-point scales. The scale assessing clarity of instructions and items ranged from 0 (*not clear*) to 4 (*extremely clear*), while the scale for item relevance ranged from 0 (*not relevant*) to 4 (*extremely relevant*). Most participants (> 90%) indicated that the items and instructions were extremely clear and that the items were extremely relevant for Black youth assessment. Minor modifications (i.e., mainly rewording) were done to fewer than a dozen items, according to these participants' suggestions.

The BACAH strength items are rated on two related 3-point Likert scales. The first Likert scale on the left side of each item is designed to rate the presence and the magnitude of each strength item: 0 (*not true* [as far as he or she knows]), 1 (*somewhat or sometimes true*), and 2 (*very true or often true*). The second scale (i.e., measuring impact of ratings given to each strength item on the adolescent's functioning) is listed on the right side of each item and is intended for clinical use. Thus, only the left scale was used in this study.

The initial factor solution for the BACAH was derived from a combined sample of referred and nonreferred African American children; however, only the nonreferred subsample was used in the current study only. Factor analyses were conducted on the ratings for each set of informants considered separately. These analyses identified cross-informant (i.e., similar across Parent-, Teacher-, and Self-Reports) dimensions of strengths for more than 60 items on each of the three forms. Findings from the factor analyses yielded two cross-informant strengths scales, labeled (a) Resilience and (b) Emotional Control and Prosocial Behavior (Lambert et al., 2005). In previous studies, items for each BACAH cross-informant scale were calibrated to permit more reliable comparisons of scores across informants (Lambert et al., 2005; Lambert et al., 2014). The items that form the cross-informant Resilience scale for parent (34 items) and youth self-reports (27 items) are the focus of this study. Resilience scale items were administered to Jamaican and African American youth and parents sampled across the two nations. Jamaican adolescents completed all but two items on the Resilience scale, which were inadvertently excluded (i.e., "59. Appropriate role models" and "60. Shows good problem solving skills"). As noted in the Data Analyses section, although the number of items administered to each group might vary, IRT, including its linking capabilities, nevertheless allows equivalent cross-group assessment (see Performing IRT Linking subsection and Figure 1 for further discussion).

To ensure that the BACAH dimensions have appropriate conceptual and language equivalence for the Jamaican populations, in a previous study (Lambert et al., 2014), we patterned recently published studies (e.g., King et al., 2011; Olsen et al., 2013). Specifically, we assembled a team of three Jamaican researchers with doctoral-level training and at the academic rank of professor in their respective institutions (i.e., a clinical psychologist, a developmental pediatrician/epidemiologist, and a sociologist). Collectively, these professionals have more than half a century of experience conducting research in Jamaica. These researchers examined the Resilience scale and determined that conceptually this dimension is appropriate for Jamaicans. Furthermore, they examined each item to determine whether changes were needed to reflect the intended meaning and idiomatic expression of Jamaicans. Finally, a pilot study was conducted with 50 Jamaican youth who stated that they had no trouble understanding items on the BACAH scales. No suggestions for changes in item content or structure emerged from the steps detailed above.

Data Analyses

IRT procedures are well-suited to the objectives of this study (see Embretson & Reise, 2000). IRT addresses the probability of a particular response to items measuring a specific trait labeled theta (θ). An IRT model was used to estimate the probability that informants would respond affirmatively to items that measure each adolescents' specific Resilience level (e.g., informants reporting on youth with higher Resilience levels would most likely endorse items that measure high levels of the Resilience construct; Panter & Reeve, 2002). Our analytic approach for this study involved: (a) testing IRT assumptions of dimensionality, conditional independence, and selecting an appropriate IRT model; (b) identifying items without significant differential item functioning (DIF) across informant and nationality groupings; and (c) using such items to link item parameters estimates across these groups.

Testing IRT assumption: Appropriate dimensionality. Multiple approaches are available to identify factors from item scores in a dataset and thus to determine appropriate dimensionality. In this study, we used IRTPRO (2001) to conduct a full information IRT factor analyses (FIFA) on the Jamaican and U.S. parent samples separately. Identical analyses were conducted on self-reports from each nation. Factor solutions with two or more factors were rotated according to the promax (oblique) criteria.

Testing IRT assumption: Conditional independence. For each Informant × Nationality group, the items that loaded on the chosen factor solution were compared with other factor solutions (e.g., one vs. two factors) and with an IRT bifactor analysis using IRTPRO (2001). Y. Liu and Thissen (2012) have demonstrated

Demonstrating Linking Items Without Significant Differential Item Functioning Across Groups

First Item Group	Second Item Group	Third Item Group	Fourth Item Group	Fifth Item Group
A	W	A	A	A
B		B	B	B
C				
D				
E	E	E		
F	F			F
G	G			
	H	H		H
	I			
	J	J		
	K	K		
	L	L		
		M	M	M
		N	N	
		O	O	
		P	P	
			Q	Q
		R	R	R
			S	S
			T	T
			U	
				V
W	W	W	W	W
X		X		X
				Y
				Z

Note. The columns represent a database with items. In each column, items are depicted in uppercase letters. Arrows connecting identical items across two adjacent groups reflect that such items have no significant DIF across those groups.These items are constrained across their respective groups in the partially constrained model. If item groups represent the same unidimensional construct, it is not necessary that all item sets across each group are identical or even have identical numbers of items. Moreover, items without significant DIF are only required to be identical across adjacent groups in the databases to which they are linked. Constraining all items across adjacent groups with nonsignificant DIF as presented in this figure would result in a partially constrained model that links items across groups and would thus place all groups of items on an equivalent metric.

Figure 1. Demonstrating linking items without significant differential item functioning (DIF) across groups.

evidence that conditional independence can be addressed by conducting bifactor analyses. The bifactor model (omnibus test) helps the researcher determine not only whether items on the factor model chosen are appropriately loading on their respective factor(s) but also whether they are simultaneously loading on a second-order bifactor (see Reise et al., 2011; Yang et al., 2013). Hence, more parameters are estimated in the bifactor model than in the first-order factor analyses. This permits the researcher to nest the less complex first-order factor models in the bifactor model (Reise et al., 2011), and to use fit statistics such as –2likelihood and the Akaike information criterion (AIC) to determine model fit. Model improvement would be evident if the bifactor model has significantly lower fit indices than the chosen factor solution. Model improvement would suggest that conditional independence is evident for specific scales. Degradation or no significant differences would suggest that the more complex bifactor model adds little or no information to the more parsimonious factor solution chosen. Conditional independence would therefore be deemed evident if the bifactor model fits the data best (Gibbons et al., 2007; Panter & Reeve, 2002).

Testing IRT assumption: Selecting the most appropriate model. Specification of the correct measurement model is important to psychometric applications including DIF testing (see Tay, Ali, Drasgow, & Williams, 2011). For this set of analyses, we tested whether the one-parameter logistic (1PL) model, where the *a* parameters (discrimination) were constrained to be equal for all items (see Embretson & Reise, 2000; X. Liu, 2010; Reeve, 2002) or two-parameter logistic (2PL; Samejima's graded: Samejima, 2010), was the best fitting model for the data. The 2PL model assumes that both *a* and *b* (location) parameters vary, and estimates both. The *a* and *b* parameter estimates were used to plot item characteristic curves for each item and a test characteristic curve (TCC) for the entire test/dimension on graphs where the *y* axis is the probability of a response and the *x* axis measures θ in standardized units. (Higher *a* parameter estimates yield steeper curves and higher *b* parameter estimates locate curves further right, indicating higher θ levels measured.) Samejima's graded model lends itself well to estimating parameters from items in ordered ratings scales (e.g., Likert), and 2PL models are more often used in psychological testing. The three-parameter logistic (3PL) model is seldom used in personality testing or measures of behavioral and emotional functioning, and is rarely used in ordered scales such as those used by the BACAH. Hence, we excluded the 3PL test from the analyses. For the above tests, we examined the $\Delta\chi^2$ statistic as well as the AIC to compare nested 1PL and 2PL models. We also examined other models such as the generalized partial credit and

the nominal models. Because the latter models estimate virtually the same number of parameters as the 2PL model, they could not be considered as nested. Because in model comparisons the AIC is interpreted merely by a numerical change in the AIC index across models (i.e., ΔAIC), where ΔAIC increases of 6 and especially 10 or more should be considered a poorer fit, we used AIC changes as the marker of fit (see Symonds & Moussalli, 2011). For such analyses, the responses from all subsamples were estimated together in an unconstrained four group model (i.e., 2 × 2 [Informant × Nationality]).

Testing for DIF across subgroups. IRT DIF was used in this study because it is especially applicable to testing for invariance at the item level (see Nye, Newman, & Joseph, 2009; Embretson & Reise, 2000). Testing for uniform and nonuniform DIF using IRT is equivalent to testing for scalar and metric equivalence using other procedures such as confirmatory factor analysis (Nye et al., 2009). IRTLRDIF (Thissen & Wainer, 2001), a software application that uses the likelihood ratio method to test for DIF was used to test for DIF across the six possible pairs of Nationality × Informant type groups (i.e., Jamaican youth and parents, African American youth and parents, Jamaican and African American youth, Jamaican and African American parents, Jamaican parents and African American youth, as well as Jamaican youth and African American parents). In IRT DIF, if the probability of a positive response to an item is uniformly higher for members of one group than a second group across the entire latent continuum when both groups have identical levels of functioning, uniform DIF (also labeled "location DIF" because *b* parameter estimates vary across groups) is evident. If the response probability is lower at one point in the latent continuum, but higher elsewhere on the latent continuum for one group than the next when both groups have identical levels of functioning, nonuniform DIF (called slope DIF because the *a* parameter estimate varies across groups) is present (Marshall, Orlando, Jaycox, Foy, & Belzberg, 2002). Once items without significant DIF were identified, data from the four subgroups were placed in a database adjacent to one another (see Figure 1 for illustration) in left to right order of Jamaican youth, Jamaican parents, African American parents, and African American youth. MULTILOG (Thissen, Chen, & Bock, 2003) was used for further analyses. This IRT software application permitted cross-group linking by constraining all items without significant DIF across each adjacent pair of respondents in the database.

Performing IRT linking. The test characteristic curve method of IRT linking, developed for the graded response model, is believed to outperform other methods because all item parameter estimates in a test are used in finding linking constants via the identification of items that do not have significant DIF (Embretson & Reise, 2000). Accordingly, we took the following five steps: (a) ensured items that were identical (in content) across parent- and self-report measures were included in each Nationality × Informant group measure; (b) collected data on each Nationality × Respondent group; (c) conducted factor analyses on responses each Informant × Nationality group provided and aggregated collected data from each group, including items that were identical across groups and those that were not; (d) focused on common items across Informant × Nationality groups and used IRT procedures to identify a subset of items without significant DIF; and (e) constrained parameter estimates of identical items without significant DIF across each adjacent Nationality × Informant group, and freely estimated parameters of identical items with significant DIF and items that were unique to each Informant × Nationality group. This procedure permitted equivalence in assessment across informant and nations. Figure 1 depicts an original test that is linked to several subtests using IRT. The original test and each subsequent subtest were completed by different respondent subgroups. Moreover, all subtests were linked by only constraining parameters for items invariant across groups (i.e., items without significant DIF across each pair of adjacent subsamples in the database). This procedure permitted cross-group equivalence in item parameter estimates and allowed comparisons of scores from informants who responded to any item subset, using combinations of new and original items.

Results

Preliminary Demographics Analysis

A log-linear analysis revealed no significant difference in the number of boys versus girls across the two nations. A one-way analysis of variance with nation (i.e., Jamaica vs. United States) as the independent variable and age as the dependent variable revealed a significant effect for age, with African American adolescents being older. Nonetheless, this difference was extremely small and accounted for less than 3% of the variance.

IRT Assumptions

Appropriate dimensionality. For both U.S. and Jamaican adolescent and parent subsamples, the FIFA results revealed that a one-factor solution fit the data best. The items that loaded on the adolescent and parent Resilience scales in the previous factor analytic study (see Lambert et al., 2005) also loaded on a single Resilience factor for African American adolescents and parents, respectively, in the current study. Moreover, the current factor loadings were virtually identical to those of Lambert et al.'s previous factor analysis. Thus, for African American adolescents and parents, all 27 and 34 items, respectively, were retained.

For the Jamaican subsamples, most items loaded virtually identically to how they loaded in the earlier African American BACAH study (see Lambert et al., 2005). That is, although the Jamaican loadings for some items differed slightly (i.e., differences ranged from .01–.02) from those of the earlier African American study, most items that met the .30 or higher loading in the previous study also did so in the present study. The one-factor solution for the Jamaican adolescent sample revealed that one item did not meet the convention of loading at .30 or higher: "53. Stays away from alcohol." For the Jamaican parent sample, the following three items did not load: "42. Involved in art, music"; "43. Involved in exercise"; and "44. Involved in sports." These items were, therefore, removed from further analyses involving these respective subgroups in the Jamaican database. For both U.S. and Jamaican subsamples, although factor analytical guidelines (e.g., amount of variance accounted for) suggested that we explore at most a two-factor solution, items emerging from such analyses revealed extensive cross-loadings. After removing items that failed to load on the Resilience factor for Jamaicans, there were 24 and 31 items, respectively, remaining for adolescents and parents on which further analyses were conducted.

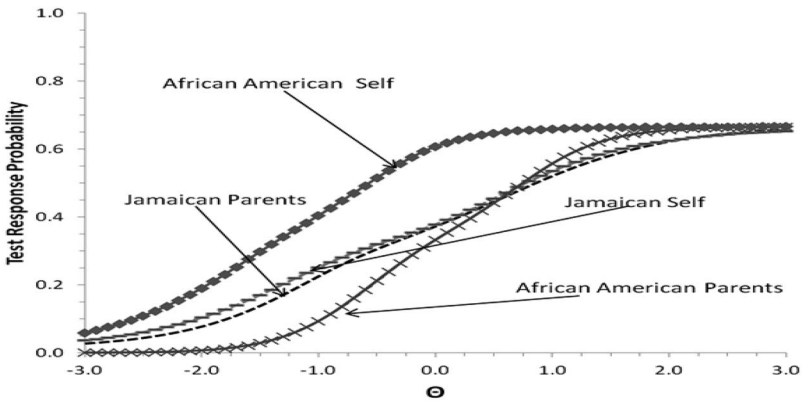

Figure 2. Behavioral Assessment for Children of African Heritage Resilience scale test characteristic curves from the unconstrained model across Jamaican and African American Parent- and Self-Reports.

Conditional independence. For each Nationality × Informant group, we compared a one-factor solution with a two-factor solution and a bifactor solution, where items loaded on both of their respective factors as well as on a bifactor. The $\Delta\chi^2$ or Δ AIC were slightly but not significantly higher for the one-factor model in all Nationality × Informant subsamples when contrasted with the two-factor solutions. Furthermore, for each Nationality × Informant group, the bifactor solution yielded a significantly higher Δ AIC than the one- or the two-factor solution. Thus, the IRT bifactor analyses revealed that for each Nationality × Informant group, the more complex two-factor or bifactor model did not fit the data better than a one-factor model. Hence, we deemed the one-factor model to be essentially unidimensional, and the findings revealed no evidence of conditional dependence.[1]

Appropriate IRT model. IRTPRO analyses of the 1PL model nested in the 2PL model revealed significant deterioration in the fit indices for the 1PL model, $\Delta\chi^2(117) = 546.45$, $\Delta AIC = 312.95$. Similar deterioration occurred for other models, such as the generalized partial credit and nominal models, where ΔAIC values were 90.74 and 113.79, respectively. This confirmed that Samejima's graded model was the most appropriate IRT model.

DIF Across Subgroups and IRT Linking

After being satisfied that the IRT assumptions were met, MULTILOG was used to estimate the parameters for all items within all groups simultaneously without constraining any item parameter estimates. TCCs derived from this analysis are displayed in Figure 2. Thus, the figure shows that TCCs for each group appear to have different slopes and especially different locations. Hence, we used the IRTLRDIF software application to identify items across pairs of respondents and across nations that were without significant DIF.

Table 1 shows that multiple items without significant DIF were identified across pairs of respondents within and across nations. Identical parameter estimates for these items without significant DIF in Table 1 were constrained across pairs of informants within and across nations. Figure 3 shows the TCCs for this partially constrained model and also reveals that all TCCs for each group have virtually indistinguishable slopes and locations. Identical TCCs emerged although some items were used to measure the Resilience construct in one group of informants but not in others, and despite there being three fewer items completed by Jamaican adolescents and parents, respectively. The TCCs demonstrate the power of using IRT linking in placing item responses from different groups of informants across nations on an equivalent metric.

Discussion

The purpose of this article was to demonstrate IRT linking across informants and nations to ensure metric and scalar equivalence in the assessment of youth. As a case example, we tested the measurement invariance of the BACAH Resilience scale across two sets of informants (i.e., parents and adolescents) residing across two nations (i.e., the United States and Jamaica). Multiple items without significant DIF emerged, allowing scale linking across nations and informants. In the remainder of this section, we discuss each major finding, highlighting the benefits of IRT linking for assessing youth adjustment across informants and nations in the current data. We then make recommendations for how future researchers can harness these benefits when using the BACAH and other measures internationally.

IRT Linking

A major difficulty in multiethnic and international cross-informant study of youth is the ability to assess youth with items that are culturally appropriate, relevant for the informant and the context, and that produce comparable scores across groups (see Lambert et al., 2005). By translating and using measures designed for youth in the United States to survey children in international contexts, many researchers (e.g., Weisz, Weiss, Suwanlert, & Chaiyasit, 2006) have not addressed these challenges. Other researchers (e.g., Rescorla et al., 2013) have relied on content-identical items on cross-informant measures and dropped other

[1] Analyses testing for conditional independence yielded extensive results. These tests merely examined one IRT assumption and are not reflective of the main focus of the study for linking the scale across informants and nations. For the sake of parsimony, details of these results are not included here. Nevertheless, they are available on request from the corresponding author.

Table 1
BACAH Cross-Informant Resilience Scale Item Parameter Estimates Linked Across Nationality and Informant Groups

	Jamaican						African American					
	Youth			Parents			Parents			Youth		
Brief item description	*a*	b_1	b_2	*a*	b_1	b_2	*a*	b_1	b_2	*a*	b_1	b_2
1. Deal with emotions	—	—	—	1.86 (0.29)	−0.93 (0.25)	1.33 (0.12)	1.95 (0.31)	−0.77 (0.21)	0.82 (0.12)	—	—	—
2. Accepts own mistakes	—	—	—	1.41 (0.27)	−1.02 (0.32)	1.38 (0.16)	1.75 (0.28)	−0.57 (0.21)	1.18 (0.15)	—	—	—
3. Adjusts to problems	—	—	—	**1.86 (0.21)**	**−0.91 (0.17)**	**0.98 (0.08)**	**1.86 (0.21)**	**−0.91 (0.17)**	**0.98 (0.08)**	—	—	—
4. Physically active	—	—	—	**0.98 (0.17)**	**−1.51 (0.37)**	**0.51 (0.14)**	**0.98 (0.17)**	**−1.51 (0.37)**	**0.51 (0.14)**	—	—	—
5. Asks for help	—	—	—	**1.66 (0.22)**	**−1.37 (0.23)**	**0.52 (0.09)**	**1.66 (0.22)**	**−1.37 (0.23)**	**0.52 (0.09)**	—	—	—
6. Attends school regularly	—	—	—	2.43 (0.28)	−1.19 (0.22)	−0.31 (0.13)	1.33 (0.22)	−1.85 (0.24)	−0.13 (0.09)	—	—	—
7. Avoids fighting	—	—	—	1.59 (0.22)	−1.85 (0.24)	−0.13 (0.09)	1.85 (0.22)	−1.13 (0.24)	0.31 (0.09)	—	—	—
8. Not make others look bad	—	—	—	1.33 (0.24)	−1.26 (0.38)	0.69 (0.14)	1.72 (0.22)	−0.85 (0.23)	0.56 (0.09)	—	—	—
9. Avoids causing trouble	—	—	—	1.53 (0.22)	−1.27 (0.24)	0.29 (0.09)	2.07 (0.28)	−0.79 (0.27)	0.41 (0.09)	—	—	—
10. Capable of learning	—	—	—	3.31 (0.55)	−1.29 (0.25)	−0.37 (0.10)	2.06 (0.30)	−2.07 (0.20)	−0.71 (0.09)	—	—	—
11. Carries self respectably	—	—	—	3.21 (0.54)	−1.16 (0.24)	−0.01 (0.08)	2.69 (0.33)	−1.37 (0.21)	−0.07 (0.09)	—	—	—
12. Communicates with adults	—	—	—	1.85 (0.28)	−1.13 (0.27)	0.31 (0.09)	2.10 (0.40)	−1.35 (0.33)	0.45 (0.12)	—	—	—
13. Concerned about others	—	—	—	2.10 (0.35)	−1.35 (0.26)	0.45 (0.10)	1.98 (0.40)	−1.41 (0.33)	0.45 (0.12)	—	—	—
14. Develops routine	—	—	—	**1.52 (0.20)**	**−0.93 (0.21)**	**1.00 (0.10)**	**1.52 (0.20)**	**−0.93 (0.21)**	**1.00 (0.10)**	—	—	—
15. Does best in school	—	—	—	2.02 (0.20)	−1.11 (0.18)	0.44 (0.07)	1.88 (0.43)	−1.02 (0.21)	0.26 (0.10)	—	—	—
16. Independent schoolwork	—	—	—	**2.02 (0.30)**	**−1.19 (0.20)**	**0.18 (0.09)**	**2.02 (0.30)**	**−1.19 (0.20)**	**0.18 (0.09)**	—	—	—
18. Expresses pride	**1.38 (0.19)**	**−1.25 (0.31)**	**0.53 (0.11)**	—	—	—	—	—	—	**1.38 (0.19)**	**−1.25 (0.31)**	**0.53 (0.11)**
19. Expresses feelings	1.47 (0.27)	−1.53 (0.33)	0.35 (0.14)	—	—	—	—	—	—	2.13 (0.42)	−0.68 (0.31)	0.54 (0.13)
25. Well with young children	**1.20 (0.19)**	**−1.95 (0.41)**	**0.38 (0.14)**	—	—	—	—	—	—	**1.20 (0.19)**	**−1.95 (0.41)**	**0.38 (0.14)**
33. Has a sense of humor	—	—	—	1.54 (0.20)	−2.00 (0.40)	0.01 (0.16)	1.92 (0.38)	−1.78 (0.40)	−0.17 (0.38)	—	—	—
34. Has variety of interests	**1.54 (0.20)**	**−1.62 (0.31)**	**0.15 (0.11)**	**1.54 (0.20)**	**−1.62 (0.31)**	**0.15 (0.11)**	**2.00 (0.27)**	**−1.30 (0.23)**	**0.07 (0.10)**	**2.00 (0.27)**	**−1.30 (0.23)**	**0.07 (0.10)**
35. Good physical health	**1.69 (0.22)**	**−1.41 (0.30)**	**0.30 (0.11)**	—	—	—	—	—	—	**1.69 (0.22)**	**−1.41 (0.30)**	**0.30 (0.11)**
36. Has a variety of friends	—	—	—	**1.02 (0.16)**	**−1.87 (0.46)**	**0.22 (0.16)**	**1.02 (0.16)**	**−1.87 (0.46)**	**0.22 (0.16)**	—	—	—
37. Has confidence	**1.56 (0.19)**	**−1.51 (0.26)**	**0.18 (0.10)**	**1.56 (0.19)**	**−1.51 (0.26)**	**0.18 (0.10)**	**1.64 (0.23)**	**−1.47 (0.33)**	**0.01 (0.13)**	**1.64 (0.23)**	**−1.47 (0.33)**	**0.01 (0.13)**
38. Has a positive attitude	2.53 (0.31)	−1.17 (0.25)	0.20 (0.17)	2.38 (0.38)	−1.59 (0.26)	0.23 (0.11)	**2.44 (0.33)**	**−0.97 (0.21)**	**0.07 (0.09)**	**2.44 (0.33)**	**−0.97 (0.21)**	**0.07 (0.09)**
40. Behavior consistent	**1.16 (0.15)**	**−1.95 (0.37)**	**1.11 (0.12)**	**1.16 (0.15)**	**−1.95 (0.37)**	**1.11 (0.12)**	**1.54 (0.20)**	**−1.51 (0.30)**	**0.33 (0.11)**	**1.54 (0.20)**	**−1.51 (0.30)**	**0.33 (0.11)**
41. History and culture interest	**0.86 (0.14)**	**−1.33 (0.36)**	**1.38 (0.17)**	**0.86 (0.14)**	**−1.33 (0.36)**	**1.38 (0.17)**	**1.59 (0.20)**	**−1.01 (0.24)**	**0.57 (0.10)**	**1.59 (0.20)**	**−1.01 (0.24)**	**0.57 (0.10)**
42. Involved in art, music	0.67 (0.19)	−1.53 (0.70)	1.65 (0.33)	—	—	—	**0.84(16)**	**−1.05 (0.40)**	**0.42 (0.19)**	**0.84(16)**	**−1.05 (0.40)**	**0.42 (0.19)**
43. Involved in exercise	0.88 (0.19)	−1.34 (0.50)	1.37 (0.22)	—	—	—	**1.28 (0.18)**	**−0.63 (0.22)**	**0.67 (0.12)**	**1.28 (0.18)**	**−0.63 (0.22)**	**0.67 (0.12)**
44. Involved in sports	0.80 (0.21)	−1.23 (0.59)	0.50 (0.24)	—	—	—	**0.91 (0.18)**	**−1.13 (0.39)**	**0.17 (0.20)**	**0.91 (0.18)**	**−1.13 (0.39)**	**0.17 (0.20)**
45. Involved socially	1.13 (0.24)	−1.66 (0.52)	0.25 (0.20)	1.00 (0.19)	−1.93 (0.26)	0.65 (0.10)	**1.52 (0.22)**	**−1.31 (0.31)**	**0.09 (0.14)**	**1.52 (0.22)**	**−1.31 (0.31)**	**0.09 (0.14)**
46. Loving/affectionate	1.65 (0.25)	−1.26 (0.25)	0.49 (0.12)	**1.71 (0.34)**	**−1.52 (0.28)**	**0.06 (0.11)**	**1.71 (0.34)**	**−1.52 (0.28)**	**0.06 (0.11)**	1.80 (0.39)	−1.74 (0.58)	0.24 (0.16)
47. Motivated, hardworking	**1.95 (0.21)**	**−1.32 (0.21)**	**0.42 (0.08**	**1.95 (0.21)**	**−1.32 (0.21)**	**0.42 (0.08)**	**2.42 (0.49)**	**−0.93 (0.18)**	**0.34 (0.08)**	**2.42 (0.49)**	**−0.93 (0.18)**	**0.34 (0.08)**
48. Obedient	**1.35 (0.16)**	**−1.71 (0.32)**	**0.72 (0.10)**	**1.35 (0.16)**	**-1.71 (0.32)**	**0.72 (0.10)**	**2.21 (0.32)**	**−1.18 (0.23)**	**0.45 (0.08)**	**2.21 (0.32)**	**−1.18 (0.23)**	**0.45 (0.08)**
49. Practices good hygiene	—	—	—	**1.71 (0.25)**	**−1.74 (0.34)**	**−0.11 (0.12)**	**1.71 (0.25)**	**−1.74 (0.34)**	**−0.11 (0.12)**	—	—	—
50. Responsible	**1.73 (0.22)**	**−1.36 (0.23)**	**0.71 (0.08)**	**1.73 (0.22)**	**−1.36 (0.23)**	**0.71 (0.08)**	2.10 (0.35)	−1.51 (0.34)	0.53 (0.12)	1.09 (0.38)	−3.13 (0.79)	−0.44 (0.41)
51. Sets appropriate goals	2.20 (0.31)	−0.99 (0.23)	0.33 (0.10)	—	—	—	—	—	—	1.28 (0.33)	−2.42 (0.67)	−0.22 (0.30)
53. Stays away from alcohol	—	—	—	—	—	—	—	—	—	1.12 (0.44)	−1.78 (0.58)	−0.85 (0.60)
54. Separate from my parents	**0.79 (0.16)**	**−1.68 (0.55)**	**0.58 (0.18)**	—	—	—	—	—	—	**0.79 (0.16)**	**−1.68 (0.55)**	**0.58 (0.18)**
55. Takes leadership roles	1.50 (0.25)	−1.09 (0.32)	0.57 (0.13)	**1.45 (0.21)**	**−1.28 (0.24)**	**0.69 (0.10)**	**1.45 (0.21)**	**−1.28 (0.24)**	**0.69 (0.10)**	2.15 (0.43)	−0.55 (0.26)	0.36 (0.13)
56. Accept differences	**1.69 (0.23)**	**−1.48 (0.28)**	**0.63 (0.09)**	—	—	—	—	—	—	**1.69 (0.23)**	**−1.48 (0.28)**	**0.63 (0.09)**
57. Try things on my own	**1.31 (0.21)**	**−2.24 (0.48)**	**0.04 (0.15)**	—	—	—	—	—	—	**1.31 (0.21)**	**−2.24 (0.48)**	**0.04 (0.15)**
59. Appropriate role models	—	—	—	—	—	—	—	—	—	2.15 (0.43)	−0.57 (0.26)	0.35 (0.13)
60. Shows problem solving	—	—	—	—	—	—	—	—	—	1.30 (0.21)	−2.27 (0.48)	0.03 (0.15)

Table 1 (*continued*)

Brief item description	Jamaican Youth a	Jamaican Youth b_1	Jamaican Youth b_2	Jamaican Parents a	Jamaican Parents b_1	Jamaican Parents b_2	African American Parents a	African American Parents b_1	African American Parents b_2	African American Youth a	African American Youth b_1	African American Youth b_2
61. Accepts spirituality	**0.74 (0.24)**	**−3.63 (0.55)**	**0.60 (0.51)**	—	—	—	—	—	—	**0.74 (0.24)**	**−3.63 (0.55)**	**0.60 (0.51)**
63. Good reasoning skills	1.61 (0.27)	−1.46 (0.38)	0.35 (0.13)	—	—	—	—	—	—	2.33 (0.50)	−0.88 (0.32)	0.53 (0.12)
64. Has hobby	—	—	—	**1.55 (0.22)**	**−1.42 (0.26)**	**0.24 (0.11)**	**1.55 (0.22)**	**−1.42 (0.26)**	**0.24 (0.11)**	—	—	—

Note. Standard error estimates shown in parentheses. Numbers on the left hand side of brief item descriptions are the item numbers as they appeared on the BACAH and are *not* the number of items that comprise the resilience scale. BACAH = Behavioral Assessment for Children of African Heritage; a = discrimination parameter estimate; b_1 and b_2 = location parameter estimates; adjacent items across any group of informants within or across any nation with identical parameter estimates are items without significant differential item functioning and are listed in boldface font. Such items were constrained to place items across all groups on an equivalent metric; — = parameters for items that were not estimated (i.e., either because they were part of the dimension but not administered to Jamaican informants, not part of the Resilience dimension for one group of informants, or were dropped because they did not load on the dimension for Jamaican informants). Specifically, Items "59. Appropriate role models" and "60. Shows problem solving" were not administered to Jamaican informants; Item "53. Stays away from alcohol" was dropped from the Jamaican adolescent reports because it failed to load on Resilience dimension in the full information factor analyses and Items "42. Involved in art, music," "43. Involved in exercise," and "44. Involved in sports" were dropped from the Jamaican parent report data because factor analyses revealed that they did not load on the Resilience factor. BACAH = Behavioral Assessment for Children of African Heritage.

items that are not identical. The present study demonstrated how IRT linking can address the stated challenges of reducing the effects of psychometric noninvariance and improving metric and scalar equivalence in cross-informant cross-national research while maximizing item retention. We demonstrated that linking item scores can place all items across informants and nations on a common scale (see Arai & Mayekawa, 2011). As demonstrated by Figures 2 and 3, the TCCs for parents and adolescents in African American and Jamaican samples are virtually identical after linking. This study, therefore, demonstrates the power of IRT linking in transforming informant-specific scores to a common scale by using a set of measures whose items are not completely identical across reporters or nations.

Methodological procedures used in this study might also be applicable to studies that use other culturally sensitive cross-informant measures to compare functioning of youth cross-nationally or across different socioethnic groups within the same nation. By employing this methodology, researchers could design new tests. They could also use existing measures of identical constructs that are culturally valid for each group studied (e.g., written to match groups' idiomatic expressions and reality). This procedure would entail the selection of items with identical content across measures for each informant and socioethnic group studied where statistical analyses reveal no significant DIF. By constraining such items, while allowing other items (i.e., those with significant DIF and items that are not identical across informants or demographic groups) to be estimated freely, all items can be placed on an equivalent metric and scale. This procedure could reduce bias caused by measurement noninvariance in culturally appropriate cross-informant assessment of youth across different socioethnic groups.

IRT and its linking procedures can also allow researchers to build and to continually modify as well as extend item test banks for computerized adaptive testing (CAT) administration (see Lai et al., 2011). Items and their precalibrated parameter estimates are entered into CAT software applications and, with the use of its algorithms, CAT software scores informant responses as they are submitted in real time. From this score, it determines whether an item reflecting lower or higher levels of functioning should be automatically administered. For example, if the informant responds positively to an item measuring moderate levels of Resilience, the CAT software would next administer items reflecting higher Resilience levels. If the informant's rating to such an item is zero, items measuring lower Resilience would then be automatically administered. The software would continue scoring each item administered up to a predetermined stopping rule (e.g., number of items positively endorsed or not or when a prespecified standard error of measurement is met). Therefore, CAT permits economical test administration that allows each respondent to be administered tailored tests based on items that reflect the adolescent's reported functioning. Yet, because items are administered from a precalibrated item bank, their scores can be compared (see Arai & Mayekawa, 2011).

Study-specific findings. As presented in the Method section (see Performing IRT Linking subsection under Data Analyses), IRT linking provides a powerful means of developing and refining item banks. Hence, it permits researchers to identify and remove items that might provide limited information as well as to add items that might provide better measurement

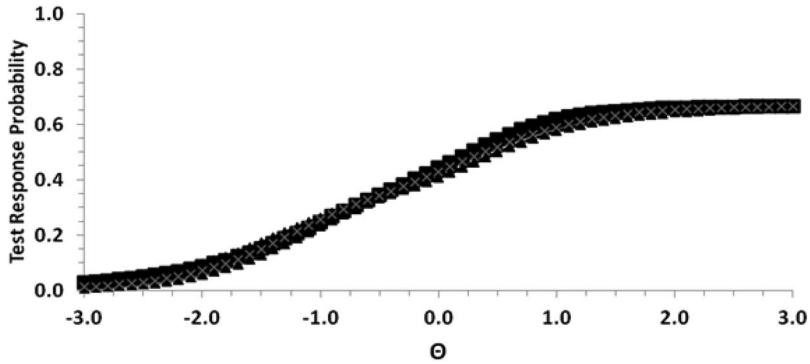

Figure 3. Behavioral Assessment for Children of African Heritage Resilience scale test characteristic curves from the partially constrained model across Jamaican and African American Parent- and Self-Reports.

precision. Items whose *a* parameter estimates 1 or higher are generally accepted as capable of capturing acceptable amounts of information (see Olino et al., 2013). Table 1 shows that most BACAH Resilience scale items meet this criterion. Nevertheless, items such as "42. Involved in art, music" for Jamaican adolescents might at least warrant further study to determine the reason for their poor discrimination. For example, we suspect that because, on the whole, Jamaican adolescents are highly immersed in the arts, this item does not provide much discrimination for Jamaican youth due to low sample variance. Because IRT discrimination parameter is similar to a factor loading, we saw that, while Item 42 did load on Jamaican adolescents' reports, this item along with Items 43 and 44 did not load on the Resilience factor for Jamaican parent reports, suggesting that they offer virtually no discrimination when parent reports are considered. Qualitative studies that interview Jamaican parents and youth about their thoughts while reading such items might provide researchers with clues about reasons for the poor performance of such items as well as whether such items might benefit from modification and further psychometric study or whether removal from the scale is warranted.

Levels of Functioning Scale Measures

In developing measures for clinical assessment and research, test developers often prefer that their instruments are capable of assessing individuals with low to high levels of the trait being measured (Fischer, Tritt, Klapp, & Fliege, 2011). Examining the value of location (*b*) parameter estimates can be especially helpful to researchers in evaluating the lower to upper bounds of trait levels the test is capable of measuring. If the lowest *b* parameter estimate is higher than the researcher would like, the measure might be considered as not being capable of measuring individuals with low trait levels. If the highest is lower than the researcher would like, the measure might be considered as not having the capacity to measure functioning of individuals with higher trait levels. Within this context, we note that, for Jamaican adolescent reporters, the items are capable of measuring respondents whose level of resilience range from approximately 2 *SD* below the mean to approximately 1 *SD* above the mean. Similar findings were also evident for Jamaican parent informants and African American parent informants. Parameter estimates for African American adolescents showed that the items seem capable of measuring adolescents whose level of functioning is as low as 3 *SD* below the mean but not more than approximately 0.5 *SD* above the mean.

These findings suggest the need for the addition of items that permit assessment of all respondents across nations, but especially for African American adolescent self-reports. Because linking allows the economical addition of items, this procedure might be especially useful in reducing ceiling effects in the BACAH Resilience scale. Linking additional items to increase the number of items in the pool might also be warranted if CAT application is also an objective. Yet, it has been demonstrated that in "personality"-type testing, where the number of potential items is limited, substantial time savings can be generated from CAT administration using item banks of relatively small numbers of Likert scale items (see Lai et al., 2011). Hence, if adolescents with low-to-moderate levels of Resilience (e.g., those clinic-referred adolescents) are assessed, reliable and valid test data could be obtained from a few items drawn from an item bank to match their levels of functioning (Lai et al., 2011).

Invariance

In our case example of BACAH data, findings revealed several items with significant DIF and multiple items without significant DIF across informants and nations. The TCCs in Figure 2 further show that ignoring cross-informant and cross-national measurement invariance concerns in international studies could be problematic, and might even lead to spurious results. That is, findings could be adversely impacted by significant DIF and might not accurately represent similarities or differences in test scores across groups (Thomas, 2011).

The absence of invariance for several items in this international cross-informant study implies that, even when researchers study groups across different nations who have similar heritage, an assumption of invariance in response styles could be erroneous. As Lambert et al. (2007) have shown, differences in response styles can be even more profound when comparing responses from youth of different heritage residing across countries (e.g., German White vs. Jamaican Black adolescents). The burden is, therefore, on researchers who study youth of similar and different heritage cross-nationally to demonstrate that their findings are at least minimally impacted by invariance-related measurement artifacts.

Further inspection of Figures 2 and 3 highlights the importance of this statement. Figure 2 shows marked differences in the TCCs for Jamaican versus African American parent- and adolescent self-report response styles. That is, examining the differences in positions of curves in Figure 1 versus Figure 2 if resilience trait levels of adolescents from both nations are held constant, African American parents overreport on their children's functioning. By contrast, Jamaican adolescents underreport their functioning on this construct. Merely comparing parent and/or adolescent self-reported raw scores across both nations without attention to such invariance issues could exaggerate differences where African American parents report significantly higher levels of Resilience in their children when little or no such differences exist. It is also possible that if Jamaican adolescents have significantly higher levels of Resilience than their African American counterparts, such differences might be obscured due to significant differences in parent reporting styles. By contrast, the converse might be true for adolescent self-reports. Figure 3 further shows that location- and even slope-related measurement artifacts are markedly reduced when scores are linked across informants and nations because the test response functions are virtually identical across informants and nations.

An examination of Figure 2 shows that TCCs for Jamaican parent- versus youth self-reports are virtually identical, whereas those for African Americans are not. Though unexpected, this finding shows more convergence in rating styles across Jamaican than African American informants. It is possible that Jamaican adolescents and parents are simply more "in tune" with each other such that parent reports of their adolescents' adjustments are closer to the adolescent self-reports than is the case among African American families. In such situations when IRT findings show that one group underreports or overreports, it is important to interpret this in light of known cultural contextual differences between groups. In the case wherein Jamaican parents primarily underreport their children's strengths, the first consideration is that both Jamaicans and African Americans originate from highly collectivistic cultures and have maintained most of this cultural mores throughout their centuries in the Americas (Allen & Bagozzi, 2001). Collectivism typically inhibits the expression of positive individual attributes and "standing out" more than one's fellow citizens. It is likely that, while both groups maintain a collectivistic stance, the influence of living in a more individualistic nation might temper the collectivistic behavior of African American adults somewhat and increase their comfort levels on reporting their children's strengths. Oyserman, Coon, and Kemmelmeier (2002) have provided empirical evidence of considerable individualism in African Americans. By contrast, Jamaicans parents are more isolated from the effects of individualism and might find it less comfortable to report on their children's strengths. The finding that, in the absence of linking the TCC, shows Jamaican youth are more likely to overreport on their Resilience is puzzling (see Figure 2). Yet more recent research has shown that many Jamaican adolescents have remotely acculturated to aspects of the White U.S. individualistic cultural stance, and exceeded their mothers in this regard (Ferguson & Bornstein, 2012). For some Jamaican youth, this attitude is therefore at odds with their parents who show a more traditional collectivistic stance.

Limitations and Future Research

Findings from the present study must be interpreted within the context of its limitations. A broad-based limitation refers to the general use of IRT procedures, including those involved in linking. Although many IRT practitioners (e.g., Embretson & Reise, 2000) have argued about its superiority to traditional (i.e., classical) test theory and the methods it informs, IRT is far more difficult and expensive to implement. That is, it requires extensive training on the part of researchers. Besides the cost associated with using larger samples required for IRT study, user-friendly IRT software is only slowly becoming available. Because IRT statistical procedures are not typically integrated in widely used software packages, specific software packages must often be purchased and many are difficult to learn (see Streiner, 2010). Indeed, besides SPSS, this study required the purchase of two packages (i.e., MULTILOG and IRTPRO) and use one shareware package (IRTLRDIF). In addition, because linking is often intricately linked to CAT administration and we have referenced this as a potential test administration application for the BACAH Resilience scale, we note that CAT also requires the availability of computers and the purchase of expensive software to implement.

Turning to study-specific limitations, first we acknowledge that the study focused on a single cross-informant strength dimension (i.e., Resilience) and its findings cannot be generalized to other dimensions of strengths and problems. Second, teacher reports were excluded from this study. Hence, information on youth in the school context is missing, which did not permit linking of scales across all three informants. Third, the study focused on nonreferred youth. Hence, the findings might not generalize to youth who are identified as having behavioral and emotional difficulties.

Another concern is that, although by some standards the size of each subsample is adequate, by others they might be viewed as relatively small (see Ghada, 2005). Simulation studies have shown that smaller sample size could impact the precision of item parameter estimates (see Kankaraš, Vermunt, & Moors, 2011). Yet others (e.g., Edelen & Reeve, 2007; Hula, Fergadiotis, & Martin, 2012) have argued and shown that, unlike large samples needed to calibrate item parameters for use in CAT, smaller samples (e.g., $\geq$200) are sufficient for hypothesis testing such as DIF across groups. In addition, other researchers have shown that, if item response data meet IRT assumptions and if parameters from samples studied are estimated simultaneously (both of which are true for this study), smaller sample sizes can be appropriate (see Chen, Revicki, Lai, Cook, & Amtmann, 2009).

It might be argued that the number of items that were constrained across adjacent groups is small and furthermore no items were constrained across all four groups. We note that, although more items eligible for constraints could potentially lead to more stable item calibrations, there are no consistent rules on the number or proportion of items in the item pool that should be eligible for constraint (Chen et al., 2009). Additionally, there is precedence for calibration of multiple item pools to form item banks where linking is merely done across adjacent subsets. Such item banks are considered well calibrated (see Arai & Mayekawa, 2011).

Finally, the BACAH forms were developed with considerable input from the African American community to measure functioning in African American youth. Multiple experienced Jamaican researchers and clinicians have repeatedly suggested that the forms

are appropriate for youth in their country. Nevertheless, it is possible that, despite our extensive efforts to make the measures culturally valid for African Americans, important items that measure Resilience in this group might have been omitted. We are aware that this probability might be even greater for Jamaicans, where we did not have such community input. Hence, it is possible that items of strengths that are culture-specific for Jamaican youth might have been excluded.

Further empirical investigation is necessary to build on the findings of this study. To address the inclusion of additional Resilience scale items that could be culturally relevant for Jamaica, qualitative research with Jamaican adolescents, parents, teachers, clinicians, and other key informants could be conducted. Information could be gathered to first determine the relevance of items on the existing Resilience scale and suggest item modifications, removal, and additions. IRT linking studies could then be conducted to place new items on an equivalent metric to existing items. Alternatively, researchers could employ similar procedures used in the development of the BACAH to create new cross-informant measures of strength and problem constructs for youth in different nations. Professionals in the respective nations could be asked to ensure that a substantial number of identical items are included across measures (e.g., across existing, modified, or entirely new measures) with the hope that many such identical items will show measurement invariance across informants within and across the nations of focus (Muraki, Hombo, & Lee, 2000). Next, identical procedures used in the present study could be employed to test for significant DIF in items across reporters within and across nations. Finally, by constraining identical items in each dimension without significant DIF, sets of items designed for two or more groups could be placed on an equivalent metric to permit unbiased comparisons of functioning across informants and across nations.

Conclusion

Despite the need for further research, the present study draws attention to the tremendous psychometric responsibility that rests on the shoulders of researchers who compare the functioning of two or more socioethnic groups of youth. That is, researchers conducting comparative studies are charged with ensuring that their findings are as free as possible of measurement artifacts emerging from the lack of cross-group metric and scalar equivalence. Cross-informant assessment, the gold standard for assessing youth, places further responsibility on researchers because it increases the number of groups that need linking. This burden is even greater when youth from different nations are studied. The present study included two sets of informants across two nations and resulted in the need to link four groups. Inclusion of teacher reports would add two additional groups. We hope our efforts to demonstrate one method of linking the measurement of a cross-informant construct in African diaspora youth across two nations can scaffold further research to reduce the bias in cross-informant cross-national assessment.

References

Achenbach, T. M., & Rescorla, L. A. (2001). *Manual for the ASEBA school-aged forms and profiles*. Burlington, VT: University of Vermont, Research Center for Children, Youth and Families.

Achenbach, T. M., & Rescorla, L. A. (2004). The Achenbach System of Empirically Based Assessment for ages 1.5 to 18 years. In M. E. Maurish (Ed.), *The use of psychological testing for treatment planning and outcomes assessment: Vol. 2*: *Instruments for children and adolescents* (3rd ed., pp. 179–213). Mahwah, NJ: Erlbaum.

Allen, R. L., & Bagozzi, R. P. (2001). Consequences of Black sense of self. *Journal of Black Psychology, 27,* 3–28. http://dx.doi.org/10.1177/0095798401027001001

Arai, S., & Mayekawa, S. (2011). A comparison of equating methods and linking designs for developing an item pool under item response theory. *Behaviormetrika, 38,* 1–16. http://dx.doi.org/10.2333/bhmk.38.1

Berry, J. W., Phinney, J. S., Sam, D. L., & Vedder, P. (2006). Immigrant youth: Acculturation, identity, and adaptation. *Applied Psychology: An International Review, 55,* 303–332. http://dx.doi.org/10.1111/j.1464-0597.2006.00256.x

Brown, T. E. (2001). *Manual for the Brown Attention-Deficit Scales*. San Antonio, TX. Psychological Corporation.

Campbell, H. L., Barry, C. L., Joe, J. N., & Finney, S. J. (2008). Configural, metric, and scalar invariance of the Modified Achievement Goal Questionnaire across African American and White university students. *Educational and Psychological Measurement, 68,* 988–1007. http://dx.doi.org/10.1177/0013164408318766

Carter, A. S., Briggs-Gowan, M. J., & Davis, N. O. (2004). Assessment of young children's social-emotional development and psychopathology: Recent advances and recommendations for practice. *Journal of Child Psychology and Psychiatry, 45,* 109–134. http://dx.doi.org/10.1046/j.0021-9630.2003.00316.x

Chatters, L. M., Taylor, R. J., Jackson, J. S., & Lincoln, K. D. (2008). Religious coping among African Americans, Caribbean Blacks and non-Hispanic Whites. *Journal of Community Psychology, 36,* 371–386. http://dx.doi.org/10.1002/jcop.20202

Chen, W. H., Revicki, D. A., Lai, J. S., Cook, K. F., & Amtmann, D. (2009). Linking pain items from two studies onto a common scale using item response theory. *Journal of Pain and Symptom Management, 38,* 515–628.

Conners, C. K. (2008). *Manual for the Conners' Rating Scales* (3rd ed.). Toronto, ON, Canada: Multi-health Systems.

de Ayala, R. J. (2009). *The theory and practice of item response theory*. New York, NY: Guilford Press.

Edelen, M. O., & Reeve, B. B. (2007). Applying item response theory (IRT) modeling to questionnaire development, evaluation, and refinement. *Quality of Life Research, 16,* 5–18. http://dx.doi.org/10.1007/s11136-007–9198-0

Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists*. Mahwah, NJ: Erlbaum.

Ferguson, G. M., & Bornstein, M. H. (2012). Remote acculturation: The "Americanization" of Jamaican islanders. *International Journal of Behavioral Development, 36,* 167–177. http://dx.doi.org/10.1007/s11136-007-9198-0

Fischer, H. F., Tritt, K., Klapp, B. F., & Fliege, H. (2011). How to compare scores from different depression scales: Equating the Patient Health Questionnaire (PHQ) and the ICD-10-Symptom Rating (ISR) using item response theory. *International Journal of Methods in Psychiatric Research, 20,* 203–214. http://dx.doi.org/10.1002/mpr.350

Georgas, J., Berry, J. W., van de Vijver, F. J. R., Kagitçibasi, Ç., & Poortinga, Y. H. (Eds.). (2006). *Families across cultures: A 30-nation psychological study*. New York, NY: Cambridge University Press.

Ghada, K. E. (2005). The effects of sample size on equating. *Education, 126,* 165–180.

Gibbons, R. D., Bock, R. D., Hedeker, D., Weiss, D. J., Segawa, E., Bhaumik, D. K., . . . Stover, A. (2007). Full-information item bifactor analysis of graded response data. *Applied Psychological Measurement, 31,* 4–19. http://dx.doi.org/10.1177/0146621606289485

Goldstein, S., & Brooks, R. B. (2013). Why study resilience? In S. Goldstein & R. B. Brooks (Eds.), *Handbook of resilience in children* (2nd ed., pp. 3–14). New York, NY: Springer Science + Business Media. http://dx.doi.org/10.1007/978-1-4614-3661-4_1

Hollingshead, A. B. (1975). *The four-factor index of social status*. Unpublished manuscript, Department of Sociology, Yale University, New Haven, CT.

Hula, W. D., Fergadiotis, G., & Martin, N. (2012). Model choice and sample size in item response theory analysis of aphasia tests. *American Journal of Speech-Language Pathology, 21*(Supp.), S38–S50. http://dx.doi.org/10.1044/1058-0360(2011/11-0090)

IRTPRO. (2011). *Version 2.1* [Computer software]. Skokie, IL: Scientific Software International.

Ivanova, M. Y., Achenbach, T. M., Rescorla, L. A., Dumenci, L., Almqvist, F., Bilenberg, N., . . . Verhulst, F. C. (2007). The generalizability of the Youth Self-Report syndrome structure in 23 societies. *Journal of Consulting and Clinical Psychology, 75,* 729–738. http://dx.doi.org/10.1037/0022-006X.75.5.729

Kankaraš, M., Vermunt, J. K., & Moors, G. (2011). Measurement equivalence of ordinal items: A comparison of factor analytic, item response theory, and latent class approaches. *Sociological Methods & Research, 40,* 279–310. http://dx.doi.org/10.1177/0049124111405301

Kim, S. Y., Wang, Y., Weaver, S. R., Shen, Y., Wu-Seibold, N., & Liu, C. H. (2014). Measurement equivalence of the Language-Brokering Scale for Chinese American adolescents and their parents. *Journal of Family Psychology, 28,* 180–192. http://dx.doi.org/10.1037/a0036030

King, K. M., Khan, N., Leblanc, P., & Quan, H. (2011). Examining and establishing translational and conceptual equivalence of survey questionnaires for a multi-ethnic, multi-language study. *Journal of Advanced Nursing, 67,* 2267–2274. http://dx.doi.org/10.1111/j.1365-2648.2011.05679.x

Lai, J. S., Butt, Z., Zelko, F., Cella, D., Krull, K. R., Kieran, M. W., & Goldman, S. (2011). Development of a parent-report cognitive function item bank using item response theory and exploration of its clinical utility in computerized adaptive testing. *Journal of Pediatric Psychology, 36,* 766–779. http://dx.doi.org/10.1093/jpepsy/jsr005

Lambert, M. C., Essau, C. A., Schmitt, N., & Samms-Vaughan, M. E. (2007). Dimensionality and psychometric invariance of the Youth Self-Report Form of the Child Behavior Checklist in cross-national settings. *Assessment, 14,* 231–245. http://dx.doi.org/10.1177/1073191107302036

Lambert, M. C., Rowan, G. T., Kim, S., Rowan, S. A., An, J. S., Kirsch, E. A., & Williams, O. (2005). Assessment of behavioral and emotional strengths in Black children: Development of the Behavioral Assessment for Children of African Heritage. *Journal of Black Psychology, 31,* 321–351. http://dx.doi.org/10.1177/0095798405278409

Lambert, M. C., Rowan, G. T., Rowan, S. A., & Mount, D. L. (2014). Behavioral and Emotional Problem scales of the Behavioral Assessment for Children of African Heritage (BACAH): Culturally valid measurement of psychological functioning in Black children. *Journal of Black Psychology, 40,* 451–482. http://dx.doi.org/10.1177/0095798413497512

Liu, X. (2010). *Using and developing measurement instruments in science education: A Rasch modeling approach*. Charlotte, NC: Information age Publishing.

Liu, Y., & Thissen, D. (2012). Identifying local dependence with a score test statistic based on the bifactor logistic model. *Applied Psychological Measurement, 36,* 670–688. http://dx.doi.org/10.1177/0146621612458174

Marshall, G. N., Orlando, M., Jaycox, L. H., Foy, D. W., & Belzberg, H. (2002). Development and validation of a modified version of the Peritraumatic Dissociative Experiences Questionnaire. *Psychological Assessment, 14,* 123–134. http://dx.doi.org/10.1037/1040-3590.14.2.123

Muraki, E., Hombo, C. M., & Lee, Y. W. (2000). Equating and linking of performance assessments. *Applied Psychological Measurement, 24,* 325–337. http://dx.doi.org/10.1177/01466210022031787

Nicolas, G., Helms, J. E., Jernigan, M. M., Sass, T., Skryzpek, A., & DeSilvia, A. M. (2008). A conceptual framework for understanding the strengths of Black youth. *Journal of Black Psychology, 34,* 261–280. http://dx.doi.org/10.1177/0095798408316794

Nye, C. D., Newman, D. A., & Joseph, D. L. (2009). Never say "always"? Extreme item wording effects on scalar invariance and item response curves. *Organizational Research Methods, 13,* 806–830. http://dx.doi.org/10.1177/1094428109349512

Olino, T. M., You, L., McKain, D. L., Forbes, E. E., Seely, J. R., Lewinsohn, P. M., & Pilkonis, P. A. (2013). Comparisons across depression assessment instruments in adolescence and young adulthood: Item response theory using two linking methods. *Journal of Abnormal Child Psychology, 41,* 1267–1277. http://dx.doi.org/10.1007/s10802-013-9756-6

Olsen, M., Jensen, N. K., Tesfaye, M., & Holm, L. (2013). Conceptual equivalence of WHOQOL-HIV among people living with HIV in Ethiopia. *Quality of Life Research, 22,* 261–367. http://dx.doi.org/10.1007/s11136-012-0141-7

Oyserman, D., Coon, H. M., & Kemmelmeier, M. (2002). Rethinking individualism and collectivism: Evaluation of theoretical assumptions and meta-analyses. *Psychological Bulletin, 128,* 3–72.

Panter, A. T., & Reeve, B. B. (2002). Assessing tobacco beliefs among youth using item response theory models. *Drug and Alcohol Dependence, 68*(Supp. 1), 21–39. http://dx.doi.org/10.1016/S0376-8716(02)00213-2

Reeve, B. B. (2002). *An introduction to modern measurement theory*. Bethesda, MD: National Cancer Institute.

Reise, S. P., Ventura, J., Keefe, R. S., Baade, L. E., Gold, J. M., Green, M. F., . . . Bilder, R. (2011). Bifactor and item response theory analyses of interviewer report scales of cognitive impairment in schizophrenia. *Psychological Assessment, 23,* 245–261. http://dx.doi.org/10.1037/a0021501

Renk, K. (2005). Cross-informant ratings of the behavior of children and adolescents: The "gold standard." *Journal of Child and Family Studies, 14,* 457–468. http://dx.doi.org/10.1007/s10826-005-7182-2

Rescorla, L., Achenbach, T., Ivanova, M. Y., Dumenci, L., Almqvist, F., Bilenberg, L., . . . Verhulst, F. (2007). Behavioral and emotional problems reported by parents of children ages 6 to 16 in 31 societies. *Journal of Emotional and Behavioral Disorders, 15,* 130–142.

Rescorla, L. A., Ginzburg, S., Achenbach, T. M., Ivanova, M. Y., Almqvist, F., Begovac, I., . . . Verhulst, F. C. (2013). Cross-informant agreement between parent-reported and adolescent self-reported problems in 25 societies. *Journal of Clinical Child & Adolescent Psychology, 42,* 262–273. http://dx.doi.org/10.1080/15374416.2012.717870

Reynolds, C. R., & Kamphaus, R. W. (2004). *Manual for Behavior Assessment System for Children* (2nd ed.). Minneapolis, MN: Pearson.

Robinson, L. (2001). A conceptual framework for social work practice with Black children and adolescents in the United Kingdom. *Journal of Social Work, 1,* 165–185. http://dx.doi.org/10.1177/146801730100100204

Samejima, F. (2010). The graded response model. In M. L. Nering & R. Ostini (Eds.), *Handbook of polytomous item response theory models* (pp. 77–107). New York, NY: Routledge.

Schmitt, N., & Kuljanin, G. (2008). Measurement invariance: Review of practice and implications. *Human Resource Management Review, 18,* 210–222. http://dx.doi.org/10.1016/j.hrmr.2008.03.003

Spencer, M. B., & Tinsley, B. (2008). Identity as coping: Assessing youth's challenges and opportunities for success. *Prevention Researcher, 15,* 17–21.

Stout, W. (2001). Nonparametric item response theory: A maturing and applicable measurement modeling approach. *Applied Psychological Measurement, 25,* 300–306. http://dx.doi.org/10.1177/01466210122032109

Streiner, D. L. (2010). Measure for measure: New developments in measurement and item response theory. *The Canadian Journal of Psychiatry/La Revue Canadienne de Psychiatrie, 55,* 180–186.

Symonds, M. R. E., & Moussalli, A. (2011). A brief guide to model selection, multimodel inference, and model averaging in behavioral ecology using the Akaike's information criterion. *Behavioral Ecology and Sociobiology, 65,* 13–21. http://dx.doi.org/10.1007/s00265-010-1037-6

Tay, L., Ali, U. S., Drasgow, F., & Williams, B. (2011). Fitting dichotomous and polytomous data: Assessing the relative model-data fit if ideal point and dominance models. *Applied Psychological Measurement, 35,* 280–295. http://dx.doi.org/10.1177/0146621610390674

Thissen, D., Chen, W. H., & Bock, R. D. (2003). *MULTILOG* (Version 7.0) [Computer software]. Chicago, IL: Scientific Software International.

Thissen, D., & Wainer, H. (2001). Overview of test scoring. In D. Thissen & H. Wainer (Eds.), *Test scoring* (pp. 1–19). Mahwah, NJ: Erlbaum.

Thomas, M. L. (2011). The value of item response theory in clinical assessment: A review. *Assessment, 18,* 291–307. http://dx.doi.org/10.1177/1073191110374797

Vlachopoulos, S. P., Ntoumanis, N., & Smith, A. L. (2010). The Basic Psychological Needs in Exercise Scale: Translation and evidence for cross-cultural validity. *International Journal of Sport and Exercise Psychology, 8,* 394–412. http://dx.doi.org/10.1080/1612197X.2010.9671960

Weisz, J. R., Weiss, B., Suwanlert, S., & Chaiyasit, W. (2006). Culture and youth psychopathology: Testing the syndromal sensitivity model in Thai and American adolescents. *Journal of Consulting and Clinical Psychology, 74,* 1098–1107. http://dx.doi.org/10.1037/0022-006X.74.6.1098

Wright, M. O., Masten, A. S., & Narayan, A. J. (2013). Resilience processes in development: Four waves of research on positive adaptation in the context of adversity. In S. Goldstein & R. B. Brooks (Eds.), *Handbook of resilience in children* (2nd ed., pp. 17–37). New York, NY: Springer Science + Business Media.

Yang, Y., Sun, Y., Zhang, Y., Jiang, Y., Tang, J., Zhu, X., & Miao, D. (2013). Bifactor item response theory model of acute stress response. *PLoS ONE, 8,* e65291. http://dx.doi.org/10.1371/journal.pone.0065291

Zhang, J., & Stout, W. (1999). The theoretical DETECT index of dimensionality and its application to approximate simple structure. *Psychometrika, 64,* 213–249. http://dx.doi.org/10.1007/BF02294536

Received September 10, 2012
Revision received April 6, 2015
Accepted April 21, 2015 ■