CONTEXTUAL ANALYSIS OF MEASUREMENT BIAS IN THE CHILD BEHAVIOR CHECKLIST FOR AFRICAN AMERICAN CHILDREN

By

Victoria K. Ngo

Dissertation

Submitted to the Faculty of the

Graduate School of Vanderbilt University

in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

in

Psychology

August, 2006

Nashville, Tennessee

Approved:

Bahr Weiss, Ph.D.

Bruce Compas, Ph.D.

David Cole, Ph.D.

Tony Brown, Ph.D.

To my amazing family for always providing me with
love, encouragement, and support to
pursue my dreams

ACKNOWLEDGEMENTS

This work would not have been possible without the financial support of the National Institute of Mental Health and the American Psychological Association Minority Fellowship Program. I am especially indebted to Dr. Nambury Raju, whose unparalleled guidance and support, made this dissertation possible. Although, Nam is no longer with us, his devotion to his students and his contribution to the field will be remembered. I am also grateful to my advisor, Dr. Bahr Weiss, for cultivating my intellectual and professional development throughout the years. I have been fortunate to have an advisor, who has always extended himself above and beyond my expectations. The many long hours spent challenging me to sharpen my research skills, reviewing my drafts, and encouraging me to continue my pursuits, have taught me what great mentorship is all about. I would also like to thank my committee members, Drs. Bruce Compas, David Cole, and Tony Brown for their invaluable feedback and insight.

Many people in Nashville have made my graduate career worthwhile. All the wonderful people from Center for Psychotherapy Research and Policy team that have created an intellectually stimulating, supportive, and fun environment for me over the years. My dear friends, Agnes Leung, Stephanie Reich, Manuel Riemer, Eban Walters, Brian Barry, Pete Eamarond, and Roberto Morales have cheered me on and provided me with much love to get to the finish line.

Most importantly I have my family who I could never thank enough for everything they have given me. Jennifer, David, and Susan, you are the best siblings any one could ask for. Thank you for selflessly allowing me to pursue my dreams while you hold down the fort with mom and dad. Most importantly, I have my wonderful parents,

who have provided me with unending inspiration. Without their unconditional love and support, this accomplishment would not be possible.

TABLE OF CONTENTS

LIST OF TABLES

Page

LIST OF FIGURES

Page

CHAPTER I

INTRODUCTION

Over the past several decades, the demographic composition of the United States has been rapidly changing, with ethnic and racial minorities expected to account for over half of the population by 2050 (1996 Census Reports). According to the Children's Defense Fund (1991), minority children are not only one of the fastest growing populations, but due to the economic and social disparities that they experience, they are one of the most vulnerable. Ethnic minority children, in particular African-Americans, have significantly higher rates of poverty than their Euro-American counterparts. In 2000, 30.6% African-American children, 28.0% Hispanic children, 14.4% Asian and Pacific Island children, and 12.9% Euro-American children lived under the Federal poverty level. The rate of severe poverty for African-Americans is three times that of Euro-Americans (U.S. Census Bureau, 1999). Problems such as school dropout, poor academic performance, violence, and teenage pregnancy also disproportionately impact ethnic minority children, in large part due to their greater exposure to poverty and stress (Mcloyd & Steinberg, 1998). The changing U.S. demographics and these racial economic disparities highlight the need for increased research focus on minority mental health, particularly that of children.

However, research on ethnic minorities has failed to keep pace with their increasing prevalence in the population, as evidenced by the small percentage (2-6%) of publications utilizing ethnic minority (predominantly African-Americans) research samples through the 1980's (Graham, 1992; Jones, 1983; McLoyd & Randolph, 1986;

Ponterotto, 1998); even fewer studies have focused on conceptual issues pertaining to culture or context (McLoyd, 1999). Through the 1980's, in fact, there was a declining trend. For example, in a content analysis between 1970-1989 of race trends in six top APA journals considered to be representative of mainstream psychology, Graham (1992) found a declining representation (5.2% in 1970-1974 compared with 2% in 1985-1989) of African-Americans in empirical articles that either focused on African-Americans or conducted race analyses. He also found that 72% of these studies were race-comparative studies, but only 10% controlled or even reported the SES of their sample, thereby confounding race and SES. The rate of declining representation of African-Americans also was present in the child literature. In a content analysis on the Journal of Child Development between 1936 to1980, McLoyd and Randolph (1985) found that representation of African-Americans peaked in the early 1970's and declined there after. Consistent with Graham (1992), they concluded that the majority of research was race-comparative and confounded race with SES.

More recently, although research on ethnic minorities appears to be slowly increasing, the body of research focused on ethnic minorities remains limited (McLoyd & Steinberg, 1998). Further, the conclusions that can be drawn from the extant research base are limited by methodological and conceptual shortcomings inherent in conducting culturally-appropriate research on ethnic minorities. Some of these shortcomings include difficulties in obtaining sufficient sample sizes, a lack of consistency and clarity in the conceptualization of key terms such as race, ethnicity, culture, and a shortage of validated psychological assessment tools for various ethnic minority populations (Sue & Okazaki, 1995).

Starting in 2002, the Office of Behavioral and Social Sciences Research of the National Institutes of Health began to set national research priorities for examining racial disparities in health and behavioral health outcomes. Their strategic plan notes that although race has been found to relate to health and mental health outcomes, the magnitude of this relation and the mechanisms that underlie these effects remain unclear. Along with the Surgeon General, they called for more research to clarify the role of race and ethnicity in the development, expression, assessment, and treatment of psychopathology. This need also has been identified in the child arena, as the National Institute of Child Health and Human Development has instituted a requirement for inclusion of a representative sample of ethnic minority persons in federally funded research (Hall, 2001). The purpose of this requirement is to foster a more systematic examination of cultural and racial issues that have been neglected in psychological and developmental research, and to increase the generalization of research findings to more diverse populations.

However, although these initiatives have led to some increased representation of ethnic minorities in samples, the current literature remains limited. The extant empirical literature focused on ethnic minorities has mainly involved race-comparative studies, examining racial or ethnic differences in psychological outcomes or processes (Allen & Mitchell, 1998; Graham, 1992; Sue & Okazaki, 1995). One central problem with race-comparative research is that the validity of its findings rests upon the validity of the measurement instruments for *all* of the populations examined. Unfortunately, most assessment instruments used in psychological research have not been evaluated across diverse cultural groups. If the instruments' validity has not been determined for all

groups involved in the research, it is impossible to ascertain whether scores produced from the instruments are reflective of the same constructs across the groups. Since many of these assessment instruments are used clinically, failure to culturally validate the instruments also has likely led to invalid assessments, inappropriate treatment planning, and thus may contribute to poorer mental health outcome for ethnic minorities.

Multicultural researchers have warned for decades that assessment instruments may not be appropriate for use in diverse populations (Ben-Porath, 1990; Hall, Bansal, & Lopez, 1999; Okazaki & Sue, 1995; Triandis & Brislin, 1984). A common criticism of the use of 'objective' personality instruments for ethnic minority populations is that they assess constructs that may be culturally bound, reflect a method that may not accurately assess the construct of interest, and result in items that are potentially biased against ethnic minorities (Dana, 2000; van de Vijver & Hambleton, 1996). This potential measurement inequivalence in items and tests of child behavior assessments, such as the Child Behavior Checklist (Achenbach & Edelbrock, 1983; Achenbach, 1991) and the factors underlying this inequivalence is the focus of the present study.

The Importance of Measurement Equivalence

It is widely understood that if an instrument measures different constructs for the populations for which it is used, the observed score differences between the groups are not comparable; i.e., differences may not be due to the construct of interests but rather to some other unknown construct. For example, if one were interested in differences in aggression for African-Americans versus European-Americans, and one used a measure of aggression for which the items had different meanings for African-Americans as

opposed to the European-Americans, mean differences on the measure would not necessarily reflect differences in 'aggression' but rather some other factor related to ethnic membership. Although the influence of ethnic membership on the perception of aggression is in itself interesting, it confounds our interpretation of the level of differences in aggression. Failure of construct validity of this sort generally is known as measurement inequivalence or differential item functioning (DIF). Conversely, a test is considered culturally equivalent when test items have the same meaning across different cultural groups; i.e., the items reflect the same underlying construct for both groups and are similarly correlated with other related constructs.

Sheppard and Camilli (1994) have outlined external and internal approaches to studying measurement inequivalence or bias[1] in the context of construct validity. The former involves the use of an external criterion to evaluate bias, whereas the latter examines internal relations within the test to determine bias. The external method can be conceptualized as an investigation of predictive or concurrent validity, where instruments are evaluated for their relations to related constructs whose validity is already established *for all groups concerned*. When an unequal predictive relation between the groups of comparison is found, test bias can be inferred. However, this predictive accuracy is dependent upon an internal criterion of the test itself, in that the items of which the test is composed must exhibit equal relations to the latent trait for both groups. Additionally,

[1]Test 'bias' has been conceptualized in various ways, including mean differences, different interpretation of items, and differential statistical relation between the performance on a test and the underlying construct (Allen & Walsh, 2000). We follow the definition that test bias refers to systematic error in the measurement of some latent trait, whereas 'fairness' refers to the use of the test for a particular purpose (Reynolds, 1982; Sheppard, 1982). This clarification is important since the term *bias* can often connote prejudice and unfairness.

the external criterion must also be bias free in order for the use of a predictive model for appropriate examination. The internal approach of evaluating test bias, in contrast, assesses the equality of the magnitude of the relation of items to the latent construct across the two different populations. Although both approaches are necessary to provide a thorough examination of measurement equivalence, internal examination is a necessary first step to establish measurement validity.

Factor analytic models have often been used to investigate construct validity, via an internal examination of the extent to which items' relations to the latent factor(s) fit a hypothesized theoretical model. Factor analytic cross-cultural examinations of constructs use several techniques, including (a) comparison of the internal factor structures across groups, (b) comparison of item loadings, and (c) comparison of error variance, to detect measurement inequivalence. Significant differences found in these comparisons are thought to reflect a difference in the meaning of the construct across the groups of comparison (Ben-Porath, 1990; Floyd & Widaman, 1995).

However, several key assumptions of the factor analytic approach limit the precision with which construct validity can be assessed, particularly at the item level (Embretson & Hershberger, 1999). First, factor analysis provides information about measurement equivalence at a scale level by its comparison of factor structures. This sort of comparison does provide some information about item level properties, such as the degree to which the items load on the latent construct (similar to item discrimination in IRT). However, in factor analytic models, the magnitude of the latent construct needed to endorse an item (item threshold) is undifferentiated from the item loadings. Another limitation of factor analytic models is that their findings are sample dependent, in that

results are based on the assumption that a sample is representative and normally distributed which is often not the case in clinic populations of outpatient or hospitalized children. Furthermore, factor analysis is based on an additive linear model that assumes that an individual's observed score is the sum of the true score plus a single source of error across different items, assessed via a common standard error. This model does not allow for variability in the error across the different levels of traits for each item, nor does it provide information in regards to item properties such as guessing. Thus, although factor analytic procedures are often used to establish the validity of a test and examine measurement equivalence, this method is limited by its assumptions about the linear relation of the data to the underlying construct, greater sampling and distributional assumptions, and the limited information regarding item properties.

Lord (Lord & Novick, 1968) revolutionized measurement theory with his model-based measurement known as Item Response Theory or Latent Trait Theory. IRT uses logistic regression to characterize the relation of an individual's item response to the underlying trait level, theta (θ). This regression curve generally is referred to as the "Item Characteristic Curve" (ICC), as it describes the monotonically increasing relation between the respondent's level of a latent trait and the item performance. This S-shape curve traces the probability of the endorsement of an item as a function of the latent trait and the item parameters: (a) slope (or discrimination, represented by the parameter α, which represents that magnitude of the non-linear relation between the item and the latent construct), (b) location (or difficulty, represented by the parameter β, which represents the level of the latent trait necessary for an individual to have a 50% probability of endorsing the item), and (c) lower asymptote (or guessing, represented by $\boldsymbol{\gamma}$**)**.

IRT assesses measurement equivalence via evaluation of differential item functioning (DIF), which represents the extent to which individual items differ across different populations in regards to various aspects of their relation to the hypothesized underlying latent construct. In the IRT framework, DIF occurs when respondents with equal latent trait but from different subgroups (e.g. ethnic groups, gender groups) do not have an equal probability of endorsing an item (Hambleton & Swaminathan, 1985; Hambleton & Rogers, 1989) due to differences in item parameters. For example, differences in the β parameter indicate that endorsement of the item reflects a different level of the trait (e.g. see Figure 1a, wherein African-Americans show a lower threshold for the CBCL item "Disobedient at home" than Euro-Americans, which can be interpreted as indicating that endorsement of this item reflects less aggression, the factor or latent trait upon which this CBCL item loads, for African-Americans than it does for Euro-Americans). This is often also called uniform DIF.

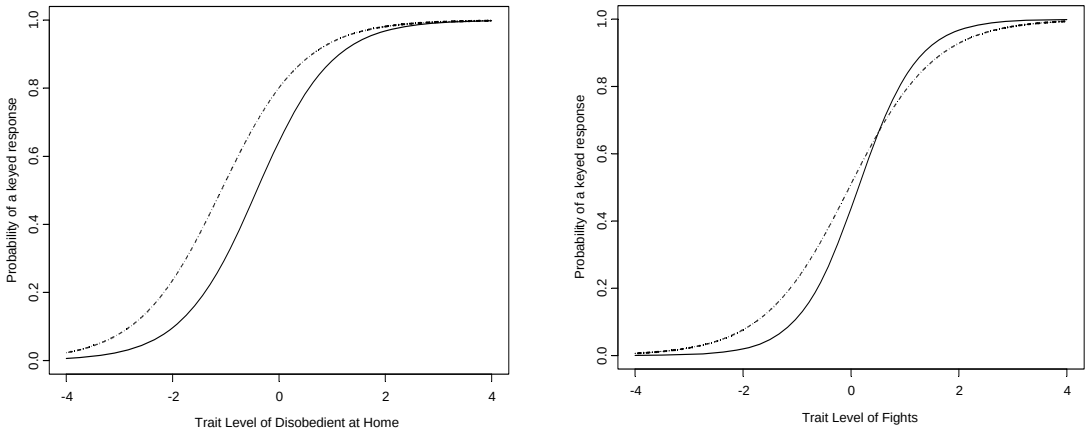


Figure 1. Uniform and Non-uniform DIF

In contrast, difference in the α parameter is called nonuniform DIF, because at different levels of the trait, the difficulty or threshold of the item is different. For example, in Figure 1b, African-Americans have a higher probability of endorsing the CBCL item "Fights" on the lower end of the continuum, but lower probability of endorsing the item at the high end of the continuum, as compared to Euro-American. In addition, the item is more discriminating for Euro-American than for African-Americans (i.e., it has a greater slope for Euro-Americans). When there is no differential functioning, the item response curves overlap, indicating that, given the same latent trait level (in this case, of aggression), both groups have a similar probability of endorsing the item. Key advantages of IRT include the disentanglement of item parameters and latent trait, assessment of variability in error distribution across the latent trait, and sample independence — all of which allows for greater precision and more information (Embretson, 2000) as compared with factor analysis. Consequently, the use of IRT for the assessment of internal validity and psychometric equivalence has come to be preferred over factor analysis (Embretson, 2000; Embretson & Reise, 2000; Sheppard & Camilli, 1994; Holland & Wainer, 1994; Allen & Walsh, 2000).

Despite this progress in measurement technology, there remains a paucity of research examining the cultural validity of psychological assessment tools (Dana, 2000; Okazaki & Sue, 1995). Much of the work investigating differential item functioning has focused on achievement and cognitive tests (Carlton & Marco, 1982; Embretson & Reise, 2000), with only a small number of studies focusing on the validity of psychological assessment for ethnic minority children.

Validity Studies on Child Psychopathology Measures for African-Americans

A review of the few existing validity studies of psychopathology assessments for African-American children suggests that there may be measurement inequivalence for African-American versus Euro-American samples. Assessment of the measurement equivalence for African-Americans on child psychopathology instruments assessing depression using the CES-D (Center for Epidemiologic Studies Depression Scale, Iwata et al., 2002), and delinquency using the Delinquency Scale (Piquero et al., 2002) have found DIF at both the item and scale levels. Iwata and colleagues (2002) examined the CES-D for DIF across three ethnic groups, Hispanics, Euro-Americans, and African-Americans. Their data showed that compared with non-Hispanic Euro-Americans with the same latent trait level, African-Americans under-endorsed items on the Depression scale but over-endorsed items on the Somatic Scale. However, there were certain limitations with the DIF methodology used in this study. First, a partial correlation method, wherein individual CES-D item and an exogenous variable for group membership were correlated, partialling out the variance shared with a matching variable, the NEG subscale of the CES-D. The authors assumed that the negative items of the scale would be less culturally loaded and therefore appropriate for use as the matching criterion. Thus, according to their analyses, DIF was identified when a unique association was found between an item and an exogenous variable for group membership, with differences in scores on the 'latent' score controlled. Given this DIF methodology, the distinction between discrimination and threshold could not be made. The assumption that the NEG subscale was culturally unbiased was untested, which also is a limitation of this study.

Piquero and colleagues (2002) used a more sophisticated approach involving Rasch models, which are a form of model within the IRT family, to examine the Delinquency Scale (Elliot et al., 1985). Using data from the first wave of the National Youth Survey, Piquero and colleagues (2002) found that several items exhibited DIF on the basis of racial or ethnic identification. Five out of twenty four items were found to show significant DIF (differences in threshold) in comparisons between African-Americans and Euro-Americans. These items assessed aggravated assault, prostitution, sexual intercourse, hitting parents, disorderly conduct, and panhandling. These findings suggest that further investigation is warranted regarding potential measurement inequivalence in assessments of psychopathology for African-American children.

Why Focus on the Child Behavior Checklist?

With hundreds of studies utilizing the Child Behavior Checklist (CBCL), this instrument is one of the most widely used child psychopathology tools. In the United States, the CBCL is viewed as the "gold standard" in the assessment of children and adolescent psychological functioning due to the breadth of literature on the psychometrics of the scale and the large supporting body of empirical research using the instrument to validate other child psychopathology or problem behavior scales.

The CBCL is one of the forms that make up the multi-informant approach to child assessment, now referred to as ASEBA (Achenbach & Rescorla, 2001). The ASEBA consists of parent (CBCL), teacher (Teacher Report Form, or TRF), and Youth Self-Report (YSR) forms, on which parents, teachers, and adolescents, respectively, rate the functioning of children. Since the first publication of the CBCL (Achenbach &

Edelbrock, 1983), the TRF (Achenbach & Edelbrock, 1986), and the YSR (Achenbach & Edelbrock, 1987), minor changes in item wording and composition were made to the forms. In pre-ASEBA versions of the CBCL (1983, 1986, 1991), age groups 4-5 were included. The ASEBA system now has a preschool form for age group 1.5 to 5 years old and another form for age group 6 to 18 years old (Achenbach & Rescorla, 2000). Norms have been developed for each age group across both genders. The three forms (CBCL, YSR, TRF) have maintained a hierarchical cross-informant (i.e., the same for ratings from parents, teachers, and adolescents) factor model across the two most recent revisions (i.e., in 1991 and 2001), in which eight first-order factors (i.e., dimensions or syndromes) load on two higher order factors, labeled Internalizing and Externalizing.

Development of the CBCL was based primarily on middle-class, Euro-Americans, yet its application has extended to ethnically diverse groups within the U.S., including African-Americans, Latinos, and Asian-Americans. The CBCL has also been used in research across at least 28 countries including Australia (Sawyer et al., 1989), Germany (Remschmidt and Walter, 1990), Israel (Margalit et al., 1989), mainland China (Li et al., 1989; Tseng et al., 1988), France (Frombonne, 1991), Netherlands (Berden et al., 1990; Verhulst et. al., 1990), Puerto Rico (Camino et al., 1990; Rubio-Stipec et al., 1990), Thailand (Weisz & Weiss, 2005), and Jamaica (Lambert, Rowan, Lyubansky, & Steinberg, 1998), and translated into 40 languages (Verhulst and Achenbach, 1995). Clinicians, teachers, parents, and researchers across the world have used the CBCL to obtain profiles of children's behaviors for the purposes of assessment, treatment planning, and theory testing.

In the US, the CBCL is commonly used to assess the psychological functioning of African-American children. However, few studies have specifically examined the validity of the CBCL for African-Americans (e.g., Latkovich, 1996; Lambert et al., 1998). Only one unpublished dissertation (Latkovich, 1996) has evaluated the construct validity of the CBCL for a clinic sample of African-Americans children, by comparing the factor structure of the CBCL with a Euro-American clinic sample. Latkovich (1996) found dissimilarities in the factor structures, with differences in item loadings for a factor termed "Schizoid-Anxious", which overlapped with the Depressed / Anxious factor typically used to describe the CBCL (Latkovich, 1996). This study also investigated concurrent and predictive validity by comparing the relation to changes in the Global Assessment Functioning scores and the Diagnostic and Statistical Manual of Mental Disorders (DSM-IIIR) diagnosis for African-Americans and Euro-American children. Latkovich (1996) concluded that the CBCL did not show predictive or concurrent validity for African-Americans. Consistent with this finding, validity of the CBCL for use with African-Americans was evaluated in a survey of clinic records for 1,605 African-American children (Lambert et al, 2002), who found poor predictive validity of the CBCL for African-Americans. In this study, Lambert and colleagues (2002) found that children's problems presented in clinic records did not overlap with problems on the CBCL. These researchers questioned the use of the CBCL for African-Americans and recommended further research on the validity of the CBCL for African-Americans.

Review of Validity Studies for the CBCL

Most factor analytic studies of the CBCL have not conducted separate analyses for different ethnic groups, and therefore results are mostly a function of the majority Euro-American sample. Similarly, in the development and norming of the CBCL, separate analyses have not been conducted by ethnic group. In the 1991 version of the CBCL, for instance, Achenbach (1991) used principal components and factor analysis to construct an eight factor cross-informant (self, parent, and teacher) model, using an approximately 80% Euro-American and 20% African-Americans sample. The most recent version of the CBCL (Achenbach & Recorla, 2001) used a multistage national probability sample to obtain a representative sample of U.S. children but also did not analyze separately by race.

Several Confirmatory Factor Analysis (CFA) studies on US clinic samples have found a 2-factor broadband model (Greenbaum & Dedrick, 1998) and an 8-factor narrowband model (Dedrick, Greenbaum, Friedman, Wetherington, & Knoff, 1997; Early, Gregoire, and McDonald, 2001). These eight syndromes have been labeled (a) Withdrawn / Depressed, (b) Anxious / Depressed, (c) Somatic Complaints, (d) Social Problems, (e) Thought Problems, (f) Attention Problems, (g) Delinquent Behavior, and (h) Aggressive Behavior. These factors load on two higher order factors, labeled Internalizing and Externalizing Problems. The first three factors load exclusively on Internalizing Problems, the last two factors load exclusively on Externalizing Problems, and the remaining three are labeled mixed, as they load approximately equally on both the Internalizing and Externalizing second-order factors.

However, in both the initial development analyses and subsequent CFA validation studies, because of the preponderance of Euro-American subjects (74-80% of the sample) results for the most part have been a function of the Euro-American sample, reflecting their majority status in the U.S. population. There have also been other U.S. studies that have supported Achenbach's 8-factor and 2-factor models, but the studies did not describe the non-Euro-American portion of the sample (e.g., Dedrick, Greenbaum, Friedman, Wetherington, & Knoff, 1997; Early, Gregoire, & McDonald, 2001).

Many of the cross-cultural factor analytic studies have compared the eight factor cross-informant model derived with the U.S sample (Achenbach & Edelbrock, 1983; Achenbach, 1991) to samples derived from different countries, including Norway (Novik, 1999), Holland and Australia (Heubeck, 2000), Netherlands (Achenbach, Verhulst, Baron, & Althaus, 1987; De Groot, Koot & Verhulst, 1994; Verhulst, Achenbach, Althaus, & Akkerhuis, 1988), and Thailand (Weisz, Weiss & Suwanlert, 2003). Results have been inconsistent, with researchers affiliated with the authors of the CBCL tending to conclude that the CBCL is comprised of essentially the same broad- and narrow-band factors regardless of the culture/country of the sample, (e.g. Achenbach, 1991; Albrecht, Veerman, Damen, & Kroes, 2001; DeGroot, Koot, & Verhulst, 1994), whereas other studies have interpreted their results to indicate a lack of adequate fit across various populations (e.g. Hartman et al., 1999; Hoge & Andrews, 1992; Lengua, Sadowski, Friedrich, & Fisher, 2001; Song, Singh, & Singer, 1994, Weisz and Weiss, 2005).

Most notable is Hartman et al.'s (1999) construct validation study across seven countries, including the United States, Israel, Netherlands, Norway, Turkey, Portugal, and Greece. Using complementary estimation procedures (maximum likelihood and

unweighted least squares estimation) for the confirmatory factor analysis, they tested Achenbach's cross-informant syndromes on 13,226 parent ratings and 8,893 teacher ratings for clinic, normal, and simulated samples (see Hartman et al., 1999 for more detail). Their study concluded that the 8-factor cross-informant model did not provide an adequate fit for data from the above countries. They also critiqued previous research by De Groot and colleagues (1994; 1996), which concluded cross-cultural equivalence for the Dutch sample, by pointing out that although the fit indices were equivalent, item composition of the factor model was not consistent with that of Achenbach's 8-factor model. The authors asserted that the use of a fit index to infer cultural equivalence, without examining the relation of the items to the latent construct, results in inappropriate conclusions (Hartmann et al., 1999).

A number of researchers have raised concerns regarding the utility of the CBCL for diverse populations, due to the lack of diversity of the research samples used to establish validity and norms (Drotar, Stein, & Perrin, 1995; Hartman et al., 1999; Lambert et al., Weisz and Weiss, 2005). Although the most recent efforts to broaden the factor derivation sample to include children from Australia, England, and Holland (Achenbach & Rescorla, 2001) and to aggregate across the three groups surveyed, the factor derivation sample continues to be primarily of European ancestry.

Further, a confirmatory factor analytic approach may not be appropriate for use with the CBCL, because of data characteristics such as skewness of the distribution and the ordinal scaling of the measure (Hartman et al., 1999; Weisz and Weiss, 2005). These researchers have raised concerns about the appropriateness of CFA results of CBCL

studies when the multivariate normality assumption is violated, calling into question the previous studies supporting cross-cultural validity of the 8-factor cross-informant model.

Although the CBCL has been supplanted by the ASEBA, research on the 1991 version remains important, for several reasons. First, the 1991 version was the basis for and continues to be used in several large longitudinal datasets, including Fast Track (CPPRG, 1992) and The Project on Human Development in Chicago Neighborhoods (PHDCN; Sampson et al., 1997), which continues to contribute to the literature on childhood psychopathology using the CBCL as the primary child outcome measure. Furthermore, changes to the CBCL have been minor, as the large majority of the items remain the same in the 1991 and ASEBA versions. Other than name changes from Withdrawn to Withdrawn / Depressed and from Delinquent to Rule Breaking Behavior, the names of both 1991 and 2001 syndromes and their higher order factors are identical. The 2001 syndromes primarily reflect the content of the 1991 first-order factors. For example, with the exception of the addition of one item (Item #5, "Enjoys little") to the 1991 Withdrawn syndrome, the items loading on the 2001 Withdrawn / Depressed syndrome are identical. Across all the 2001 syndromes, 74 of the 94 items overlapped with the 1991 cross-informant syndromes and 20 did not. Six items that were rarely endorsed in the 1991 version were replaced by new items in the 2001 version, including item #2. "Drinks alcohol without parents' approval;" #4. "Fails to finish things he/she starts;" #5. "There is very little he/she enjoys;" #28. "Breaks rules at home, school, or elsewhere;" #78. "Inattentive or easily distracted;" and #99. "Smokes, chews, or sniffs tobacco." The new items #4 and #78 assess the same as the items bearing these numbers on previous versions of the CBCL 6-18, (Achenbach, 1991, Achenbach &

Edelbrock, 1986). Due to changes to item #2 about alcohol, item #105 (which previously asked about both alcohol and substance abuse) no longer includes alcohol. Thus, IRT analyses on the 1991 version is also applicable to the ASEBA as well as other measures of child problem behaviors, as many of the items will likely be similar to other measures.

Only one study has used IRT to evaluate the CBCL (Lambert et al., 2003). This study employed both CFA and IRT to assess the 1991 syndrome scales of the youth self report version of the CBCL for Jamaican sample. They found that approximately 3/4 of the items provided little information for Jamaicans, with items failing to discriminate well between the levels of various psychopathology traits. Given that there are substantial cultural differences between Jamaicans and African-Americans (Lambert et al., 2003), it is not possible to generalize the results of this IRT analysis to American samples. However, findings from this study do raise concerns regarding the universal utility of the CBCL, which is often assumed from previous factor analytic studies.

Previous research that we have conducted, using an aggregate of three clinic samples, has found differential functioning for African-Americans on the CBCL, using a dichotomous two-parameter IRT analysis (Ngo & Weiss, 2003). Using a criterion that significant differential functioning be found across multiple DIF indices (signed area, unsigned area, and non-compensatory Differential Item Functioning-NCDIF; Ngo & Weiss, 2003), numerous items showed differential functioning. Differential functioning also was found for six CBCL scales, including the Aggressive Behavior, Social Problems, Attention Problems, Delinquent Behaviors, Thought Problems, and Somatic Complaints subscales. The Aggression scale showed the largest DIF whereas the Somatic

Complaints scale showed the smallest DIF. On average, the African- and Euro-Americans differed by .71 on the observed scale score for the Aggression scale, .44 on the Social Problems scale, .32 on the Hyperactivity and Attention Problems scale, .27 on the Delinquent Behavior scale, and .10 for the Thought Problems and Somatic Complaints scales. This suggests that when the Aggression scale, for example, is scored dichotomously (i.e., 0/1), the average total score difference is .71 for this 20-item scale when African-Americans are calibrated on Euro-American parameters versus their own parameters (i.e., when they are at the same latent trait level).

These findings may have significant implications for the use of the CBCL with African-American populations, because differential functioning at the item or test level indicate other dimensions may impact the responses pattern for this group other than the latent trait, which in turn implies differential validity (i.e., the endorsement of the items may not reflect the same underlying construct for African-Americans as for Euro-Americans). However, although this study (Ngo & Weiss, 2003) represented an important step in applying IRT framework to the CBCL for African-Americans, it is limited in several respects. First, it utilized a dichotomous IRT model, which although parsimonious, provides less information on the scale than a polytomous model and does not allow for the detection of differential functioning at each response option of the scale. Second, it did not examine the factors that may underlie the differential functioning; given that in this sample race and SES were confounded, it was unclear whether the bias was due to race, or to socioeconomic status, or to some other factor correlated with but not a fundamental part of race.

Contextual Perspective of Measurement Inequivalence

Findings of differential response patterns indicate that there are between-group differences in how individuals respond to items. However, differential functioning alone does not provide information regarding the correlates or processes that underlie these differential responses. Race is itself not a process, but rather a marker for a variety of social and cultural experiences (e.g., economic marginalization, racial discrimination, social disadvantage, racial socialization, cultural values) that are believed to have important effects on a wide variety of factors (Betacourt & Lopez, 1993; Jones, 1990; Williams & Williams-Morris, 2000; Yee, Fairchild, Weizmann, & Wyatt, 1993). Although there are several general theoretical explanations for race differences (e.g., genetic factors; cultural effects; socio-political differences), these theories have seldom served as a foundation for empirical investigation of race effects in health and mental health research. Thus, although empirical studies often collect and analyze race data, these studies seldom consider the processes through which race may influence other variables.

For example, in a review of the use of race as a variable across 30 years of health services research, Williams (1994) found that despite the frequent use of race or ethnicity to stratify or adjust results, the terms seldom were defined, nor was the inclusion of these variables theoretically justified or explained. He recommended that researchers provide a more explicit conceptualization of the term, and that they identify the social and cultural factors presumed to underlie the effects of race. Ethnic minority mental health researchers have echoed the need for a theoretical framework for examining race effects and have often raised the same critique of psychological research. Betacourt and Lopez

(1993), for example, suggested "that when behavioral variations are studied in relation to race, the so-called racial variable under study should be defined, measured, and the proposed relationship tested. The role of specific cultural and social variables should be clearly separated from that of biological and other variables…the important point is that the research be on the relevant variables and not on racial groups alone" (pg. 631).

When considering measurement inequivalence across different races, it likewise is important to explicate a theoretical model for potential racial differences. In this study, we employed a socio-cultural perspective on race effects, assuming that race effects are most strongly influenced via ethnic membership, which in turn reflects differences in socio-cultural experiences for the individual, for the individual's community, and for the ethnic group historically. Culture is conceptualized as incorporating ethnic heritage and historical adaptation to economic and sociopolitical experience, with one's cultural context, which includes societal, neighborhood and family environment, impacting on and reflecting the values and expectations about behaviors in the home and in the community. These values and beliefs may influence what is considered normative versus pathological, both in terms of children's behavior as well as adult behavior that influences the behavior.

Specific Contextual Factors Underlying Race Differences

It generally is assumed that a number cultural factors underlie racial differences on psychological constructs (e.g., Betancourt & Lopez, 1993), including such factors as ethnic identity (Phinney,1992), experience with discrimination (Simons et al., 2002), parenting values and practices, cultural values about communication and expression of

emotions. These factors are likely to not only impact directly on child behavior but also on parent's evaluation and response to their children's behavior. People from different backgrounds or contexts may well have different beliefs in regards to what constitutes normative versus non-normative behaviors, as well as differences in their threshold of distress for different child behaviors. Their contextual or cultural background thus may impact how and whether parents view their children's behaviors as problematic. For example, parents who value hierarchy and teach their children to not question their authority will perceive their child as abnormal or problematic if their child asks too many questions about rules, particularly if the parents view this behavior as oppositional and disrespectful. On the other hand, a parent that values autonomy and critical thinking may expect their child to examine and question the world, and therefore will encourage this behavior in their children.

In the present study, we focused on three factors: Neighborhood context (Levanthal & Brooks-Gunn, 2000), parenting (McLoyd et al., 2000; Deater-Deckard & Dodge, 1997), and SES (McLoyd, 2002), because as discussed below these three factors have consistently been found to be related to both externalizing behavior problems and race, and appear to be central and proximal factors in the processes connecting psychopathology—particularly externalizing problems—and race.

In the United States, race and SES are highly correlated, with Euro-Americans typically occupying higher socio-economic status than ethnic minorities, particularly African-Americans. Historical trends from the 1980s to 1998 show that Euro-Americans have the highest employment rates, lowest percentage of unemployed, and lowest percentage of children living in poverty, whereas African-Americans have the lowest

percentage employed, highest unemployment, and highest percentage of children living in poverty (White & Rogers, 2000). These economic disparities are longstanding and pervasive. According to U.S. Census Reports in 1990, 7% of Euro-Americans fell below poverty level compared whereas 26% of African-Americans were below the poverty line. This trend generally continued in 2002, with 10% of respondents who identified themselves as Euro-American below poverty, whereas 24% African-Americans reported living below the federal poverty threshold.

Further, African-Americans are not only more likely to experience poverty than any other group, but are also more likely to remain in poverty longer (Duncan & Rodgers, 1988; McLeod & Shanahan, 1997; McLeod & Nonnemaker, 2000). According to findings from the Panel Survey of Income Dynamics, 20% of non-African-Americans spent one to four years in poverty and 6% spent more than five years in poverty. In contrast, 32% of African-Americans spent one to four years in poverty, and 43% spent more than five years in poverty. Of the African-American children living below poverty level, only 21% were later able to break above the poverty threshold (White & Rogers, 2000). A similar disparity was found in the National Longitudinal Survey of Youth (NLSY) sample, with African-American and Hispanic children likely to spend significantly more time in persistent poverty (McLeod & Shanahan, 1997). The race difference is even more striking when considering the other end of the continuum, with only 21% of African-American children having never lived in poverty, in contrast to 54% of non-Hispanic Euro-American children (McLeod & Shanahan, 1997).

In addition to the poverty rates experienced by African-Americans, the nature of socio-economic disadvantage also differs for African-Americans and Euro-Americans.

McLeod and Shanahan (1993) have argued that the psychological meaning of severe economic deprivation, such as that experienced by impoverished African-Americans, is different from that experienced by Euro-Americans. African-Americans are more likely to live in isolated urban ghettos with a higher concentration of poverty (Farley, 1987; Wilson, 1987), and are less likely than poor Euro-Americans to have friends and family with financial resources (McLoyd, 1990). Their experience of economic disadvantage is thus more pervasive with fewer prospects for economic improvement. Further, family structure (single mother status) and female head-of-household mostly accounted for low SES in African-American communities. African-American families are on average larger than Euro-American families (65% versus 54% of families with three or more members) but also are more likely to be headed by a single parent (i.e., 38% of African-American children were living in 2-parent household compared with 69% of all other children in the U.S.). Furthermore, African-American one-parent households are also more likely to be headed by females (92% versus 69%) than other ethnic groups (U.S. Census Bureau, 2001).

Many of the effects of SES occur at the neighborhood level; i.e., numerous studies have found that a child's socioeconomic background is correlated with neighborhood structural dimensions, such as concentration of poverty, residential instability, availability of resources, cohesion among neighbors, exposure to violence, and other environmental threats such as drug sales. Neighborhoods where African-Americans reside are more likely to have crowded housing, poor quality schools, and the presence of prostitution, homelessness, violence, drugs, and firearms (Brooks-Gunn, Duncan, & Aber, 1997; Duncan, 1991; McLoyd, 1998; Sampson, Morenoff, & Earls, 1997; Tolan, Guerra, &

Montaini-Klovdahl, 1997). Overall, African-American status is significantly associated with children's exposure to violence, both in terms of the total amount and the severity of violence (Oya, 2000).

African-American ethnicity is also significantly associated with residential instability (r=.32), structural disadvantage (r=.69), and less concentrated affluence (r=-.39; Beyers, Bates, Pettit, & Dodge, 2003). It is not surprising, given these objective indicators of disadvantage, that African-Americans perceive their neighborhoods to be lacking in resources and safety, with greater exposure to gang violence than Euro-Americans (Duncan, Strycker, Duncan, & Okut, 2002). When research studies have investigated the effects of exposure to violence in inner-cities, samples are predominantly African-American (e.g., DuRant et al., 1994; Fitzpatrick & Boldizar, 1993; Gorman-Smith, & Tolan, 1998; Kliewer et al., 1998; Osofsky et al., 1993; Richter & Martinez, 1993; Schwab-Stone et al., 1995) and Latino-American (Attar, Guerra, & Tolan, 1994), which highlights the correlation between negative neighborhood characteristics and ethnic minority status.

Some evidence suggest that as a response to this neighborhood context, as a group parents of African-American children tend to use different parenting behaviors than Euro-American parents, with African-American parents using more authoritarian, high controlling, harsh, physical discipline strategies than Euro-Americans (Deater-Deckard et al, 1997; Bradley et al., 2001; Garcia Coll, 1990; Gils-Sims, Straus, & Sugarman, 1995; Hurd, Moore, & Rogers, 1995; Krishnakumar, Buehler, & Barber, 2003; McLeod & Nonnemaker, 2000; Shumow, Vandell, & Posner, 1998; Steinberg, et al., 1992; Whiteside-Mansell et al., 2003). It has been suggested that these differences represent

African-Americans' adaptation to their group experiences of slavery, discrimination, and historic economic marginalization, ultimately resulting in differences in cultural norms (McAdoo, 2002; McLoyd, 1999; Steinberg, 1991). This long-term history of African-Americans' exposure to racism, poverty, and violence in their communities may have resulted in parents' perceiving a need for strict obedience to protect their children from harsh societal conditions (McLoyd, 1999). African-American families' cultural norms place greater value on respect for hierarchy, collectivism and cooperation in the family, and in a higher reliance on extended family members and community members (Gaines, Marelich, Bledsoe, & Steers, 1997; Hill, 1995; Ogbu, 1981). These group differences in SES, neighborhood context, and parenting values are in turn believed to impact upon parents' tolerance and expectations for normal and abnormal behavior in their children. Thus, cultural factors may explain for why parents' of African-American children may respond differently to psychopathology assessment instruments.

It is important to note, however, that a number of authors have raised the issue that there often is more variability within cultural groups than between cultural groups (Betacourt & Lopez, 1993). Although it clearly is important to recognize that culture is heterogeneous even within a particular cultural group, differential shared group experiences are nonetheless relevant to group-level outcomes. Thus, although individuals differ in their reactions, the legacy of slavery and the history of discrimination and racism shared by the African-American community as a whole play an important role in shaping this group's experience in American society. Adaptations to this marginalized economic and political status, such as high spirituality, collectivism, and extended family

orientation, are also a part of the cultural experience of African-American communities (Chatters, Taylor, Lincoln, & Schroepfer, 2002; Hill, 1995), as a group.

Purpose of Present Study

The present study extended upon our previous research in several ways. First, we used IRT polytomous scale techniques to analyze the CBCL, which provides greater information by (a) mapping the full response range of the polytomous CBCL (e.g., it is possible that African- and Euro-Americans' responses do not differ for the 0/1 response but do differ for the 1/2 response), as well as (b) more accurately assessed the overall IRT characteristics of the CBCL, given that the CBCL is polytomous. Second, the study examined the relation between (a) the magnitude of the measurement inequivalence for psychopathology and (b) various contextual factors that may underlie race effects, which provides information regarding possible mediators of race effects. An understanding of how poverty, neighborhood factors, and parenting differentially impact parental definitions of problem behavior across racial groups will provide insight into how we can better design prevention and intervention programs that take into consideration context and culture. We predict (a) that there will be significant differential functioning on the CBCL for African-Americans compared with European-Americans at both the item and scale levels, and (b) that this difference in response patterns will be related to socio-cultural differences between these two groups.

Findings of item and test differential functioning on the CBCL would have significant implications for the field of child psychopathology. First, differential functioning at the item or test level would indicate that other dimensions impact the

response pattern for African-American children, which would imply deferential validity; i.e., the endorsement of items may not indicate the same underlying construct for African-Americans and Euro-Americans. This would suggest that these constructs are culturally bound, that is, the same behaviors (listed in the items) across two groups may have a different meaning for each group. Such information is essential for drawing meaningful conclusions from past and future research using the CBCL (and other child psychopathology assessment instruments as they likely have similar items) with African-American children. At a practical level, findings of DIF and DTF will also help identify items or scales in need of modification so as to increase the validity of the measurement of emotional and behavioral problems among African-American children. Finally, an examination of the sources of bias that may influence a person's response to the test items will represent the first study to move beyond race differences to examining the role of cultural context in measurement inequivalence.

CHAPTER II

METHODS

Overview

This is a secondary analysis of data obtained from the Fast Track Study (Conduct Problem Prevention Research Group, 1992). The key objectives of this study were: (a) to examine the psychometric properties for the narrowband CBCL factors (i.e., syndrome scales) using IRT methods; (b) conduct differential item functioning analyses across African- and Euro-American groups for those scales that showed adequate IRT psychometric properties; and (c) relate contextual variables to differential item functioning to suggest factors that may underlie race differences.

Sample

Data for the present investigation came from the Fast Track study, a national longitudinal multi-site youth violence prevention program for high-risk children. Data have been collected for 3 cohorts over a 10-year period and include a diverse and representative sample of high risk and normative children from Seattle, Nashville, Durham, and central Pennsylvania. In the present study, African-American and Euro-American participants were selected from pretreatment data, which includes 6 year olds assessed during the 1st grade between years 1991-1993. We utilized cross-sectional data of CBCLs obtained from the Fast Track pre-treatment phase, including 579 African-American and 572 Euro-American across the three cohorts. One individual from the African-American sample was deleted due to incomplete CBCL form (with 70 items

missing). Therefore, only 578 participants from African-American group were used in the IRT analysis. The participants were obtained from three risk groups: 1) normative (A-A=165, E-A=198), 2) low-risk (A-A=128, E-A=158), and 3) high risk (A-A=214, E-A=217). The criteria for these risks groups can be found in the procedures section below. Gender breakdown are as follows: A-A (Female=198, Male=380) and E-A (Female=226, Male=346).

Assessment Domains and Measures

Child Emotional and Behavioral Problems

The Child Behavior Checklist (CBCL; Achenbach, 1991) was used to assess children's emotional and behavioral problems. It is a parent-report checklist of 118 behavioral and emotional problems (e.g., "cruel to animals," "sad, unhappy, or depressed"). For each item, parents report whether their child has the problem by circling 0 ("Not True"), 1 ("Somewhat or Sometimes True"), or 2 ("Very True or Often True"). The CBCL produces two broadband or higher order scales (Internalizing and Externalizing problems) as well as eight narrowband scales. In the present study, the first-order factors, which are the eight narrow-band cross-informant subscales (Anxious / Depressed, Somatic Complaints, Thought Problems, Withdrawn, Social Problems, Hyperactivity and Attention Problems, Delinquent Behavior, and Aggressive Behavior) were used (Achenbach, 1991). The items for the CBCL 1991 version are listed in the Appendix.

Neighborhood Characteristics

Relative to non-minorities, ethnic minority families are more likely to live in high-crime neighborhoods with increased exposure to violence and restricted access to resources (Garcia Coll et al., 1996; Levanthal & Brooks-Gunn, 2000; Simons et al., 2002). Residence in such neighborhoods is associated with increased risk for depression, anxiety disorders, aggression, and delinquency (Brody et al., 2003, Gorman-Smith & Tolan, 1998; McLoyd, 1990; Richters & Martinez, 1993), all of which impact on parents' beliefs and perceptions of appropriate behavior in their children. Therefore, neighborhood environment is an important contextual factor relevant to child psychopathology, and in the present study was assessed using the 16-item Neighborhood Questionnaire (Greenberg, M. & Lengua, 1995), which assesses parent's perception of sociability in the neighborhood, neighborhood's stability, quality of public services (e.g., police, schools), neighborhood safety and violent crime, etc. The questionnaire includes responses on dichotomous, three-point, four-point, five-point, and six-point scales; one item has nine response choices. Items ask about satisfaction level, quality, duration, quantity, frequency, and level of involvement. This measure produces 3 subscales: Neighborhood Safety, Neighborhood Social Involvement, and Public Services.

Neighborhood Disadvantage

U.S. Census data for 1990 was used to provide information regarding neighborhood disadvantage as defined by the percentage of individuals within the census tract: with income below the poverty level, public assistance, with female-headed households, percentage unemployed adults, African-American residents, and stability (no moves within last 5 years or since 1985). This information was gathered using the

address information of participants, which was linked to the census tract (or block numbering area, where applicable) of residence for all of the participants for whom it was possible to verify a match. The geocoded files were linked to characteristics of the census tract or BNA, using tabular information from the 1990 Census Summary Tape File 3.

Parenting Behaviors

Cultural differences in how African-American parents discipline and interact with their children, reflecting differential parenting styles and beliefs (Harwood, et al., 1996; Dodge, Bates, & Pettit, 1990), may have a significant impact on what parents perceive to be appropriate / inappropriate behavior in their children (McLoyd, 1990). The Parenting Practices Inventory is a 17-item measure developed for Fast Track to assess the parent's permissiveness of their discipline, the effectiveness of their discipline and the consistency of their discipline efforts. The items are coded on a 4-point scale describing specific frequency ratings: "never (1)", "almost never (2)", "sometimes (3)", "often (4)." This measure produces three sub-scales. The first, *Effectiveness*, assesses the parents' perceptions of the effectiveness of their discipline. The second, *Consistency*, assesses parents' perceptions of the consistency of their discipline. The third scale, *Punitiveness*, measures parents' perceptions of the punitiveness of their discipline.

Family Expressiveness

The Family Expressiveness Questionnaire is a 13-item revision of the original 40-item questionnaire developed by Halberstadt (1986). Halberstadt's questionnaire contains four subscales: positive-dominant (PD), positive-nondominant (PS), nonpositive-dominant (ND), and nonpositive-nondominant (NS). Greenberg et al. (1995)

revised the scale to include ten of Halberstadt's original items and three new items. Analysis of responses on the 13-item questionnaire supported two subscales – Positive Expression and Negative Expression (Greenberg, 1995).

Each item asks how often a particular scenario of emotional expression occurs in the respondent's family. Six of the items describe negative scenarios, e.g., "How often does someone in your family try to cheer up another family member who is sad?" and seven items express positive scenarios, e.g., "How often does someone in your family praise someone in the family for good work?" Items loading on the Negative Expression subscale reflect of openness and responsiveness to negative content (e.g., conflicts, disappointment, etc.) whereas items on the Positive Expression subscale indicate expressivity and responsiveness towards positive events. Responses are coded on a four-point scale: "Never happens (0)," "Sometimes happens (1)," "Happens a lot (2)," and "Happens all the time (3)."

Parenting Values

The Values Questionnaire is a 35-item measure that explores parents' values about their children's peer relations. In year 1, cohort 1 was given the whole measure and cohort 2 was given items #1-#10. In year 2, cohort 1 was given items #1-#10. This measure was never administered to cohort 3.

Each item describes a preference that parents may have about children's social behavior. Examples of statements are: "It is important for you to know who your child's friends are," and "If other children won't play with your child, he/she should stand up for himself/herself by showing who's boss." Responses are coded on a five-point scale and include "Strongly Disagree (0)," "Disagree (1)," "Neutral-Unsure (2)," "Agree (3)," and

"Strongly Agree (4)." Two sub-scales are generated from this measure. The Socialization subscale is the mean of items #6-#10, focusing on the child's social abilities and how important it is to the parents that they be involved in the child's social life. The Physical Defense subscale is created by taking the mean of items #1-#5. These items assess parents' desire that their children be able and willing to defend themselves physically.

Family Demographics

The Family Information Form (CPPRG, 1992) was used for deriving demographic information, information concerning family structure, and socioeconomic status. SES scores were obtained from The Hollingshead (1979) Four Factor Index of Social Status, which uses a weighted average of the education and occupational scale values of the participant's parents. Two variables were used in the study, including a socioeconomoic status continuous code and a family occupational code.

The Socioeconomic Status Continuous Code was created using the scoring formula derived by Hollingshead (1975). The score is "calculated by multiplying the scale value for an occupation by a weight of five and the scale value for education by a weight of three" (Hollingshead, 1975). These scores are then added together. The score may then be divided by two if both parents work.

The Family Occupation Code is derived from questions answered earlier in the survey. The score is equal to the higher of the occupation codes (female type of job or male type of job) of the head(s) of household. For two parent families where the mother's job is rated as a more highly ranking occupation on the Hollingshead (1975) scale, where there is a single mother, or where there is information for a female head of household but

missing data indicating a male head of household, this value is equal to the answer to question regarding the female's type of job. If the opposite conditions are true (2 parent household where dad's occupation is more highly rated, a single male headed household, or when there is information for a male head of household but missing data indicating a female head of household, this value is equal to the male's type of job.

Procedures

The participants of Fast Track study were identified through a multi-stage screening process. High-risk schools first were identified in four different areas of the United States, using crime records, poverty statistics, and high school dropout rates. The schools were matched on size, ethnic composition, achievement scores, and percentage of free lunch recipients. Half of the schools were randomly assigned to receive intervention services and the other half was control schools. In the spring of 1991, 1992, and 1993, teachers rated the behavior problems of each of the kindergarten children in the 55 participating elementary schools, using a screening instrument measuring aggressive and oppositional behavior. Children who scored in the top ten percent of the combined teacher and parent screen were considered to be the high-risk target group, and were invited to participate in the Fast Track longitudinal study. A randomly selected sample of children was chosen from the control schools to serve as a non-high risk normative sample. Within each of the four sites, children were stratified to represent the population according to sex, race, and decile score on the teacher screen measure, and then chosen randomly to participate in the study (CPPRG, 1992).

Data Analysis

Overview. There were six steps to data analysis: (a) Classical Test Statistics, (b) Assessment of IRT assumptions; (c) IRT parameter estimation; (d) Assessment of model fit, (e) DFIT analyses; and (f) assessment of relation between contextual variables and DIF.

Classical Test Statistics

Internal Consistency. An important attribute of the measurement of a construct is that the measure's items are internally consistent. A popular method of the internal consistency approach to reliability estimation was developed by Cronbach (1951) and often is referred to as Cronbach's alpha. Cronbach's alpha is an estimate of the proportion of the scales variance due to the common factors among the items. SAS Proc Corr procedure was used to compute Cronbach's alpha as an estimate of reliability of these items for each of the CBCL subscales.

IRT Model Assumptions

In order to be appropriate for IRT analysis, scales must meet unidimensionality and local independence assumptions (Hambleton et al., 1991). Unidimensionality refers to the assumption that item covariation arises predominantly from a single underlying dimension or common factor. Local independence refers to the assumption that all dependency among item responses is due to this common factor; i.e., for each factor, when covariance with this common factor is controlled, responses across items are independent of one another. If a single factor underlies a scale, both the unidimensionality and local independence assumptions are met (McDonald, 1999). However, it has been noted that this assumption cannot be met fully because there

inevitably will remain some level of shared variance related to readability of items, participant's reading level, etc. For the applied purpose of IRT analysis, unidimensionality assumption is met when (a) scree plot exhibits clear first dominant factor, (b) the eigenvalue of first factor is significantly larger than remaining factors, and (c) first factor accounts for over 20% of the total variance (Reckase, 1979; Hambleton et al., 1991).

Exploratory factor analyses were conducted to determine the dimensionality of each of the eight CBCL narrow-band scales. In these analyses, we employed principle factor analysis using squared multiple correlations on the matrix diagonal. Because CBCL item scaling is not interval level, we first attempted to use polychoric correlations; however, for several scales (e.g., depression, thought problems) models did not converge and the use of polychorics was not possible. Polychorics typically require N > 1,000 to produce stable estimates, and also have an assumption of multivariate normality. Given that our sample was less than 600 for each group, we were unable to use polychorics correlations. Consequently, Pearson correlations were used for the factor analyses. Following recommendation by Gorsuch (1983), a scree plot of the eigenvalues was used to determine the dimensionality of the scale. Following the above logic, exploratory factor analysis and scree plots were used to test the unidimensionality and local independence of each scale.

Samejima Graded Response Model

IRT is based on nonlinear regression that relates the latent trait level (in the present case, "Aggression", "Delinquency", etc.) to the probability of a keyed response (i.e., the probability of endorsing a specified CBCL item). The regression curve,

generally referred to as the "Item Response Function" (IRF) or "Item Characteristic Curve" (ICC), describes the relation between the level of the latent trait and the probability of a specified response, in this case, on the CBCL items.

Samejima (1969, 1972) has recommended that when there are more than two response categories, the response type be referred to as polytomous or graded responses. Having a model that incorporates multiple response categories is recommended for analyzing likert scale data, such as the CBCL. Because the data are polytomous and scaling is ordinal, we employed Samejima's Graded Response Model (GRM; Samejima, 1969) to estimate item response parameters for the Child Behavior Checklist (Achenbach & Edelbrock, 1981) scales. This model allows for the estimation of the probability of endorsing each of the response options (0- not true, 1-sometimes true, 2-almost always true). The extent of pathology or distress is reflected by increasing value of the categories.

According to the SGR model, the probability of selecting option k on item #i is

$$P(v_i = k|\theta = t) = 1 / 1 + \exp[-1.7_{\alpha i}(t-\beta_{jk})] - 1 / 1 + \exp[-1.7_{\alpha i}(t-\beta_{jk}+1) \quad (1)$$

where:

v denotes the person's response to the polytomously scored item #i

k is the particular option selected by the respondent (*k* = 1 to s, where "*s*" refers to the number of options for that item)

α is the item discrimination parameter, assumed to be the same for each option within a particular item

β is the threshold parameter that varies from option to option

θ represents the value of the latent trait (e.g., aggression)

$p(\theta)$ represents the probability of a positive response

In the polytomous case, the Item Response Function is replaced with Boundary Response Function (BRF), which can be defined by up to three parameters: threshold location (or difficulty, represented by the parameter β), slope (or discrimination, represented by the parameter α) and lower asymptote (or guessing, represented by the parameter γ). For psychopathology measurement, the lower asymptote / guessing parameter is not appropriate since there is no "correct" answer that can be obtained by guessing, therefore only the discrimination and threshold are used to define the BRF for psychopathology data. With psychopathology data, the y-axis represents the probability of endorsing a particular option, rather than the probability of endorsing the correct answer, as in dichotomous IRT of achievement measures.

In polytomous IRT Models, the discrimination parameter is the slope of the curve and assumed to be equal across options. Items with high discrimination parameters, all else being equal, have relatively more distinct BRFs (i.e., are more peaked) compared to items with smaller discrimination parameters. Because of this, items with high discriminations are better at differentiating among individuals of varying thetas (theta equaling the level of the latent trait for an individual or group) compared to items with low discriminations, thus providing greater information. If the slope is flat, on the other hand, then the item is less efficient at discriminating between individuals with different latent trait levels.

Baker (2001) provides general guidelines for interpreting the discrimination parameter for logistic models:

α value	Interpretation
.01-.34	Very low
.35-.64	Low
.65-1.34	Moderate
1.35-1.69	High
>1.70	Very high

In contrast to the dichotomous data wherein there is a single discrimination parameter for each item, the number of threshold parameters is equal to one less than the number of options for polytomous data. Therefore, a three-option item will have two between-option threshold response parameters. These between-option threshold parameters reflect the overall "difficulty" of items and correspond to the level of the latent trait (of psychopathology) at which there is a 50 percent probability of an informant endorsing the specific option (or item if model is dichotomous). These added features of the GRM result in a more precise model than the dichotomous model used in our previous analyses.

Boundary Response Functions

Cohen et al. (1993) define graded response items as those items that have m_j, ordered response categories (e.g., for three categories; m, = 3) where a rater is only able to choose one category in to a single item. To handle this type of data, Samejima (1969) developed a probability function called a Boundary Response Function (BRF). The BRFs are consistent in their *shape* as they are cumulative probability functions that are characterized by two parameters (i.e., a discrimination parameter α_j, and a location or threshold parameter β_{jk}). That is, for each item, the number of functions is one less than

the number of categories. For an item with m_j response categories, there are $m_j - 1$ cumulative dichotomies. Therefore, a series of dichotomies must be calculated for each item.

The first cumulative dichotomy in the set is whether a rater gives a rating in any category other than the first category. For a three category item ($m_j = 3$), the rating is scored as a 0 if a ratee has a rating in the first category and as a 1 if a ratee has a rating in the second and third category. The second dichotomy is for a rating greater than the first two categories. Again for a three category item ($m_j = 3$), the rating is scored as a 0 if a ratee has a rating in the first or second category and as a 1 if a ratee has a rating in the third category. Note that there is no need to calculate the last cumulative dichotomy (i.e., $m_j = 3$) because the item would be scored as a 0 when the rater marks any category and that probability will always be 0.

After the set of dichotomies have been created, a BRF is calculated for each dichotomy using the following probability function:

$$P_{jk}(\text{choosing above a response category} > k|\ \theta) = P_{jk}(\theta) = \frac{e^{\alpha j\,(\theta-\beta jk)}}{1 + e^{\alpha j\,(\theta-\beta jk)}} \qquad \textbf{(2)}$$

where $P_{jk}(\theta)$ is the probability that a randomly chosen ratee with ability θ will receive a rating score to item j with a rating greater than the k response category, β_{jk} is the boundary parameter between category k and k - 1, and α_j is the item discrimination parameter (see Cohen et al., 1993 for more details). This results in a set of monotonically increasing curves for each item. It is important to note that while α_j may *vary* between items, within Samejima's (1969) model, an additional assumption made for the graded response models is that the α's across the response categories within an item are equal

($\alpha_{j1} = \alpha_{j2} = \alpha_{j3} = \alpha_{jk}$) (Maurer et al., 1998; Flanagan, 1997; and Collins et al., 1997).

This means that the reasoning process of the individual rater is homogeneous or constant throughout the set of responses for an item. This results in all BRFs having equal slopes for each category in an item which ensures no crossing of the curves (Flowers, Osbima, & Raju, 1995). Figure 1 is an example of the BRFs for the CBC item #3 "Argues a lot" (Aggression Scale) with 3 response categories. In this example, the BRFs for both groups are plotted (Black representing the E-A and red dotted line representing A-A). For the African-American group, the $\alpha = 1.178$, $\beta_1 = -1.53$, $\beta_2 = 1.275$ and for the European-American group ($\alpha = 1.427$, $\beta_1 = -2.673$, $\beta_2 = 0.574$). In this example, African-Americans require greater levels of the latent trait (aggression) in order to endorse the item. Specifically, African-Americans require -1.50 standard deviations below the mean level of aggression to endorse option 1 or above (BRF1) and 1.75 standard deviations above the mean level of aggression to endorse option 2 (BRF2). In contrast, European-Americans require -3.0 standard deviations below the mean to endorse option 1 or higher and .75 standard deviations above the mean to endorse option 2. As can be seen in the BRFs, the difference between the two groups is greatest in the first response dichotomy (i.e., endorsement of option 1 or 2).

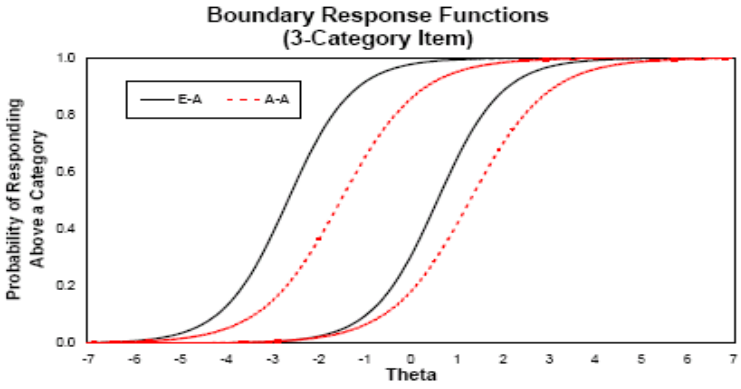


Figure 2. Example of Boundary Response Functions

IRT Parameter Estimations

Parscale 4.2 (Muraki & Bock, 2003) was used to calibrate item and person parameters. Parameter estimates were obtained separately for each scale and each race group. For example, the CBCL Aggressive Behavior syndrome consists of 20 items, which results in 60 parameters (20 discrimination parameters, 40 threshold parameters) for each group. These parameters and functions were used to examine the psychometric properties of the CBCL scales to determine utility of items and scales for the African-American and Euro-American groups.

Differential Functioning Analyses

IRT assesses measurement equivalence via evaluation of differential item functioning (DIF), which represents the extent to which individual items differ in various aspects of their relation to the hypothesized underlying latent construct. DIF occurs when respondents with equal latent trait from different subgroups (e.g. ethnic groups, gender

groups) do not have an equal probability of endorsing an item (Hambleton & Swaminathan, 1985; Hambleton & Rogers, 1989). Thus, measurement inequivalence is found when one or more parameters are different between the African-American and Euro-American groups.

Equating Procedure

In order for IRT parameters to be comparable, it is necessary to equate or link the parameters of each group to the same metric. For the purposes of this study, parameters were equated to the metric of the African-American group (because African-Americans are the focal group whereas Euro-Americans are the reference group). The EQUATE 2.1 program (Baker, 1995) was used for linking the metric of the two groups using a linear transformation. Linking of metric should be based on non-dif items as they impact the transformation equation, therefore the additive and multiplicative coefficients were estimated iteratively before and after detection of DIF, following recommendations of Candell and Drasgow (1988). In other words, DIF analyses are iteratively conducted, with items as identified as showing DIF dropped, until a DIF-free subset of items are available for estimating the additive and multiplicative coefficients used in the transformation equation.

Differential Functioning of Items and Tests (DFIT)

After the parameters from both groups are equated using non-dif items, a comparison of the threshold and discrimination parameters across the race groups was made using the Differential Functioning of Items and Tests framework (DFIT: Raju et al., 1995). The DFIT framework allows for an examination of differential test functioning

(DTF), compensatory differential functioning (CDIF), and non-compensatory (NCDIF) item functioning.

In the DFIT framework, differential functioning is defined by comparing the true score of the focal group based on parameter estimation for the reference group and the focal group. Thus, each member of the focal group will have two true scores – one as a member of the focal group and one as a member of the reference group. When this true score is summed across the test, expected true test scores can be calculated for each focal group examinee.

For polytomously scored data, an expected score (ES_{si}) for item #*i* can be computed for examinees as

$$ES_{si} = \sum_{k=1}^{m} P_{ik}(\theta_s) X_{ik} \tag{3}$$

Where,

X_{ik} is the score or weight for category *k*

m is the number of categories

P_{ik} is the probability of responding to category *k* (see Equation 2)

This is referred to as the *expected item score function* or the IRF. Summing the expected item scores across a test will result in the *expected test score function* for each examinee as

$$T_s = \sum_{i=1}^{n} ES_{si} \tag{4}$$

In this study, the African-American group was considered the focal group, therefore two sets of expected true scores were estimated for the African-American examinees, one using their own parameters (T_{sF}) and the other using the equated

European-American parameters (T_{sR}). The greater the difference between the two expected scores, the greater the differential functioning. If the item is functioning differentially, the two expected scores will not be equal.

The same reasoning can be applied at the test level and is referred to as Differential Test Functioning (DTF). DTF represents the average squared difference in true scores between the two groups at the subscale level. When item parameters are equal across the two groups, measurement equivalence is achieved because expected true score for a particular ability level should be the same irrespective of group membership. The DTF reflects the sum of this difference across items. However, the additive nature of DTF allows for possible cancellation at the test level. This occurs when one item displays DIF in favor of one group and another item displays DIF in favor of the other group. This combination of DIF items will have a canceling effect on the overall DTF. The square root of the DTF approximates the mean (across participants) absolute true score difference at the scale level.

$$DTF = \mathbb{E}_f (T_{sF}-T_{sR})^2 = \mathbb{E}_F (D_i)^2 = \sigma_D^2 + \mu_D^2 \quad (5)$$

Where the expectation ϵ is taken across the entire focal group.

Two indicators of differential item functioning can then be determined using DTF. Compensatory DIF estimates DIF under no assumptions about the extent of DIF in other items in the scale. The sum of all CDIF items across a test or scale is equal to the overall DTF. The formula for CDIF is equal to:

$$CDIF_i = \mathbb{E} (d_{is,} D_s) = \mathbb{E}((P_{iF} (\theta) - P_{iR} (\theta)), T_{sF}-T_{sR})) \quad (6)$$

where $P_i(\theta)$ is the probability that individuals with ability level θ answer item *i* correctly,

$$d_{is} = \epsilon_{iF}(\theta) - \epsilon_{iR}(\theta), \text{ and } D_s = T_{sF} - T_{sR} \quad (7)$$

The sum of the CDIF indices reflects the net directionality. For practical applications, a test developer could examine the DTF, then determine which item(s) should be eliminated based on its CDIF value and its overall contribution to DTF.

Non-compensatory DIF (NCDIF), however, is a special case of CDIF in which DIF is estimated under the assumption that all other items in the set of items are free from DIF. NCDIF is calculated by

$$NCDIF_i = \epsilon_f (P_{iF}(\theta_i) - P_{iR}(\theta_i))^2 = \epsilon_f d_i^2 = \sigma_{di}^2 + \mu_{di}^2 \quad (8)$$

According to Raju et al. (1995) the NCDIF index is the most useful criterion of differential functioning at the item level as it indexes the absolute value of DIF at the item level, while DTF is the most useful criterion of differential functioning at the scale level.

The NCDIF estimates the latent trait based on African-American derived versus Euro-American derived parameters, by addressing the following questions: (a) what is the person's true score on an item when he or she is viewed as a member of the African-American (focal) group? (b) What is the person's true score on an item when he or she is viewed as a member of the Euro-American (reference) group? When the true scores are identical, the difference between them is zero. Therefore, differential item functioning occurs when the difference in true scores is significantly different from zero, which results from differences in item parameters. Thus when NCDIF is zero, the IRFs must

also be identical. Conversely, when item parameters are not equal, NCDIF will be different from zero, and differential functioning is identified.

DFIT Significance Testing

Chi-square tests of significance are provided for the DTF and NCDIF indexes of the DFIT Framework (Raju et al.,1995). A significant χ^2 indicates that the observed DTF measure is significantly different from zero and indicates that one or more items are functioning differentially. However, Fleer (1993) points out that the χ^2 tests of significance in the DFIT Framework are overly sensitive with large sample sizes and may result in incorrect identification of differential functioning. The large sample size required for IRT analysis typically result in statistically significant χ^2 when NCDIF is very small. Fleer (1993) suggested empirically establishing a critical (cutoff) value for NCDIF and DTF. Based on Monte Carlo studies, Raju, van der Linden, and Fleer (1995) recommend .024 as the cut-off for assessing whether or not a NCDIF index is significant for a polytomous 3 category rating scale. Thus, if an item has an NCDIF index greater than .024 and a statistically significant chi-square it is said to have significant DIF.

The square root of the NCDIF and DTF index may be viewed as an approximation to the average absolute difference between the two (item) true scores. For the recommended NCDIF cut-off of .024, a scaled score of .15 difference on a 3 category item is considered significant and meaningful.

Person-level Differential Item Functioning (PDIF)

The following procedure was used to assess the degree to which various person-level contextual variables (e.g., SES) were related to the magnitude of differential functioning in the African-American sample. For each African-American participant, we

(a) estimated the true score for each CBCL item on each of the CBCL scales, based on the Euro-American parameters, (b) estimated the true score for each of those items based on the African-American parameters, and (c) computed the difference between these two true score estimates (because we are using a two parameter model, individuals with the same total observed score can have different estimated true scores). The difference between these two estimated true or latent scores represents differential functioning at the person level, which we refer to as Person Differential Item Functioning (PDIF).

To determine the relation between PDIF and the person-level contextual factors (e.g., SES; parenting values), we correlated each of the contextual factors with the PDIF from each CBCL item showing significant DIF. This provided an estimate of the total relation between the differential item functioning, and the contextual factor. To estimate the unique relation of the contextual predictors within each set of contextual factors (as well as to determine the relation between sets of contextual factors and PDIF), the six sets of contextual factors (SES, census block neighborhood characteristics, parental perceptions of the neighborhood, family expressivity, parenting values, and parenting practices) were regressed on each CBCL item's PDIF. As with the PDIF / Contextual Factors correlations, only CBCL items showing significant DIF were analyzed.

CHAPTER III

RESULTS

Classical Test Statistics

Estimates of Cronbach's alpha reliabilities, means, and variances, kurtosis, and skewness for each group are presented below.

Table 1. Scale Reliabilities for Euro-American CBCL Subscales

Subscales	N	# of Items	Alpha	Scale Mean	Scale Variance	Kurtosis	Skewness
Aggression	572	20	.89	.67	.12	.16	.54
Anxiety	572	14	.82	.35	.09	1.65	1.21
Attention	572	11	.75	.46	.09	.20	.71
Delinquency	572	12	.69	.27	.04	1.97	1.23
Social	572	8	.62	.41	.08	.21	.71
Somatic	572	9	.61	.15	.03	2.81	1.59
Thought	572	7	.57	.16	.04	4.12	1.80
Withdrawn	572	9	.65	.28	.06	1.04	1.08

Table 2. Scale Reliabilities for Afro-American CBCL Subscales

Subscales	N	# of Items	Alpha	Scale Mean	Scale Variance	Kurtosis	Skewness
Aggression	578	20	.88	.67	.12	-.34	.23
Anxiety	578	14	.79	.30	.07	1.74	1.19
Attention	578	11	.76	.48	.10	.41	.69
Delinquency	578	12	.72	.31	.05	1.14	.94
Social	578	8	.53	.44	.08	-.22	.51
Somatic	578	9	.73	.14	.05	11.73	2.59
Thought	578	7	.55	.17	.05	2.45	1.62
Withdrawn	578	9	.70	.29	.07	1.16	1.12

As can be seen in Tables 1 and 2, the Aggressive Behavior (E-A=.89, A-A=.88) and Anxious / Depressed (E-A=.82, A-A=.79) scales showed the highest reliability, with similar reliabilities across the two groups, whereas the Social Problems (E-A=.62, A-A=.53) and Thought Problems (E-A=.57, A-A=.55) scales showed the lowest reliability across both groups. Of note was the race difference in reliability for the Somatic Complaints scale, showing greater internal consistency for the African-American (.73) than the Euro-American group (.61). The Somatic Complaints scale was also highly kurtotic for the African-American group but not for the Euro-American group, which suggests differential functioning of this scale when comparing African-Americans and Euro-Americans.

Race Differences

CBCL

Significant race differences were found in both the raw scale and T scores for the Anxious / Depressed, Social Problems, and Delinquency scales, with African-Americans observed to have higher observed scores for Social Problems and Delinquency scales and lower scores for the Anxious / Depressed scale.

Table 3. Total Raw Score Differences Between Race Groups

Scales	A-A (N=578)	E-A (N=572)	*F value*
Anxious/Depressed	4.14	4.85	-3.16[1**]
Social Problems	3.52	3.22	2.22*
Attention Problems	5.26	5.11	.74
Delinquency Problems	3.34	2.94	2.69[1**]
Aggressive Behaviors	13.11	13.29	-.45
Thought Problems	1.19	1.13	.6[1]
Withdrawn / Depressed	2.59	2.54	.31[1]
Somatic Complaints	1.29	1.32	-.20[1]

1=unequal variances, used Satterthwaite t-tests, *<.05, **<.01, ***<.001

Table 4. Total T-Score Differences Between Race Groups

Scales	A-A (N=578)	E-A (N=572)	*F value*
Anxious/Depressed	55.79	57.00	-2.80[1**]
Social Problems	58.85	57.92	2.00*
Attention Problems	58.94	58.66	.58
Delinquency Problems	60.15	58.81	2.79**
Aggressive Behaviors	60.10	60.27	-.34
Thought Problems	57.09	57.06	.06
Withdrawn / Depressed	56.24	55.94	.74[1]
Somatic Complaints	55.63	55.59	-.42[1]

1=unequal variances, used Satterthwaite t-tests
*<.05, **<.01, ***<.001

Contextual Variables

Summary Statistics for context variables and T-test comparisons between race groups can be found in Table 5.

Table 5. Descriptive Statistics for Contextual Variables

	African-Americans			European-Americans			
Variables	N	Means	SD	N	Means	SD	T value
SES							
SES	570	23.23	13.04	568	27.38	12.52	-5.48****
FAMOC	570	2.65	2.48	568	3.97	2.35	-9.21***
Neighborhood Disadvantage							
POVERTY	569	0.31	0.16	560	0.12	0.08	24.1[1]***
PUB ASSIST	569	0.16	0.08	560	0.07	0.05	22.11[1]***
FEM HOUSE	569	0.45	0.17	560	0.17	0.09	34.57[1]***
UNEMPLOY	569	0.11	0.06	560	0.06	0.03	16.32[1]***
BLACK	569	0.69	0.31	560	0.07	0.14	43.04[1]***
NOMOVE	569	0.47	0.10	560	0.58	0.11	-16.31***

Variables	N	Means	SD	N	Means	SD	T value
Neighborhood							
SOCIAL	571	14.41	9.27	569	18.07	9.78	-6.48***
SERVICE	569	15.36	4.13	560	14.01	4.36	5.37***
SAFETY	571	28.63	12.78	569	36.28	8.54	-11.89[1]***
Family							
CONSIST	428	2.16	0.57	451	2.36	0.59	-5.04***
EFFECT	428	1.78	0.54	451	1.81	0.48	-.074
PUNITIVE	428	2.49	0.61	451	2.59	0.50	-2.55[1]*
POSITIVE	430	2.20	0.52	457	2.17	0.47	.92
NEGATIVE	430	1.91	0.59	457	1.82	0.51	2.21[1]*
DEFENSE	414	9.66	3.50	457	8.16	3.17	6.61[1]***
SOC	414	17.24	2.38	457	17.53	2.12	-1.86

Labels: SES = SES Continuous Code, FAMOC = SES Higher Family Occupation Code, SOCIAL = Neighborhood Social Connectedness, SERVICE = Neighborhood Public Services, SAFETY = Neighborhood Safety, POVERTY = % poverty, PUB ASSIST = % public assistance, FEM HOUSE = % female headed household, BLACK = % of African-American residents, UNEMPLOY = % unemployment, NOMOVES = % no moves in 5 years, NEGATIVE= Negative Expression, POSITIVE = Positive Expression, CONSIST = Consistent parenting, EFFECT = Effective parenting, PUNITIVE = Punitive parenting, DEFENSE = Value for Self Defense, and SOC = Value for social Involvement.
Note: 1=unequal variances, used Satterthwaite t-tests, *<.05, **<.01, ***<.001

Using Hollingshead index, which defines socioeconomic status by occupation and education, African-American families in this sample were of significantly lower SES compared to the European-Americans (A-A=23.23, E-A=27.38, p<.001) when the mean SES for both parents was calculated. This difference was even greater when the higher occupation of the two parents were used as an index for SES (FAMOC; A-A=2.65, E-A=3.97, p<.001). However, both groups fell within the same social strata, semi-skilled workers and machine operators.

Significant socio-economic differences between the two groups also occurred at the neighborhood level, based on U.S. census tract data. On average, African-Americans lived in neighborhoods (defined by tracts) where 31% were living below poverty level in 1990, 16% of their neighbors were on public assistance, 45% lived in female headed households, and 11% of adults over 16 years old were unemployed. They also lived in predominantly African-American communities (69%) and had 47% neighborhood stability (defined by neighbors who did not move within 5 years).

In comparison, European-Americans only had 12% of their neighbors living below poverty level, 7% on public assistance, and 17% female headed households, and 7% neighborhood unemployment. European-Americans also had higher percentage of neighbors who did not move between 1985 to1990. African-Americans rated their neighborhoods to have lower social involvement and connectedness and sense of safety than European-Americans. However, they perceived their neighborhoods to have higher public services.

In terms of parenting behaviors, African-American parents rated themselves using punitive parenting behaviors less frequently and to be less consistent in their parenting; no differences were found in their levels of perceived effectiveness. Significantly higher levels of openness and responsiveness towards negative events were reported by African-Americans, but no significant differences were found for the expression of positive affect. In addition, African-American parents placed greater value on their children's ability and need for self-defense in their environment, but not for the degree to which they (the parents) should be involved in their child's social relationships.

IRT Model Assumptions

Unidimensionality. The eigenvalues and percent contribution of each item as indicated by the EFA of each rater source are presented in Table 6. The percent of total variance accounted for by the first component in each analysis ranged from 88% to 100%, well above the 20% criteria suggested by Reckase (1979). Also, the eigenvalue ratios (1st to 2nd eigenvalue) all were greater than 3 (Waller et al., 1990), except for Thoughts Problems for African-Americans (eigenvalue ratio = 2.90). Scree plots for all scales (presented in the appendix) also showed clear unidimensionality. The results from the EFA, eigenvalue ratios, and scree plots indicate unidimensionality for all but the Thought Problem subscale for African-Americans. Thus, for almost all of the CBCL scales, both unidimensional and local independence assumption were met for IRT analysis.

Table 6. Dimensionality for CBCL Subscales for European-Americans

Subscales	N	# of Items	1st Eigen	2nd Eigen	1:2 Ratio	Variance Accounted
Aggressive Behavior	572	20	6.06	.92	6.59	81%
Anxious / Depressed	572	14	3.50	.60	5.83	85%
Attention Problems	571	11	2.59	.65	3.98	80%
Delinquent Behavior	569	12	2.27	.40	5.68	100%
Social Problems	570	8	1.50	.27	5.56	100%
Somatic Complaints	568	9	1.74	.23	7.57	100%
Thought Problems	572	7	1.22	.37	3.30	100%
Withdrawn / Depressed	570	9	2.36	.22	10.73	100%

Table 7. Dimensionality for CBCL Subscales Afro-Americans

Subscales	N	# of Items	1st Eigen	2nd Eigen	1:2 Ratio	Variance Accounted
Aggressive Behavior	577	20	5.43	.80	6.79	81%
Anxious / Depressed	574	14	3.10	.53	5.85	85%
Attention Problems	575	11	2.57	.53	4.85	83%
Delinquent Behavior	572	12	2.31	.57	4.05	80%
Social Problems	578	8	1.29	.29	4.45	100%
Somatic Complaints	570	9	2.37	.22	10.77	100%
Thought Problems	572	7	1.13	.39	2.90	100%
Withdrawn / Depressed	572	9	1.85	0.16	11.56	100%

Eliminated Scales

Based on lack of fit to a 1-factor model, the Thought Problem scale was eliminated because it did not fit for African-Americans. However, it showed adequate fit for the European-Americans. Although we did not conduct DFIT analyses on the Thought Problem Scale, the difference in fit also suggests differential functioning. Developmentally, the Thought Problem scale probably is not appropriate for 6 year olds, who are unlikely to exhibit psychotic symptoms, and not surprisingly, this scale showed low variability. Consequently, this scale was not included in subsequent analyses.

In addition, the Delinquency Scale also was dropped from analyses because a number of its items also are developmentally inappropriate (e.g., “running away”, “vandalizing”, “thinking about sex”, “using illegal drugs”) for six year olds. These items had limited variability, with no parents endorsing the “2” option (“all of the time”), suggesting a dichotomous model rather than the polytomous model used in this study.

Eliminated Item

On the somatic scale, one item ("Vomits") also was not endorsed at the "2" level by any parents. Because this was the only item on this scale that showed a limited range, the item rather than the scale was dropped. This pattern was found in both groups.

IRT Parameter Estimation

PARSCALE (Muraki & Bock, 1993) was used to estimate the two-parameter logistic model for each scale of the CBCL, except for the Thought Problems and Delinquency scales. The unequated (across groups) item and category parameters from PARSCALE are reported in Table 8 for each scale.

Table 8. Unequated Parameter Estimates for CBCL Items.

		African-Americans			European-Americans		
Agg	Items	α	β1	β2	α	β1	β2
3	*Argues*	1.178	-1.53	1.275	1.431	-2.489	0.749
7	*Brags*	0.802	-0.47	3.077	0.685	-1.019	3.661
16	*Cruelty*	0.373	-2.64	1.658	1.59	0.261	2.837
19	*Demands attention*	1.035	-1.6	0.705	1.148	-1.414	1.176
20	*Destroys own things*	1.307	-0.29	1.797	1.398	0.368	2.462
21	*Destroys others things*	1.646	0.018	2.005	1.387	0.554	2.722
22	*Disobedient at home*	1.376	-1.32	1.997	1.663	-1.613	1.713
23	*Disobedient at school*	1.249	-1.2	2.119	1.001	-0.09	3.799
27	*Easily jealous*	0.861	-1.53	1.559	0.84	-1.281	2.169
37	*Gets in many fights*	1.085	0.408	2.992	1.616	0.557	2.731
57	*Physically attacks others*	0.198	-2.13	0.949	1.713	1.29	3.763
68	*Screams*	1.139	0.804	3.073	1.261	0.239	2.347
74	*Showing off*	1.184	-1.4	1.378	1.193	-1.759	1.573
86	*Sullen*	1.144	-1.23	1.706	1.304	-1.781	1.316
87	*Sudden changes in mood*	0.823	0.004	3.562	1.023	0.013	3.151
93	*Talks too much*	0.957	-1.47	0.746	0.871	-0.958	2.046
94	*Teases a lot*	1.347	-0.46	2.03	1.224	-0.286	2.665
95	*Temper tantrums*	1.263	-0.4	1.669	1.446	-1.023	1.404
97	*Threatens people*	1.447	1.186	3.529	1.847	1.159	3.312
104	*Loud*	1.143	-0.53	1.704	1.341	-0.135	2.24
Anx							
12	*Lonely*	0.607	1.005	4.416	0.899	0.539	3.296
14	*Cries*	0.623	0.456	3.905	1.025	0.561	3.278
31	*Fearsbad*	1.109	1.036	3.797	0.307	-2.19	0.978
32	*Perfect*	1.278	0.943	2.772	1.09	0.569	2.752

Anx	Items	α	β1	β2	α	β1	β2
33	*No love*	1.116	1.027	3.765	1.166	0.169	3.109
34	*Others are out to get*	2.431	1.26	2.793	1.791	0.962	2.99
35	*Worthless*	2.502	1.445	2.882	1.57	1.043	3.341
45	*Nervous*	0.923	0.976	3.503	0.829	0.395	3.102
50	*Fearful*	1.243	0.696	3.076	1.258	0.385	3.104
52	*Guilty*	1.749	1.347	3.417	1.223	1.769	5.678
71	*Self-conscious*	1.082	-0.2	2.492	1.109	-0.903	1.99
89	*Suspic*	0.918	1.126	3.441	0.156	-2.786	1.556
103	*Unhappy*	1.777	1.133	3.951	0.222	-3.069	1.724
112	*Worries*	1.676	1.35	3.401	1.319	-0.006	2.874
Att							
1	*Acts too young*	0.771	0.279	3.459	0.789	-0.079	3.393
8	*can't concentrate*	2.053	-1.094	0.961	2.063	-1.126	0.833
10	*can't sit still*	1.475	-1.707	0.379	2.005	-1.085	0.477
13	*Confused*	0.013	-34.537	33.355	0.246	-1.833	0.507
17	*Daydreams*	0.776	0.593	4.136	0.444	0.566	8.639
41	*Impulsive*	1.121	-0.645	2.241	1.52	-1.223	1.612
45	*Nervous*	0.987	0.796	3.275	1.254	0.159	2.159
46	*Twitches*	0.181	-1.798	0.819	0.064	-2.375	1.375
61	*poor school work*	0.988	0.727	3.153	0.25	-2.672	1.595
62	*Clumsy*	0.272	-2.124	1.169	0.034	-14.118	12.991
Soc							
1	*Acts too young*	0.596	0.471	4.387	0.686	0.236	4.068
11	*Clings*	1.023	-0.548	1.819	0.723	-0.226	3.01
25	*Doesn't get along w/other kids*	0.706	0.033	5.08	1.103	-0.076	4.431
38	*Gets teased*	1.336	-0.355	2.092	1.269	0.094	2.477
48	*Not liked by other kids*	0.177	-2.737	1.519	2.445	0.99	3.307
55	*Overweight*	0.055	-1.901	1.901	0.132	-1.212	-0.141
62	*Clumsy*	0.186	-2.889	1.898	0.028	-16.906	15.729
64	*Prefers younger kids*	0.431	-0.543	6.018	0.499	0.383	6.736
Soma							
51	*Dizzy*	0.131	-1.351	-0.445	0.066	-1.375	-0.609
54	*overtired*	0.699	1.09	4.709	0.443	1.177	7.625
56a	*Aches*	0.206	-1.708	0.13	2.273	1.097	3.128
56b	*headaches*	2.305	0.747	2.599	1.551	1.084	3.889
56c	*Nausea*	0.221	-1.537	-0.074	0.16	-2.323	0.805
56d	*Eyeprob*	1.422	1.731	2.752	0.075	-1.63	0.067
56e	*skinprob*	0.2	-1.73	0.629	0.131	-1.737	0.142
56f	*stomachaches*	1.95	0.751	2.648	1.756	0.859	3.272
With							
42	*Prefers to be alone*	1.069	1.012	3.661	0.724	0.936	3.334
65	*Refuses to talk*	1.08	1.065	3.533	1.109	0.173	2.34
69	*Secretive*	0.153	-3.726	2.464	1.162	0.154	1.906
75	*Shy*	0.827	0.084	3.544	0.765	-0.44	2.344
80	*Stares blankly*	0.043	-7.737	6.472	0.98	0.609	3.377
88	*Sulks*	0.864	0.225	2.948	0.979	-0.01	1.86
102	*Underactive*	1.543	1.818	4.417	0.711	1.883	3.99
103	*Unhappy*	1.669	1.231	4.18	1.045	0.15	3.025
111	*Withdrawn*	1.399	1.751	4.052	2.238	0.068	1.532

Prior to the DFIT analysis the item parameters for the samples were equated on the African-American metric using Equate (Baker, 1993). As described in the Methods section, the equating of parameters is an iterative process performed by removing parameters for items found to have DIF or poor functioning (i.e., low discrimination parameter) and re-equating based on non-dif items. Items are said to display DIF if the NCDIF value is greater than 0.024 and is significant at the 0.01 level. This process is continued until the same items are identified as having DIF across two consecutive iterations (Flanagan, 1997). These DIF items are then removed from the equating process. The final equate constants are based on final set of good functioning (high discrimination) and non-dif items, which was used to anchor the metric for all items on a particular scale.

Differential Test and Item Functioning

The DFIT analyses were conducted with the African-American sample considered to be the focal group and European-Americans the reference group. Three IRT based differential functioning index were used, the DTF, CDIF, and NCDIF. At the item level, NCDIF was considered practically significant when the index was above .024 (equivalent to .15 scaled score difference between the two groups) for a 3-category item and accompanied by a significant chi-square test where the p-value was less than .01 (Fleer, 1993; Flowers et al., 1995, & Collins et al., 1997). Significance for the CDIF is determined by the effect it has on the DTF (i.e., the items that contributing most to DTF [for the CDIF] was deleted until DTF reached non-significant levels). This method was

used to determine which items contributed the most to the scale level differential functioning.

Differential Test Functioning

The DTF, square root of the DTF, and critical cut-off for the DTF for the six CBCL subscales used in the DIF analyses are presented in Table 9. Included in this table are also the list of items in the final iteration of equating and the associated linking coefficients. Scales with the largest DTF index are presented first. Following Raju's (1995) recommendations, the DTF index was considered practically significant if the index is greater than the cut-off for each scale and the associated significance level of the chi-square was less than .01. Following these criteria, all six scales analyzed showed differential functioning at the test level, except for the Somatic Complaints scale. It is important to note that the square root of the DTF represents the scaled scored difference between the two groups, if they were at the same latent trait. The Anxious / Depressed scale (DTF=4.32) and the Attention Problems scale (DTF = 3.50) had the largest DTF. This was followed by the Aggressive Behavior scale (DTF = 1.56), Withdrawn scale (DTF=1.31) and the Social Problems scale (DTF = .74). The Somatic Complaints Scale did not show significantly different scale level (DTF=.034, p-value of χ^2=.423), however, there were two items that showed differential functioning in opposite directions, thus canceling out the scale level DIF. Since the square root of the DTF represents the mean difference in observed scores for Euro-Americans and African-Americans for a given trait level, on average the two groups differed by 2.08 on the observed scale score for the

Anxiety scale, 1.87 on the Attention scale, 1.25 on the Aggression scale, 1.15 on the Withdrawn scale, and .86 on the Social Problems scale.

Table 9. Differential Test Functioning of CBCL Scales and Items used for Equating Metric and Linking Coefficients

Scale	N	DTF	rDTF	Cut-off	Equate Items	A	K
Anxiety	578	4.32*	2.08	.336	12,32-35,45,50,71	.894	.502
Attention	578	3.50*	1.87	.264	1,8,41,45	.888	-.325
Aggression	578	1.56*	1.25	.480	7,19,21,22,27,37, 74,87,94,97,104	1.003	-.177
Withdrawn	578	1.31*	1.15	.216	42,65,75,88	1.040	.693
Social	578	.744*	.86	.192	1,11,25,38	.985	-.250
Somatic	578	.034	.184	.192	51,54,56b,56c, 56e,56f	.685	.036

*Note. *p<.001*

Differential Item Functioning

This section summarizes the item level differential functioning in tables and illustrated in DIF graphs. Patterns of differential functioning are also described in greater detail in this section. Substantive interpretation of results is reserved for the Discussion section. In interpreting these results, the discrimination parameter should be viewed to as an estimation of the magnitude of the non-linear relation between the CBCL item and the latent construct represented by the CBCL scale. When discrimination parameters are above .5, threshold parameters should be viewed as an estimate of the extent to which the CBCL item is viewed as a serious sign of the latent trait represented by the CBCL factor. Higher thresholds indicate that an item is seen as a more serious sign or indicator of the latent construct. For instance, the discrimination parameters for items #19 and #20 ("Demands attention" and "Destroys own things") were well above .5, in Baker's (2001) Moderate to High ranges, for both groups. The thresholds for the 0 / 1

response option for both groups for "Demands attention" were approximately -1.6, whereas the thresholds for the "Destroys own things" were notably higher, -.3 and +.2 for the African-Americans and the European-Americans, respectively. This indicates that both groups, in particular the European-Americans, see a child who receives a "1" on the "Destroys own things" CBCL item as evincing higher levels of the underlying latent trait of "Aggressive Behavior" than children who receive a "1" on the "Demands attention" item. When discrimination parameters are below .5, threshold parameters should not be interpreted as the low discrimination parameter indicates that the items is not seen as a sign of the latent construct.

Anxious / Depressed

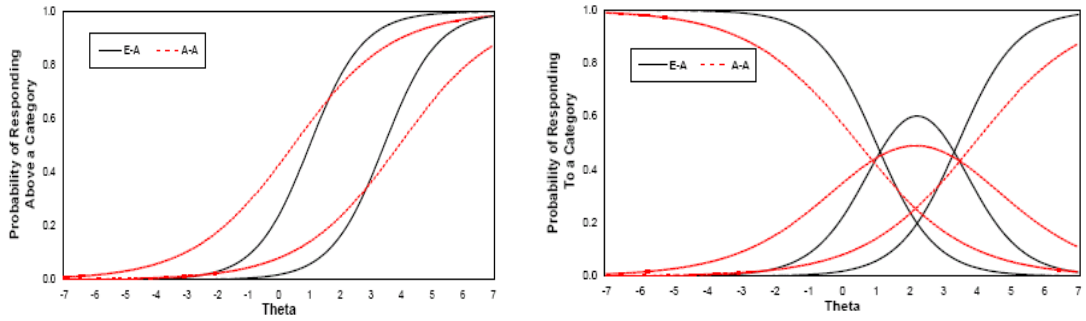
Four of 14 items on the Anxious / Depressed Scale showed differential functioning based on the NCDIF and CDIF. The item with the largest magnitude was item #103 "Unhappy" (NCDIF=.66, CDIF=1.68), which did not discriminate well for European-Americans (Ra=.25) but did discriminate well for African-Americans (Fa=1.78). Another way to conceptualize this is that the item "Unhappy, sad or depressed" was unrelated to the latent construct for the Euro-Americans but strongly related for the Afro-Americans. The threshold was also lower for European-Americans than for African-Americans, suggesting that it requires more the latent of the anxiety/depression factor for African-Americans to endorse this item. That is, African-Americans must be experiencing higher levels of the anxiety and depression, relative to the Euro-Americans, before they endorse this item.

Once this item was removed from the scale, the DTF was no longer significant, indicating that this item was responsible for most of the scale level differential functioning. Three other items also exhibited DIF, but did not contribute to overall DTF. These items included item #31 "Fears doing bad," item #89 "Suspicious," and item #14 "Cries." Items #31 and #89 exhibited a similar pattern to #103, in that they also exhibited poor discrimination for European-Americans and higher thresholds for African-Americans. Thus, unlike for their European-American counterparts, behaviors reflective of fear and paranoia were highly related to anxiety. In contrast, item #14 "Cries a lot," discriminated poorly for African-Americans, while discriminating well for European-Americans. For the exception of item #14, this scale functioned better for African-Americans than their European-American counterparts. The BRFs (Boundary Response Functions) and CRFs (Category Response Function) for the Anxious/Depressed scale are presented below in Figure 3.

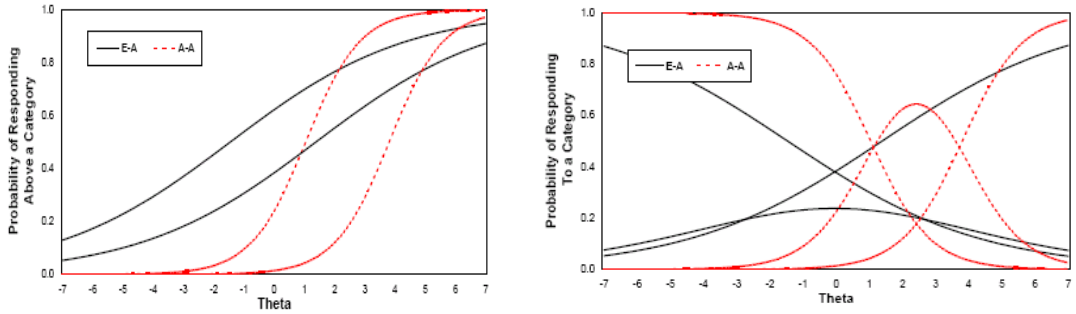
Table 10. DFIT Results for Anxious/Depressed Scale

Item	Labels	Fa	Fb1	Fb2	Ra	Rb1	Rb2	CDIF	NCDIF
12	*Lonely*	0.607	1.005	4.416	1.006	0.984	3.449	-0.196	0.012
14	*Cries*	0.623	0.456	3.905	1.147	1.004	3.433	-0.447	.047*
31	*Fears bad*	1.109	1.036	3.797	0.343	-1.456	1.376	1.456	.494*
32	*Perfect*	1.278	0.943	2.772	1.219	1.011	2.962	-0.022	0.001
33	*No love*	1.116	1.027	3.765	1.304	0.653	3.281	0.115	0.006
Item	**Labels**	**Fa**	**Fb1**	**Fb2**	**Ra**	**Rb1**	**Rb2**	**CDIF**	**NCDIF**
34	*Others are out to get*	2.431	1.26	2.793	2.003	1.362	3.175	0.001	0.002
35	*Worthless*	2.502	1.445	2.882	1.756	1.434	3.489	0.073	0.003
45	*Nervous*	0.923	0.976	3.503	0.927	0.855	3.275	0.06	0.001
50	*Fearful*	1.243	0.696	3.076	1.407	0.846	3.277	-0.125	0.004
52	*Guilty*	1.749	1.347	3.417	1.368	2.083	5.578	-0.131	0.018
71	*Self Conscious*	1.082	-0.2	2.492	1.24	-0.305	2.281	0.051	0.001
89	*Suspicious*	0.918	1.126	3.441	0.174	-1.989	1.893	1.383	.451*
103	***Unhappy***	**1.777**	**1.133**	**3.951**	**0.248**	**-2.242**	**2.043**	**1.68**	**.657***
112	*Worries*	1.676	1.35	3.401	1.475	0.497	3.071	0.416	0.053

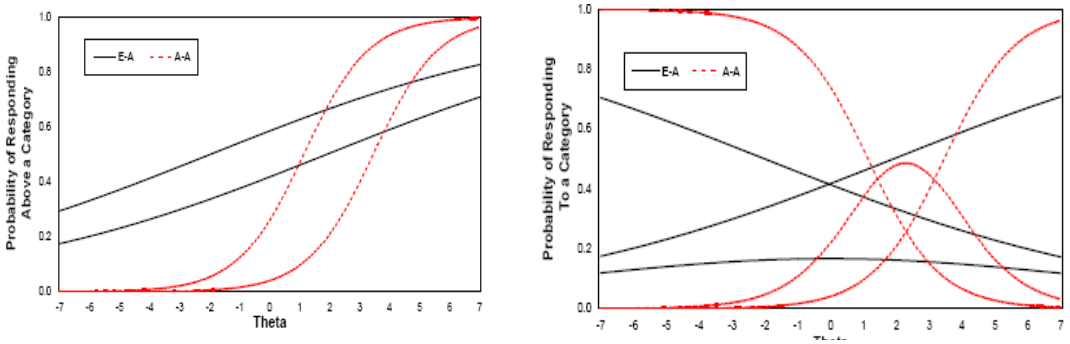
Note. * NCDIF >.024



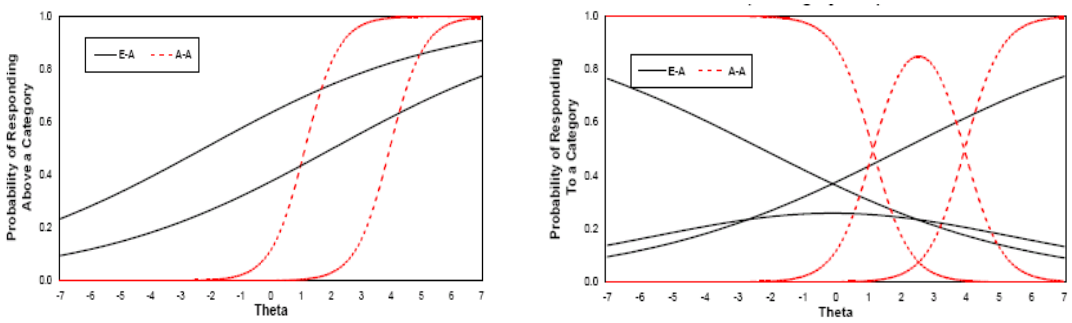
Item #14 "Cries a lot"



Item #31 "Fears he/she might think or do something bad"



Item #89 "Suspicious"



Item #103 "Unhappy, sad, or depressed"

Figure 3. BRFs and CRFs for Anxious/Depressed Scale

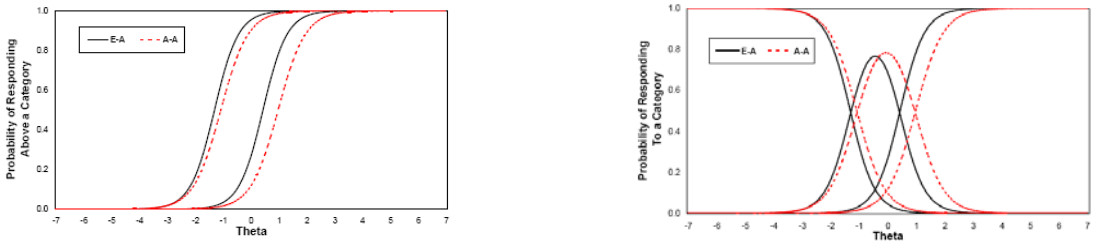
Attention Problems

Five items (items #8, #15, #41, #45, #61) on the Attention Problems scale showed significant differential functioning, with three items contributing to scale level DIF. These included item #61 "Poor school work," item #45 "Nervous movements," and item #41 "Impulsive", in order of greatest magnitude to least. The DIF contributed by item #61 ("Poor school work") was the largest and reflected poor discrimination for European-Americans whereas this item reflecting poor school work was highly related to attention problems for African-Americans, particularly for those with above average levels of attention problems. Both items #41 and #45, which may involve more hyperactivity or hyperkinetic sensitivities, were less discriminating and more "difficult" (higher thresholds) for African-Americans. The two other DIF items (which did not contribute to scale level DIF), item #8 ("Can't concentrate") and item #13 ("Confused") were both also not as discriminating and exhibited higher thresholds for African-Americans. Furthermore, the parameters for item #13 (Fa= .01, Fb1=-34.54, Fb2=33.35) were found to be highly extreme for African-Americans, suggesting that this item is a very poor item for African-Americans due to its remarkably low discrimination and consequently extreme threshold parameters. As can be seen in the Figure 4, items on this scale generally functioned more poorly (lower discrimination) and required higher thresholds for African-Americans (except for items #8 and #61) than for their European-American counterparts. However, it is important to note that for this scale, numerous items showed poor functioning for both groups (Items #13, #17, #46, #62, and #80).

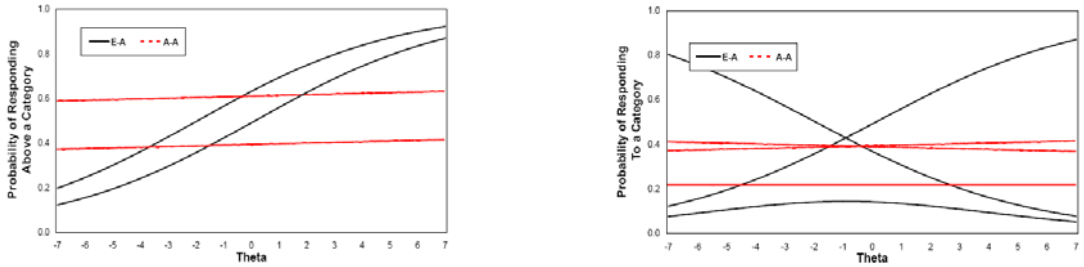
Table 11. DFIT Results for Attention Problems Scale

Item	Labels	Fa	Fb1	Fb2	Ra	Rb1	Rb2	CDIF	NCDIF
1	Acts too young	0.771	0.279	3.459	0.889	-0.39	2.693	0.286	0.024
8	Can't concentrate	2.053	-1.094	0.961	2.323	-1.32	0.42	0.372	.041*
10	Can't sit still	1.475	-1.707	0.379	2.258	-1.283	0.104	0.124	0.021
13	Confused	0.013	-34.537	33.355	0.277	-1.948	0.13	0.248	.029*
17	Daydreams	0.776	0.593	4.136	0.5	0.183	7.351	0.091	0.009
41	Impulsive	1.121	-0.645	2.241	1.712	-1.406	1.111	0.576	.097*
45	Nervous	0.987	0.796	3.275	1.412	-0.179	1.597	0.603	.121*
46	Twitches	0.181	-1.798	0.819	0.072	-2.429	0.901	-0.043	0.003
61	poor school work	0.988	0.727	3.153	0.282	-2.693	1.096	1.242	.488*
62	Clumsy	0.272	-2.124	1.169	0.038	-12.857	11.216	-0.113	0.014
80	stares blankly	0.085	-4.275	3.074	0.199	-2.406	0.701	0.119	0.006

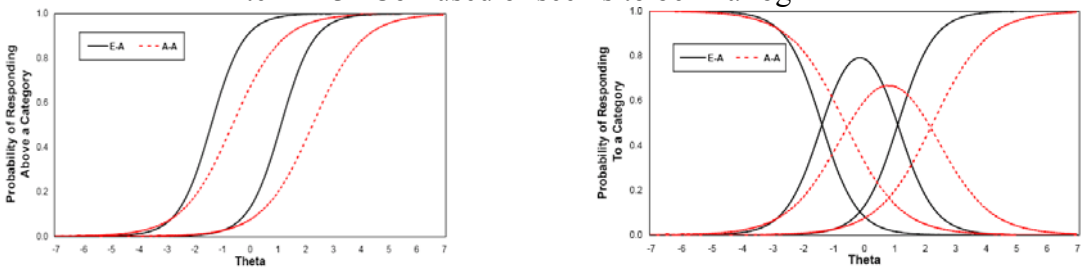
Note. * = NCDIF >.024



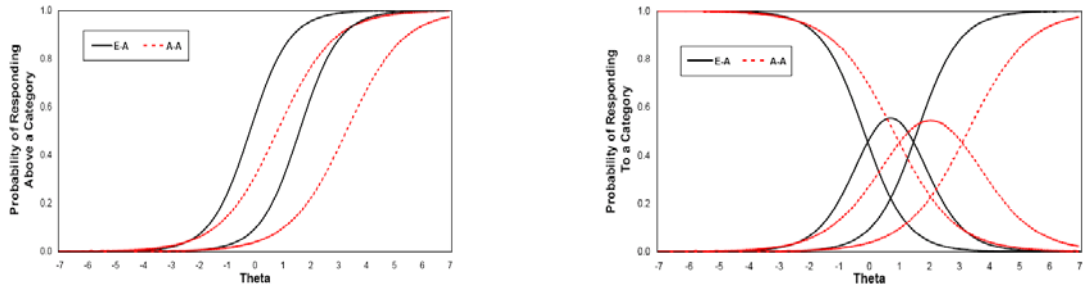
Item #8 "Can't concentrate, can't pay attention for long"



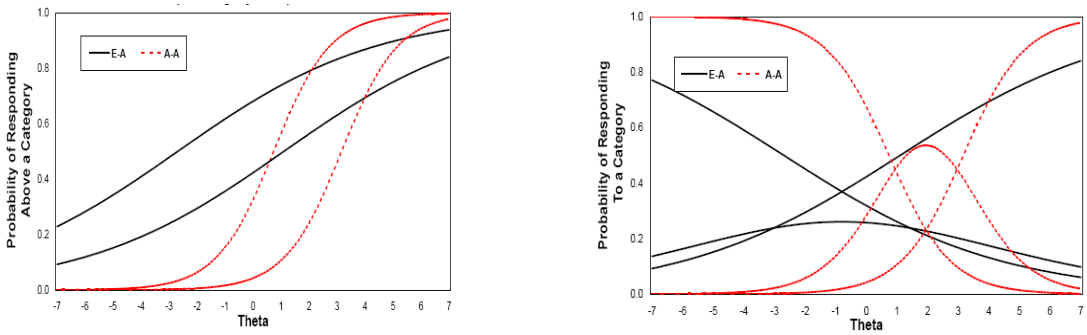
Item #13 "Confused or seems to be in a fog"



Item #41 "Impulsive or acts without thinking"



Item #45 "Nervous, high-strung, or tense"



Item #61 "Poor school work"

Figure 4. BRFs and CRFs for Attention Problems Scale

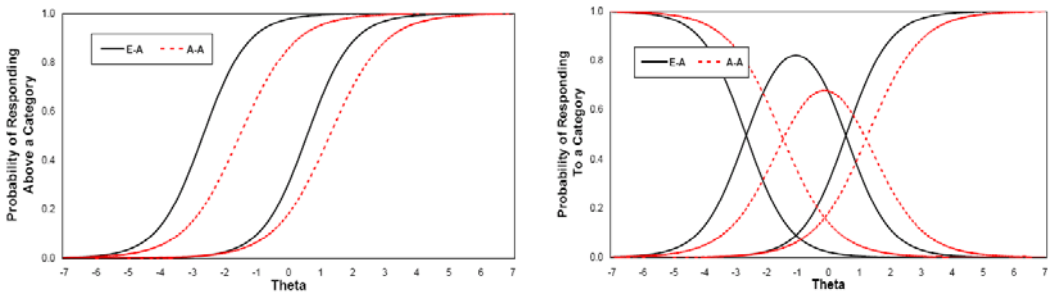
Aggressive Behavior

Nine of the 20 items (#3, #16, #20, #23, #57, #68, #86, #93, #95) were found to show significant differential functioning on the aggression scale. However, only two items contributed to scale level DIF (item #57 "Physically attacks" and item #16 "Cruelty, bullying, meanness to others." Both of these items showed very low discrimination (based on Baker's [2001] guidelines) for African-Americans, indicating that they were not related to the latent construct defined by the "Aggressive Behavior." In contrast, these two items exhibit very high discrimination for European-Americans, suggesting that behaviors that reflect physically aggression may mean something different for African-Americans. With the exception of "Disobedient at school" the remaining items all had lower discrimination for African-Americans. In contrast, the

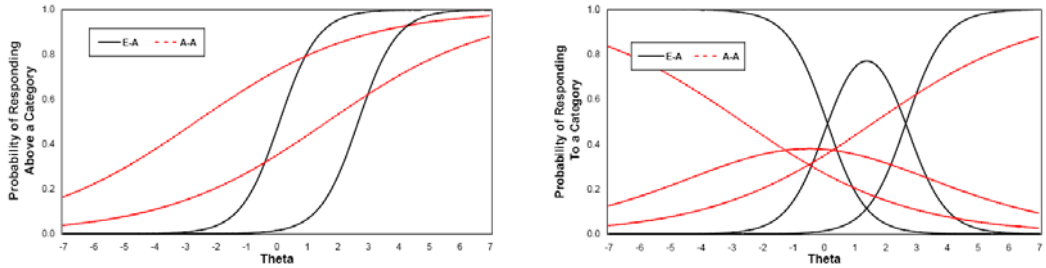
thresholds had mixed directions, with some items showing higher thresholds for African-Americans ("Argues a lot," "Screams a lot", "Stubborn, sullen, irritable," and item #95 "Temper tantrums"), whereas other items showed lower thresholds ("Destroys own things," "Disobedient at school," and "Talks too much"). Thus, African-Americans tended to require greater levels of aggression to endorse behaviors related to temper tantrums (e.g., items #3, #68, #86, and #95), whereas they had lower thresholds for items disobedience at school. In summary, the aggression scale exhibited lower discrimination for the African-Americans, particularly for the two largest DIF items, whereas threshold differences involved both directions.

Table 12. DFIT Results for Aggressive Behavior Scale

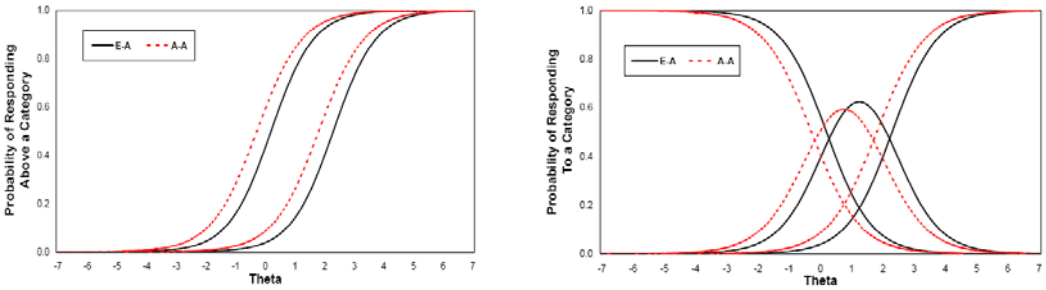
Item	Labels	Fa	Fb1	Fb2	Ra	Rb1	Rb2	CDIF	NCDIF
3	***Argues a lot***	**1.178**	**-1.53**	**1.275**	**1.427**	**-2.673**	**0.574**	**-0.325**	**.073***
7	*Bragging*	0.802	-0.47	3.077	0.683	-1.199	3.495	-0.126	0.012
16	***Cruelty***	**0.373**	**-2.64**	**1.658**	**1.585**	**0.085**	**2.669**	**0.705**	**.341***
19	*Demands attention*	1.035	-1.6	0.705	1.145	-1.595	1.003	0.068	0.003
20	***Destroy own things***	**1.307**	**-0.29**	**1.797**	**1.394**	**0.192**	**2.292**	**0.224**	**.035***
21	*Destroys others things*	1.646	0.018	2.005	1.383	0.379	2.553	0.14	0.019
22	*disobedient home*	1.376	-1.32	1.997	1.658	-1.795	1.541	-0.152	0.018
23	***disobedient at school***	**1.249**	**-1.2**	**2.119**	**0.998**	**-0.267**	**3.633**	**0.333**	**.077***
27	*Jealous*	0.861	-1.53	1.559	0.837	-1.462	1.999	0.077	0.005
37	*Fights*	1.085	0.408	2.992	1.611	0.382	2.562	0.049	0.005
57	***physically attacks***	**0.198**	**-2.13**	**0.949**	**1.708**	**1.117**	**3.597**	**1.051**	**.718***
68	***Screams a lot***	**1.139**	**0.804**	**3.073**	**1.257**	**0.063**	**2.177**	**-0.255**	**.053***
74	*Showing off*	1.184	-1.4	1.378	1.189	-1.941	1.401	-0.089	0.007
86	***Stubborn***	**1.144**	**-1.23**	**1.706**	**1.3**	**-1.963**	**1.143**	**-0.24**	**.039***
87	*Change in mood*	0.823	0.004	3.562	1.02	-0.164	2.983	-0.042	0.003
93	***talks too much***	**0.957**	**-1.47**	**0.746**	**0.868**	**-1.138**	**1.875**	**0.276**	**.055***
94	*teases a lot*	1.347	-0.46	2.03	1.22	-0.464	2.496	0.038	0.003
95	***Temper tantrums***	**1.263**	**-0.4**	**1.669**	**1.442**	**-1.203**	**1.231**	**-0.297**	**.058***
97	*threatens people*	1.447	1.186	3.529	1.841	0.985	3.145	-0.006	0.002
104	*Usually loud*	1.143	-0.53	1.704	1.337	-0.312	2.07	0.132	0.012



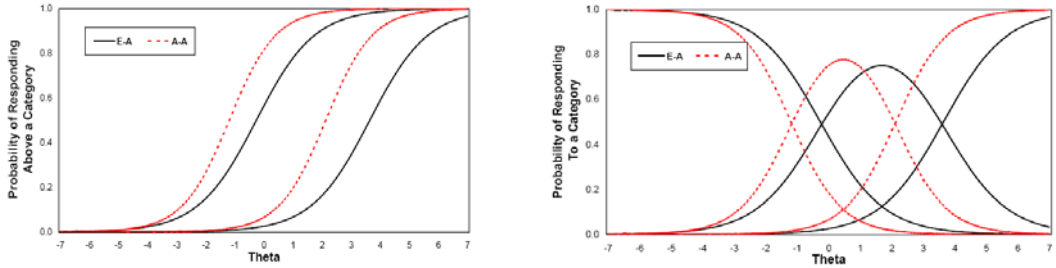
Item #3 “Argues a lot”



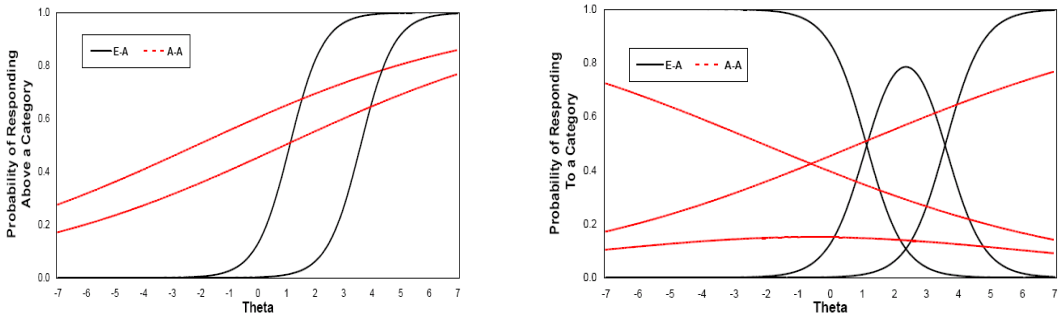
Item #16 “Cruelty, bullying, or meanness to others”



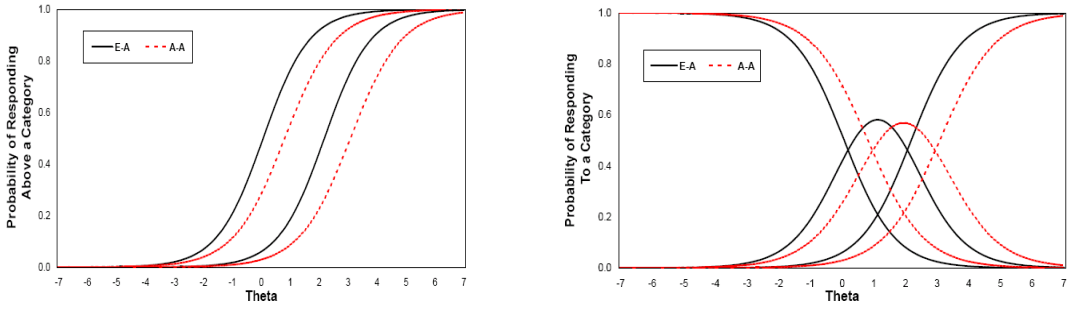
Item #20 “Destroys his/her own things”



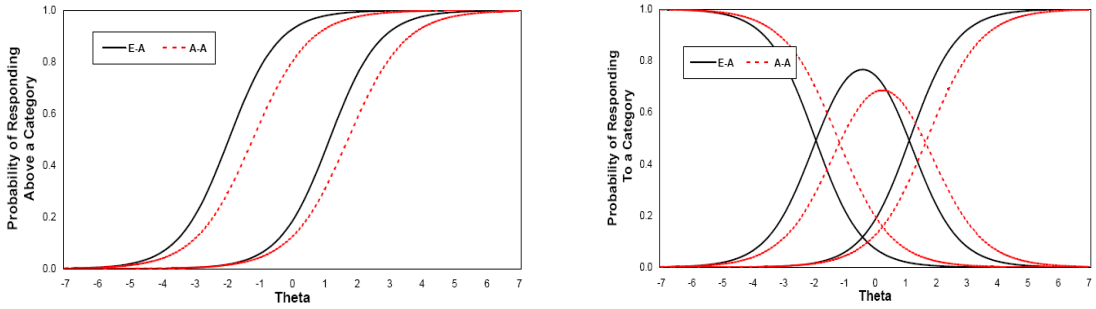
Item #23 “Disobedient at school”



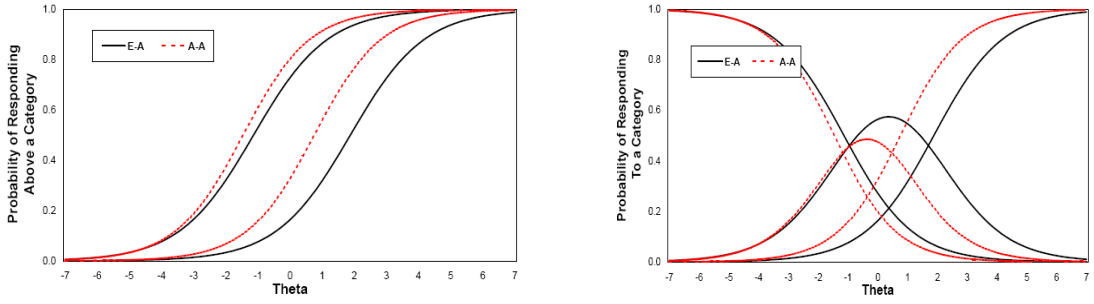
Item #57 “Physically attacks people”



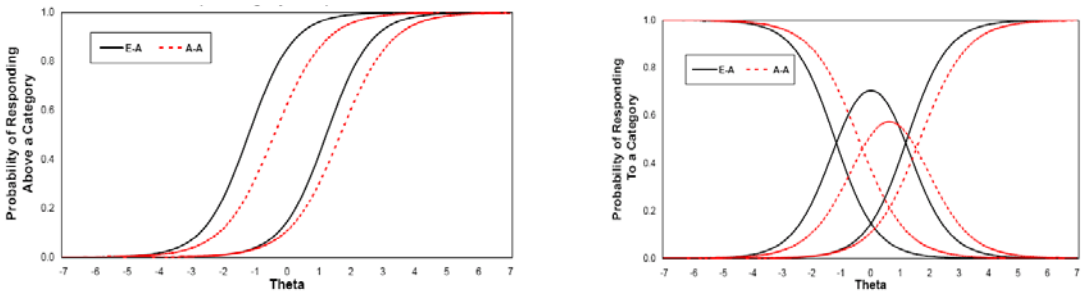
Item #68 “Screams a lot”



Item #86 “Stubborn, sullen, or irritable”



Item #93 “Talks too much”



Item #95 "Temper tantrums or hot temper"

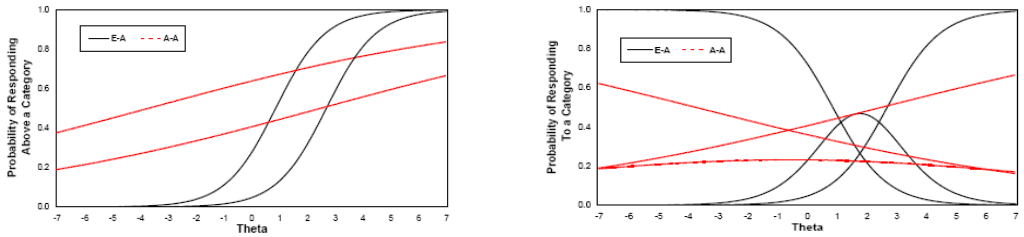
Figure 5. BRFs and CRFs for Aggressive Behavior Scale

Withdrawn / Depressed

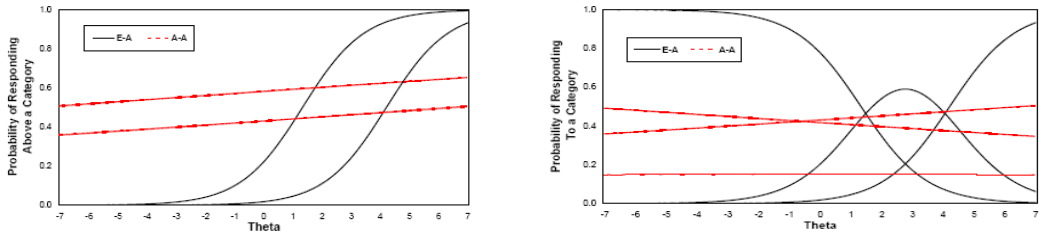
Four items exhibited DIF (items #69,# 80, #103, #111) on the Withdrawn scale. The two items with the largest magnitude (which also contributed to scale level DIF) were the items "Stares blankly" and "Secretive," both of which were very poor functioning items (i.e., provide little information about the latent Withdrawn factor) for African-Americans. Item #111 "Withdrawn" also exhibited lower discrimination (although still adequate) and higher thresholds for African-Americans. In contrast, the item "Unhappy, sad, or depressed" showed significantly higher discrimination for African-Americans. However, like item #111, it also required higher levels of the latent Withdrawn factor for the item "Unhappy, sad, or depressed" to be endorsed.

Table 13. DFIT Results for Withdrawn / Depressed Scale

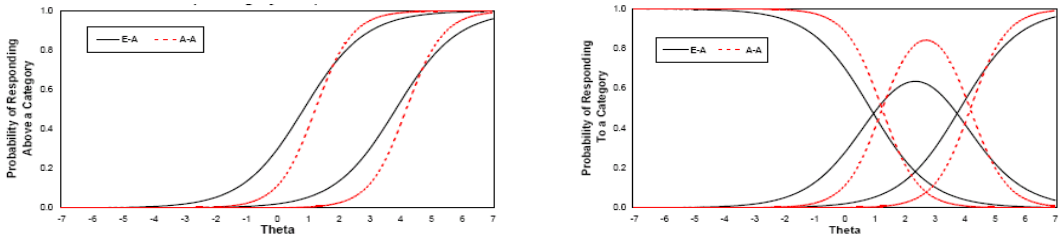
Item	labels	Fa	Fb1	Fb2	Ra	Rb1	Rb2	CDIF	NCDIF
42	*Prefers to be alone*	1.069	1.012	3.661	0.696	1.666	4.16	-0.036	0.004
65	*Refuses to talk*	1.08	1.065	3.533	1.066	0.873	3.127	-0.038	0.004
69	***Secretive***	**0.153**	**-3.726**	**2.464**	**1.117**	**0.853**	**2.675**	**0.778**	**.482***
75	*Shy*	0.827	0.084	3.544	0.736	0.234	3.131	-0.017	0
80	***Stares blankly***	**0.043**	**-7.737**	**6.472**	**0.942**	**1.326**	**4.205**	**0.822**	**.565***
88	*Sulks*	0.864	0.225	2.948	0.941	0.684	2.627	0.092	0.007
102	*Underactive*	1.543	1.818	4.417	0.684	2.651	4.843	-0.098	0.009
103	***Unhappy***	**1.669**	**1.231**	**4.18**	**1.005**	**0.849**	**3.839**	**-0.15**	**.027***
111	***Withdrawn***	**1.399**	**1.751**	**4.052**	**2.152**	**0.764**	**2.286**	**-0.042**	**.075***



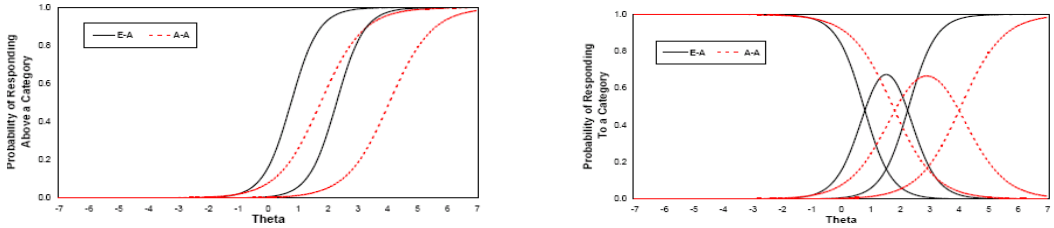
Item #69 "Secretive"



Item # 80 "Stares Blankly"



Item #103 "Unhappy, sad, or depressed"



Item #111 "Withdrawn"

Figure 6. BRFs and CRFs for Withdrawn Scale

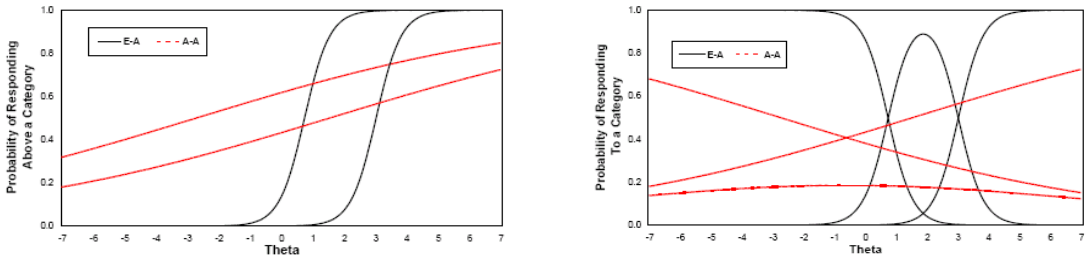
Social Problems

Only one item "Not liked by other kids" exhibited DIF, with this item showing poor functioning for African-Americans but good functioning for European-Americans.

However, it is important to note that many of the other items showed poor functioning across both groups (items #1, #55, #62, #64) but did not exhibit DIF. For African Americans, social problems appear to be defined by immaturity whereas for European Americans, social rejection appears to be more central to this construct.

Table 14. DFIT Results for Social Problems Scale

Item	labels	Fa	Fb1	Fb2	Ra	Rb1	Rb2	CDIF	NCDIF
1	*Acts too young*	0.596	0.471	4.387	0.696	-0.018	3.757	-0.053	0.006
11	*Clings*	1.023	-0.548	1.819	0.734	-0.473	2.715	0.04	0.013
25	*Doesn't get along w/other kids*	0.706	0.033	5.08	1.12	-0.325	4.115	-0.04	0.007
38	*Gets teased*	1.336	-0.355	2.092	1.288	-0.157	2.19	0.045	0.003
48	***Not liked by other kids***	**0.177**	**-2.737**	**1.519**	**2.482**	**0.725**	**3.007**	**0.694**	**.650***
55	*Overweight*	0.055	-1.901	1.901	0.134	-1.444	-0.389	-0.044	0.005
62	*Clumsy*	0.186	-2.889	1.898	0.028	-16.902	15.243	0.013	0.007
64	*Prefers younger kids*	0.431	-0.543	6.018	0.507	0.127	6.385	0.088	0.011



Item #48 "Not liked by other kids"

Figure 7. BRFs and CRFs for Social Problems Scale

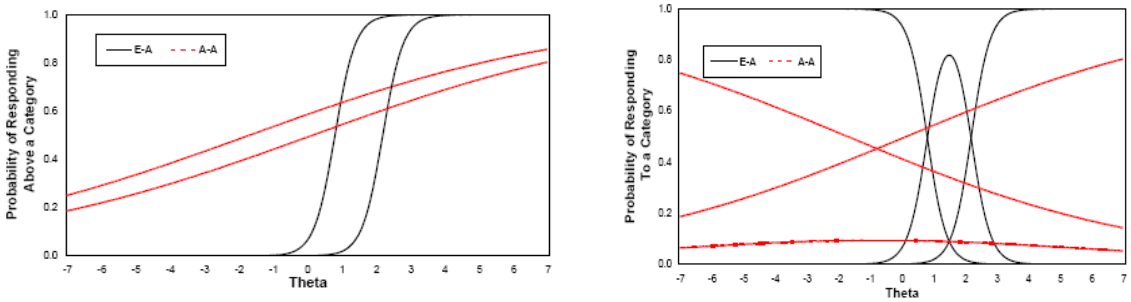
Somatic Problems

Two items were found to show differential functioning (items #56a and #56d) on the Somatic Complaints scale; however, due to the opposing direction of the DIF on these two items, scale level DIF was non-significant. Specifically, the item "Aches and pains" showed high discrimination for Euro-Americans, particularly in the theta range slightly

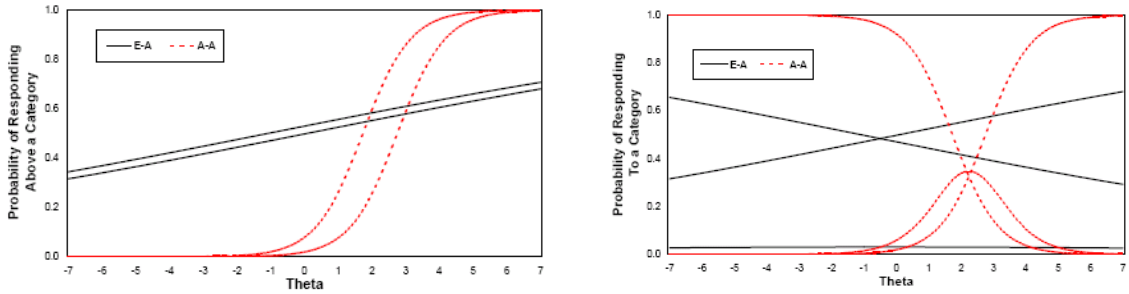
above the mean, whereas the item "Eye problems" showed higher discrimination for African-Americans, particularly slightly above the mean for theta (i.e., the level of the latent Somatic Problems factor). With the exception of "Headaches" and "Stomachaches" the remaining non-DIF items showed low discrimination for both groups, suggesting that they do not provide very much information about somatic problems.

Table 15. DFIT Results for Somatic Complaints Scale

Item	labels	Fa	Fb1	Fb2	Ra	Rb1	Rb2	CDIF	NCDIF
51	Dizzy	0.131	-1.351	-0.445	0.096	-0.906	-0.381	-0.003	0.001
54	Overtired	0.699	1.09	4.709	0.646	0.843	5.263	-0.001	0.002
56a	Aches	0.206	-1.708	0.13	3.316	0.788	2.18	0.047	.775*
56b	Headaches	2.305	0.747	2.599	2.263	0.779	2.702	-0.002	0
56c	Nausea	0.221	-1.537	-0.074	0.233	-1.556	0.588	0	0.001
56d	Eyeprob	1.422	1.731	2.752	0.109	-1.081	0.082	-0.018	.754*
56e	Skinprob	0.2	-1.73	0.629	0.191	-1.154	0.134	-0.001	0
56f	Stomachaches	1.95	0.751	2.648	2.562	0.625	2.279	0.013	0.005



Item #56a "Aches and pains"



Item #56d "Problems with eyes"

Figure 8. BRFs and CRFs for Somatic Complaints Scale

Person Differential Item Functioning

For each African-American participant, we computed (a) the estimated true score for each item based on the Euro-American parameters, (b) the estimated true score for each item based on the African-American parameters, and (c) the difference between these two true score estimates. The difference between these two estimated true or latent scores represents differential functioning at the person level, which we refer to as Person Differential Item Functioning (PDIF). Table 16 shows the means and standard deviations of the computed PDIF for significant DIF items. The average PDIF ranged from -.83 to +.83. Directionality of DPIF indicates whether the item true scores are under-estimated (DPIF is negative) or over-estimated (DPIF is positive) when African-Americans are estimated on the European-American parameters rather than their own parameters. That is, a positive PDIF indicates that use of the Euro-American parameters to estimate the item score for African-Americans leads to an over-estimate of the item score for African-Americans. Items with the largest mean PDIF include item #56 "Physically attacks (PDIF x=.-83)," item #103 "Unhappy, sad, or depressed (PDIF=.80)," "Fears he/she might do something bad (PDIF=.70)," "Suspicious (PDIF=.65)," " Poor school work (PDIF=.69)," "Cruelty (PDIF=-.56)," "Secretive (DPIF=-.65)", "Stares blankly (PDIF=-.73)," "Aches and pains (PDIF=.80), and "Eye problems (PDIF=-.83)."

Table 16. Means and Standard Deviations for CBCL PDIF (N=578)

	Items	Means	SD
Anxiety			
D14	*Cries a lot*	-.21	.06
D31	*Fears he/she might think or do something bad*	.70	.09
D89	*Suspicious*	.65	.15
D103	*Unhappy, sad, or depressed*	.80	.15
Attention			
D8	*Can't concentrate, can't pay attention for long*	.19	.07
D13	*Confused or seems to be in a fog*	.12	.13
D41	*Impulsive or acts without thinking*	.30	.08
D45	*Nervous, high-strung, or tense*	.30	.17
D61	*poor school work*	.69	.13
Aggressive			
D3	*Argues a lot*	.27	.04
D16	*Cruelty, bullying, or meanness to others*	-.56	.18
D20	*Destroys his/her own things*	-.18	.05
D23	*Disobedient at home*	-.27	.05
D57	*Physically attacks people*	-.83	.15
D68	*Screams a lot*	.21	.09
D86	*Stubborn, sullen, or irritable*	.20	.02
D93	*Talks too much*	-.23	.06
D95	*Temper tantrums or hot temper*	.24	.04
Withdrawn			
D69	*Secretive, keeps things to self*	-.65	.24
D80	*Stares blankly*	-.73	.19
D103	*Unhappy, sad, or depressed*	.16	.04
D111	*Withdrawn, doesn't get involved with others*	.17	.21
Social			
D48	*Not liked by other kids*	-.77	.25
Somatic			
D56a	*Aches or pains (not headaches)*	-.80	.37
D56d	*Problems with eyes*	.83	.24

The direction of the PDIF is inversely related to the CDIF. When the CDIF is negative, the PDIF is positive. This means that when African-Americans are scored on the European-American parameters, their estimated true scores on the latent factor will be

higher than when they are scored on their own parameters, suggesting that the differential functioning on this item is leading to an over-estimation of the level of pathology. Negative values on the PDIF suggest that the differential functioning on item is leading to an under-estimation of their true level of pathology. For example, the item "Physically attacks" (from the Aggression scale) shows, on average across African-Americans individuals, the largest magnitude of differential functioning (DPIF=.-.83). That is, when African-American children are scored on their own parameters they have a higher estimated true score on this item than when they are scored on the European-American parameters. Given that theta (i.e., the level of the latent factor) is a weighted sum of the estimated true item scores, use of the European-American parameters would lead to an under-estimate of the true level of the latent Aggression factor in African-Americans.

Analysis of Contextual Factors

The Fast Track study (the source of the present study's data) focused on the aggressive behavior and conduct problems, and hence selected contextual factors theoretically and empirically linked to aggression. Therefore, our discussion of contextual factors' influence on PDIF will focus primarily on the CBCL Aggressive Behavior scale, as our contextual factors are most closely theoretically linked to that scale. However, for the sake of completeness, we analyzed all of the CBCL scales included in this study, with the results of these analyses reported below.

In these analyses, it is important to consider precisely what the dependent variable, PDIF, represents conceptually. PDIF is the difference between estimated item true scores based on the European-American versus African-American parameters. This

difference represents "bias" in that item true scores estimated for African-Americans using their own parameters are unbiased (i.e., their expected value is theta) whereas those estimated using the European-American parameters represent theta + dif. It is important to reiterate that this is bias in the statistical sense, a directionally consistent difference between a parameter and its estimate, not bias in a social sense. If European-Americans were scored based on African-American parameters, the European-Americans' estimates undoubtedly also would be biased.

This bias results from factors exogenous to the latent construct operationalized by the CBCL scale influencing the observed item and / or scale scores differentially for the European- and African-Americans. In the present case, these exogenous factors can include social and cultural experiences, parent beliefs about psychopathology and 'normality,' parental values about what is desirable or undesirable in children, the meaning of "Somewhat True" (a response option on the CBCL), etc. By correlating contextual factors with this bias, operationalized through the PDIF, we hoped to identify possible exogenous factors underlying the bias. In the analyses below, the correlation analyses give estimates of the total effects of the contextual factors on PDIF, whereas the regression analyses give estimates of the unique (controlling for other contextual factors within the set) effects as well as effects of sets (e.g., parenting practices) of contextual factors.

Correlations between PDIF and Contextual Factors

We first computed Pearson correlations between PDIF and the various contextual factors, for those CBCL items showing significant DIF. These results are reported in

Table 17 below. With the exception of items #14, #103, #95, all PDIF variables were significantly related to family SES (r=.08 to .22). Although weaker than family SES, neighborhood level SES (as measured by Census indices) also showed significant associations with PDIF (range of r=.08 to .13) for all but items from the Anxious/Depressed scale. In addition, the percentage of residents in the neighborhood who are African-American (excluded from table) was also not related to the PDIF for any of the CBCL items.

Parental perceptions of the neighborhood also were related to PDIF. The magnitude of the correlations was largest for the Safety variable, particularly in relation to the PDIF for the CBCL aggression items (r = .-11 to.17). However, PDIF for CBCL item #95 ("Temper tantrums") was significantly related to none of the contextual variables except Parental Effectiveness (r=.12). Parent's perception of neighborhood social connectedness was only related to PDIF (r = -.09 to .09) for CBCL items from the Anxious / Depressed scale (except for the item "Cries a lot'). Parental perception of social services / resources was inconsistently associated with PDIF

Among the contextual factors, parenting and family factors were the most strongly and consistently associated with PDIF. In general, family positive expression was more strongly related to PDIF than family negative expression. Of the parenting / family variables, parent's perception of their punitiveness was most strongly related to PDIF (r = . -37 to .40) except for one item ("Hot temper" from the Aggressive Behavior Scale). Parental perception of their effectiveness was also strongly related (r = -.37 to .36) to the PDIF for items from of all of CBCL scales except the Somatic Complaints Scale. Parent's perception of their consistency in their parenting was less strongly related

to the PDIF variables (both in terms of magnitude of the correlations and the items with significant p-values). Parent's value for involvement and monitoring of their children's social life was only related to PDIF for items from the Anxious / Depressed scale (r=.08 to .11) and Withdrawn Scale (r = -.11). Finally, parent's value for self-defense in their children was related to PDIF for items in the Anxiety / Depression scale (r = -.09 to .09), the Aggressive Behavior scale (r = -.16 to .14), and the Withdrawn scale (r =.13).

Table 17. Correlation Matrix for DPIF and Contextual Variables

	SES		Neighborhood			Census – Neighborhood SES					Expressivity			Practices		Values	
Anx	**S1**	**S2**	**N1**	**N2**	**N3**	**C1**	**C2**	**C3**	**C4**	**C6**	**P1**	**P2**	**P3**	**P4**	**P5**	**P6**	**P7**
D14			.09*	.08*								-.10*	.21***	.16***	.22***	.09*	.08*
D31	.08*	.08*	-.09*	.09*	.08*							.11*	-.23***	-.18***	-.24***	-.09*	.09*
D89	.09*	.08*	-.08*	.11**	.10*							.15**	-.26***	-.23***	-.30***	-.08*	.11**
D103			-.08*	.09*	.09*							.11*	-.23***	-.19***	-.25*	-.08*	.09*
Att																	
D8	-.12**	-.12**			-.10**	.10**			.09*		-.16**	-.18***	.18***	.22***	.19***	.13**	
D13	-.13**	-.12**		-.11**	-.10**	.09*				-.09		-.15**	.13**	.31***	.33***		
D41	-.15***	-.15***			-.10*	.12**		.08*	.10**	-.10*		-.18***	.14**	.32***	.35***		
D45	-.12**	-.12**		-.12**	-.10*	.08*				-.09*		-.14**	.14**	.29***	.32***		
D61	.08*			.12**										-.22***	-.29***		
Agg																	
D3	.12**	.13**		.10*	.14***	-.10*		-.09*	-.09*			.14**	-.17***	-.35***	-.33***		
D16	-.16***	-.16***		-.08*	-.16***	.13**	.10*	.12**	.08*			-.14**	.17***	.34***	.40***	.14**	
D20	.09*	.10*		.10*	.13**							.13**	-.19***	-.32***	-.33***		
D23	.12**	.12**			.13**	-.10**		-.10*	-.09*			.15**	-.18***	-.37***	-.37***		
D57	-.15***	-.13***			-.11**	.11**		.11**	.09*			-.10*		.27***	.32***	.10*	
D68	-.14***	-.15***		.10*	-.17***	.11**	.07*	.10**	.08*			-.15**	.21***	.36***	.40*	.13*	
D86	.12**	.13**			.16***	-.09*		-.09*					-.14**	-.19***	-.22***	-.16***	
D93	.12**	.13**			.15***	-.10*		-.08*					-.21***	-.34***	-.37***	-.11*	
D95														.12*			
With																	
D69	-.22***	-.21***			-.12**	.12**	.12**	.12**	.10*		-.13**	-.17***	.27***	.17***	.23***	.13**	-.11*
D80	-.22***	-.21***			-.12**	.12**	.12**	.12**	.10*		-.13**	-.17***	.27***	.17***	.24***	.13**	-.11*
D103														.09*	.10*		
D111	-.22***	-.21***			-.14***	.12**	.13**	.12**	.10*			-.17***	.26***	.18***	.22***	.13**	-.11*
Soc																	
D48	-.12**	-.11**		-.10**	-.10**	.10*			.12**				.13**	.10*	.15**		
Soma																	
D56a	-.10**	-.09*		-.09*	-.09*	.08*			.12**				.12*		.13**		
D56d	.09*	.09*							-.11**				-.11*		-.10*		

Note: SI = SES Continuous Code, S2 = SES Higher Family Occupation Code, N1 = Neighborhood Social Connectedness, N2 = Neighborhood Public Services, N3 = Neighborhood Safety, C1 = % poverty, C2 = % public assistance, C3 = % female headed household, C4 = % unemployment, C6 = % no moves in 5 years, P1= Negative Expression, P2 = Positive Expression, P3 = Consistency, P4 = Effectiveness, P5 = Punitiveness, P6 = Self Defense, and P7 = Social Involvement. C5 (% Black) was not included in this table due to non-significant correlations.
P<.05*, P<.01**, P<.001**

Regression models for the Contextual Factors and PDIF

A series of multiple regressions were used to examine the unique variance of PDIF accounted for by predictors within each set of contextual factors. Six sets of analyses were completed for each item showing significant PDIF: (a) Family SES, (b) Neighborhood level SES (U.S. census), (c) Parent's perception of neighborhood qualities, (d) Family expressivity, (e) Parenting practices, and (f) Parenting values.

Results for the regression models of PDIF regressed separately on the six sets of contextual variables are presented in the tables below (with standardized betas). In order to restrict our interpretation of the most meaningful (in terms of effect size) relations, our interpretation focuses on items with a mean absolute value PDIF of .50 or greater; this includes CBCL items #31, #89, #103, #61, #16, #57, #69, #80, #56a, and #56d. Results for the complete set of regression analyses can be found in the Appendix.

Anxiety/Depression

Table 18. Regression Models for Neighborhood SES and PDIF for CBCL items from the Anxious / Depressed scale (N=568)

CBCL	Item		R^2	*Beta*	*F(df), p*
D31	*Fears he/she might think or do something bad*	Model	.02		2.37 (6, 561)*
		POVERTY		.14	.94
		PUBLIC		-.12	-.21
		FEMALE		-.21	-1.60
		UNEMPLOY		-.18	-2.0*
		BLACK		.26	3.0**
		NO MOVES		-.00	-.04
D89	*Suspicious*	Model	.02		2.34 (6, 561)*
		POVERTY		.12	.78
		PUBLIC		-.01	-.11
		FEMALE		-.22	-1.64
		UNEMPLOY		-.15	-1.74

		BLACK	.25	2.92**
		NO MOVES	.01	.24
D103	*Unhappy, sad, or depressed*	Model	.03	2.61 (6, 561)*
		POVERTY	.14	.91
		PUBLIC	-.01	-.12
		FEMALE	-.23	-1.74
		UNEMPLOY	-.18	-2.03*
		BLACK	.27	3.17***
		NO MOVES	-.00	-.09

Notes: Poverty = % resident living below poverty, Public = % resident living on public assistance, Female = % resident with female headed households, Unemploy = % resident of unemployed, Black = % resident of African-American descent, No Moves = % resident without moves within 5 years. *=p<.05; **=p<.01; ***=p<.001

Family level SES was not significantly related to PDIF for any items in the CBCL Anxious / Depressed scale (see Appendix for model results). However, neighborhood level SES factors, as measured by census block data, as a model were significantly related to PDIF for items in the CBCL Anxious / Depressed scale (R^2 ranged from .02 to .03). Across all items within this scale (except item #89, "Suspicious"), the census block data factors (a) percentage of unemployed individuals in the neighborhood and (b) percentage of neighborhood residents who are African-Americans were related to PDIF for items on the Anxious / Depressed scale. Specifically, the greater the percentage of residents in a subjects' neighborhoods who were African-American, the larger the over-estimation of their latent anxiety / depression level. In contrast, the percentage of residents in the neighborhood who were unemployed was inversely associated with over-estimation of CBCL items #31 ("Fears might do something bad") and #103 ("Unhappy, sad, or depressed").

Table 19. Regression Models for Neighborhood Factors and PDIF for CBCL items from the Anxious/Depressed scale (N=568)

CBCL	Item		R^2	Beta	*F(df), p*
D31	Fears he/she might think or do something bad	Model	.02		3.92(3, 564)**
		SOC		-.09	-2.15*
		PUB		.08	1.96*
		SAFE		.06	1.44
D89	Suspicious	Model	.03		4.89(3, 564)**
		SOC		-.08	-1.94*
		PUB		.10	2.40*
		SAFE		.08	1.82
D103	Unhappy, sad, or depressed	Model	.02		3.78(3, 564)**
		SOC		-.08	-1.89
		PUB		.08	1.99*
		SAFE		.07	1.56

Notes: SOC = Neighborhood social connectedness, PUB = Neighborhood public service, SAFE = Neighborhood safety
*=p<.05; **=p<.01; ***=p<.001

Parental perceptions of neighborhood qualities also were significantly related to PDIF for items from the CBCL Anxious / Depressed scale (R^2= .02 to .03). Parental perception of social connectedness related negatively to PDIF and public service related positively PDIF for items from the Anxious / Depressed.

Table 20. Regression Models for Parenting Practices and PDIF for CBCL items from the Anxious / Depressed scale (N=427)

CBCL	Item		R^2	Beta	F(df), *p* t value, *p*
D31	Fears he/she might think or do something bad	Model	.08		12.70(3, 423)***
		CON		-.13	-2.46**
		EFF		-.08	-1.66
		PUN		-.16	-3.08**
D89	Suspicious	Model	.12		19.91(3,423)***
		CON		-.13	-.58**
		EFF		-.12	-2.43*
		PUN		-.21	-4.05***

D103	Unhappy, sad, or depressed	Model	.09		13.40(3, 423)***
		CON		-.13	-2.46**
		EFF		-.09	-1.80
		PUN		-.16	-3.15**

Notes: CON = parental consistency, EFF = parental effectiveness, PUN = parental punitiveness
*=p<.05; **=p<.01; ***=p<.001

Parenting Practices were significantly related to the over-estimation for items from the Anxious / Depressed scale for African-Americans (R^2 ranged from .08 to.12). Although both consistency and punitiveness related to PDIF (as well as parental perceptions of their parenting effectiveness, with CBCL Item #89), punitiveness contributed the most unique variance to the prediction of PDIF (beta ranged from -.16 to .21) for items from this scale. That is, for African-Americans, parental use of punitive discipline approaches was negatively associated with the over-estimation of the latent trait of anxiety/depression.

Table 21. Regression Models for Family Expression and PDIF for CBCL items from the Anxious/Depressed scale (N=429)

CBCL	Item		R^2	Beta	F(df), p
D31	Fears he/she might think or do something bad	Model	.01		2.89(2, 426)
		NEG		-.04	-.70
		POS		.14	2.26*
D89	Suspicious	Model	.02		4.81 (2, 426)**
		NEG		-.04	-.67
		POS		.17	2.82**
D103	Unhappy, sad, or depressed	Model	.01		2.75(2, 426)
		NEG		-.04	-.60
		POS		.13	2.17*

Notes: NEG=family negative expression, POS = family positive expression
*=p<.05; **=p<.01; ***=p<.001

Family Expression was only significantly related to PDIF for one item from the CBCL Anxious / Depressed scale, Item #89 ("Suspicious"), with model R^2 = .02, with positive expression positively related to PDIF in this item.

Table 22. Regression Models for Parenting Values and PDIF for CBCL items from the Anxious / Depressed scale (N=414)

CBCL	Item		R^2	Beta	F(df), p
D31	Fears he/she might think or do something bad	Model	.02		4.00(2, 411)*
		DEF		-.14	2.81**
		SOC		-.01	.29
D89	Suspicious	Model	.02		5.01(2, 411)**
		DEF		-.15	-3.14***
		SOC		.02	.37
D103	Unhappy, sad, or depressed	Model	.02		3.74(2, 411)*
		DEF		-.13	-2.69**
		SOC		.02	.43

Notes: DEF = parental value for self-defense, SOC = parental value for social involvement
*=p<.05; **=p<.01; ***=p<.001

Parenting values were significantly related to PDIF (R^2 = .02) for all items in the CBCL Anxious / Depressed scale, with parents' valuing the importance of self defense negatively related to the over-estimation of items related to anxiety / depression for African-Americans (betas range from -.15 to .13).

Aggressive Behavior

Two items (Items #16 and #57) on the Aggressive Behavior Scale showed PDIF with means greater than 0.5. Both of the items had negative mean PDIF (Item #16 x = -.56 and Item #57 x = -.83), indicating under-estimation of the item for African-Americans

when calibrated on European-American parameters. For both of these items, all models of contextual variables were related significantly to the under-estimation (R^2 ranged from .02 to .21). Contextual models for each of these items are presented separately below (see Table 23 and 24).

Table 23. Regression Models for Context Factors and PDIF for Item #16 (from the Aggressive Behavior scale)

D16	Cruelty, bullying, or meanness to others		**R^2**	**Beta**	**F(df), p**
	Family SES (N=578)				7.65 (2, 566)***
		Model	.03		
		SES		-.10	-.83
		FAMOC		-.06	-.50
	Neighborhood Factors (N=578)	Model	.03		6.01 (3, 564)***
		SOC		.05	1.18
		PUB		-.06	-1.48
		SAFE		-.15	-3.45***
	Census Neighborhood (N=568)	Model	.03		2.60 (6, 561)*
		POVERTY		.13	0.85
		PUBLIC		-.08	-0.89
		FEMALE		.25	1.86
		UNEMPLOY		-.03	-.37
		BLACK		-.18	-2.13*
		NO MOVES		.04	0.76
	Family Expression (N=429)	Model	.03		7.67 (2, 426)***
		NEG		.16	2.66**
		POS		-.23	-3.90***
	Parenting Practices (N=427)	Model	.21		37.56 (3, 423)***
		CON		.07	-1.07
		EFF		.25	5.43***
		PUN		.34	6.99***
	Parenting Values (N=414)	Model	.02		4.29 (2, 411)*
		DEF		.14	2.93**
		SOC		-.00	-0.06

Notes: *=p<.05; **=p<.01; ***=p<.001

For the item #16 ("Cruelty"), the strongest model (R^2 = .21) was for the parenting practices, with parental perception of their parenting effectiveness (beta = .25) and use of punitive discipline (beta = .34) both having unique effects on the under-estimation of this item for African-American children. Predictors that were positively related to PDIF for item #16 ("Cruelty") included parental valuing of self-defense (beta =.14) and Negative Expression (beta = .16); in contrast, Positive Expression (beta = -.23), percentage of residents of African-American descent in neighborhood (beta = -.18), and parental perception of safety (beta = -.15) were negatively related to under-estimation of this item for African-American children.

Table 24. Regression Models for Context Factors and PDIF for Item #57 (from the Aggressive Behavior scale)

D57	Physically attacks people		***R²***	**Beta**	**F(df), p**
	Family SES (N=578)				5.74 (2,
		Model	.02		566)**
		SES		-.16	-1.26
		FAMOC		.02	-.14
	Neighborhood Factors				3.12 (3,
	(N=578)	Model	.02		564)***
		SOC		.06	1.34
		PUB		-.04	-.90
		SAFE		-.10	-2.34*
	Census Neighborhood	Model	.02		1.77(6, 561)*
	(N=568)	POVERTY		-.07	.18
		PUBLIC		.14	-.20
		FEMALE		-.25	1.53
		UNEMPLOY		-.05	.16
		BLACK		.18	.11
		NO MOVES		-.03	.32

Family Expression (N=429)	Model	.02		5.00 (2, 426)**
	NEG		.14	2.37**
	POS		-.19	-3.10***
Parenting Practices (N=427)	Model	.15		23.29 (3, 423)***
	CON		-.10	-2.04
	EFF		.22	4.46***
	PUN		.30	5.79***
Parenting Values (N=414)	Model	.01		2.09 (2, 411)*
	DEF		.10	2.02*
	SOC		.02	.34

Notes: *=p<.05; **=p<.01; ***=p<.001

Similar to pattern of Item #16, regression models were significant for all contextual factors (R^2 = .01 to .15), with the strongest model for parenting practices (R^2= .15). Again, parental use of punitive strategies (beta = .30) and parental perceptions of their parenting efficacy (beta = 4.46) showed unique contribution to PDIF for item #57 ("Physically attacks people"). Patterns of significant predictors, except for percentage of African-American residents in the neighborhood (which was insignificant in the neighborhood SES model), and associated direction of the betas were consistent with that of item #16.

Attention Problems

Although five items from this scale showed significant DIF, only one passed our PDIF criterion of 0.5 (item #61, "Poor school work," PDIF = .69). Only two context models, parental perception of neighborhood qualities and parenting practices (R^2 = .02 and .11 respectively) were significant for this item,. For the parenting model, both effectiveness (beta = -.17) and punitiveness (beta = -.28) contributed uniquely to the

prediction of PDIF for item #61. Specifically, parental perception of both their effectiveness and their punitiveness with their children were inversely related to the over-estimation of attention problems, while the perception of neighborhood public assistance (beta = .12) related positively.

Table 25. Regression Models for Context Factors and PDIF for Item #61(from Attention Problems scale)

D61	Poor school work		R^2	**Beta**	**F(df), p**
	Family SES (N=578)	Model	.01		2.13 (2, 566)
		SES		.13	1.07
		FAMOC		-.05	-.42
	Neighborhood Factors (N=578)	Model	.02		3.08 (3, 564)*
		SOC		-.00	-.10
		PUB		.12	2.78**
		SAFE		.03	.78
	Census Neighborhood (N=568)	Model	.01		.72(6, 561)
		POVERTY		.12	.75
		PUBLIC		-.01	-.11
		FEMALE		-.12	-.90
		UNEMPLOY		-.07	-.73
		BLACK		.06	.68
		NO MOVES		.07	1.24
	Family Expression (N=429)	Model	.01		2.00 (2, 426)
		NEG		-.08	-1.30
		POS		.12	2.00*
	Parenting Practices (N=427)	Model	.11		17.32 (3, 423)***
		CON		.10	1.89
		EFF		-.17	-3.41***
		PUN		-.28	-5.37***
	Parenting Values (N=414)	Model	.00		.83 (2, 411)
		DEF		-.06	-1.28
		SOC		.00	-.09

Notes: *=p<.05; **=p<.01; ***=p<.001

Withdrawn /Depressed

Tables 26 and 27 show results for the two CBCL items (#69 "Secretive" and #80 "Stares Blankly") from the Withdrawn / Depression subscale passing our PDIF criterion, with PDIF = -.65 and -.73, respectively. These two items tend to under-estimate level of withdrawal / depression for African-Americans. Consistent with the other PDIF models, parenting practices were the most predictive model (R^2= .10). However, in contrast to models for other CBCL scales and items, parental consistency contributed more uniquely to the variance of the bias than punitive strategies. Other predictors significantly contributing to PDIF included neighborhood safety (beta = -.11), Positive Expression in the family (beta = -.15 to -.14), parental valuing self defense (beta = .12 to .13) and social involvement (beta = -.10).

Table 26. Regression Models for Context and PDIF for Item #69 (from Withdrawn/Depressed scale)

D69	Secretive, keeps things to self		R^2	**Beta**	**F(df), p**
	Family SES (N=578)	Model	.05		14.96 (2, 566)***
		SES		-.22	-1.76
		FAMOC		-.01	-0.07
	Neighborhood	Model	.02		3.83 (3, 564)**
	Factors	SOC		.06	1.51
	(N=578)	PUB		-.04	-0.83
		SAFE		-.11	-2.66**
	Census Neighborhood	Model	.02		1.98 (6, 561)*
	(N=568)	POVERTY		-.06	-0.41
		PUBLIC		.08	0.89
		FEMALE		.16	1.18
		UNEMPLOY		.05	0.59
		BLACK		-.11	-1.27
		NO MOVES		-.01	-0.26
	Family Expression	Model	.03		6.54 (2, 426)**

	(N=429)	NEG	-.04	-.67
		POS	-.15	-2.43*
	Parenting Practices	Model	.10	14.85 (3, 423)***
	(N=427)	CON	.19	3.72***
		EFF	.06	1.26
		PUN	.13	2.59**
	Parenting Values	Model	.03	5.71 (2, 411)**
	(N=414)	DEF	.13	2.58**
		SOC	-.10	-2.14*

Notes: *=p<.05; **=p<.01; ***=p<.001

Table 27. Regression Models for Context and PDIF for Item #80 (from the Withdrawn/Depressed scale)

D80	Stares blankly		**R2**	**Beta**	**F(df), p**
	Family SES (N=578)	Model	.05		14.93 (2, 566)***
		SES		.21	-1.75
		FAMOC		.01	0.01
	Neighborhood	Model	.02		3.72 (3, 564)**
	Factors	SOC		.06	1.50
	(N=578)	PUB		-.03	-0.82
		SAFE		-.11	-2.61**
	Census Neighborhood	Model	.02		1.92 (6, 561)*
	(N=568)	POVERTY		-.06	-0.39
		PUBLIC		.08	.87
		FEMALE		.16	1.18
		UNEMPLOY		.05	.54
		BLACK		-.11	-1.25
		NO MOVES		-.01	-0.26
	Family Expression	Model	.03		6.50 (2, 426)**
	(N=429)	NEG		-.04	-.71
		POS		-.14	-2.40*
	Parenting Practices	Model	.10		14.97 (3, 423)***
	(N=427)	CON		.20	3.65***
		EFF		.07	-1.30
		PUN		.14	2.68**

Parenting Values (N=414)	Model	.03		5.65 (2, 411)**
	DEF		.12	2.56**
	SOC		-.10	-2.14*

Notes: *=p<.05; **=p<.01; ***=p<.001

Social Problems

The item #48, "Not liked by other kids" from the Social Problems scale exhibited large DIF (PDIF = .77) and over-estimated social problems for African-Americans compared with European-Americans. This item was significantly predicted by all context models (R^2 ranged from .04 to .07) except for family expression (R^2 = .01), with parenting practices as the strongest model (R^2 = 07). Similar to other PDIF models, both effectiveness (beta = .17) , punitiveness (beta=.17), and value for self defense (beta = .18) contributed positively to the over-estimation of social problems using this item. Other SES and neighborhood factors, such as the average SES for the family (beta = -.25), public services (beta = -.10), neighborhood safety (beta = -1.0), percentage of African-American residents (beta = -.17) related negatively to the difference in true scores on this item.

Table 28. Regression Models for Context Factors and PDIF for Item #48(from Social Problems scale)

D48	Not liked by other kids		**R^2**	**Beta**	**F(df), p**
	Family SES (N=578)				12.31(2,
		Model	.04		566)***
		SES		-.25	-2.0*
		FAMOC		.05	.37

Neighborhood Factors (N=578)	Model	.04		6.78 (3, 564)***
	SOC		.10	2.46
	PUB		-.10	-2.36*
	SAFE		-1.0	-2.45*
Census Neighborhood (N=568)	Model	.04		3.58(6, 561)**
	POVERTY		.12	.78
	PUBLIC		-.08	-.87
	FEMALE		.18	1.37
	UNEMPLOY		.08	.87
	BLACK		-.17	-2.02*
	NO MOVES		.03	.53
Family Expression (N=429)	Model	.01		2.15 (2, 426)
	NEG		.08	1.27
	POS		-.13	-2.07*
Parenting Practices (N=427)	Model	.07		10.64 (3, 423)***
	CON		-.01	-.23
	EFF		.17	3.31***
	PUN		.17	3.17**
Parenting Values (N=414)	Model	.04		7.69 (2, 411)***
	DEF		.19	3.85***
	SOC		-.03	-.66

Notes: *=p<.05; **=p<.01; ***=p<.001

Somatic Complaints

Tables 29 and 30 report results of the context models for items #56a and #56b from the Somatic Complaints scale. These two items showed large PDIF in opposite directions. Use of Euro-American parameters leads to under-estimation (PDIF = -.80) of item #56a ("Aches and pains") for African-Americans whereas item #56d ("Problems with eyes") is over-estimated (PDIF = .83) for African-American children when they were calibrated on European-American parameters. PDIF for Item #56a was related to

several contextual models, including family SES, neighborhood characteristics, and parenting practices; item #56d was unrelated to context factors except for neighborhood characteristics. None of the predictors within the model uniquely explained variance of the bias in these two items.

Table 29. Regression Models for Context Factors and PDIF for Item #56a (from Somatic Complaints scale)

D56a	Aches or pains (not headaches)		**R2**	**Beta**	**F(df), p**
	Family SES	Model	.01		3.00 (2, 569)*
		SES		-.12	-.99
		FAMOC		.02	.18
	Neighborhood Factors				3.63 (3,
		Model	.02		564)**
		SOC		.08	1.84
		PUB		-.08	-1.91
		SAFE		-.07	-1.58
	Census Neighborhood	Model	.02		1.59 (6, 561)
		POVERTY		.04	0.26
		PUBLIC		-.27	-0.29
		FEMALE		-.05	-0.38
		UNEMPLOY		.15	1.65
		BLACK		-.02	-0.18
		NO MOVES		.03	0.54
	Family Expression	Model	.00		.39 (2, 426)
		NEG		.01	.18
		POS		-.05	-.79
	Parenting Practices	Model	.02		3.19 (3, 423)*
		CON		.07	1.28
		EFF		.02	.34
		PUN		.10	1.81
	Parenting Values	Model	.00		0.82 (2, 411)
		DEF		.05	1.07
		SOC		-.03	-0.69

Notes: *=p<.05; **=p<.01; ***=p<.001

Table 30. Regression Models for Context Factors and PDIF for Item #56d (from the Somatic Complaints)

D56d	Problems with eyes		R2	Beta	F(df), p
	Family SES	Model	.01		2.24 (2, 569)
		SES		.07	.54
		FAMOC		.02	.17
	Neighborhood Factors	Model	.01		2.57 (3, 564)*
		SOC		-.07	-1.62
		PUB		.07	1.61
		SAFE		.05	1.26
	Census Neighborhood	Model	.02		1.64 (6, 561)
		POVERTY		-.07	-0.46
		PUBLIC		.06	0.65
		FEMALE		.04	0.31
		UNEMPLOY		-.14	-1.61
		BLACK		.04	0.45
		NO MOVES		-.05	-0.89
	Family Expression	Model	.00		0.43 (2, 426)
		NEG		-.02	-0.31
		POS		.05	0.88
	Parenting Practices	Model	.02		2.37 (3, 423)
		CON		-.08	-1.46
		EFF		-.03	-0.51
		PUN		-.06	-1.06
	Parenting Values	Model	.00		0.97 (2, 411)
		DEF		-.06	-1.14
		SOC		.04	0.78

Notes: *=p<.05; **=p<.01; ***=p<.001

CHAPTER V

DISCUSSION

In this discussion, we first provide an overview of the findings, and then review the patterns of the differential functioning found in the CBCL to provide insight into how the latent constructs differ across African-American and Euro-American groups. Then, we review findings from the contextual analyses, in order to develop hypotheses regarding the reasons for these differences in the latent constructs. Finally, we discuss limitations and make recommendations for direction of future research.

Overview of Findings

The results of our item response analysis of the CBCL support the hypothesis that CBCL items as well as some of its scales are not equivalent for African-American and Euro-American six year olds. In the present study, significant differential functioning (NCDIF >.024, raw score difference of .15 of a point on each item) was found for numerous (25) items when 6 year old African-Americans were scored on their parameters versus those of European-Americans. Significant differential functioning also was found for five of the six CBCL scales analyzed: Anxious / Depressed, Attention Problems, Aggressive Behavior, Social Problems, and Withdrawn / Depressed. The Somatic Complaints scale also showed significant differential item functioning, but DIF items operated in opposite directions canceling out scale level bias. Although the Thought Problems and Delinquent Behavior scales were not examined within the DFIT framework, both these scales showed group differences in eigenvalue ratios, suggesting

that there are differences in the underlying factor structure and measurement inequivalence.

In general, DIF can result from between-group differences in: (a) the discrimination parameter, (b) the threshold parameter(s), or (c) both sets of parameters. When extreme differences in the discrimination parameter occur, there also must be substantial differences in the thresholds, as the group with the lower slope (the lower discrimination parameter) will have more extreme thresholds; i.e., the distance between BRF1 and BRF2 will be greater due to the shallower slope of the discrimination parameter. Thus, the location of the BRFs is not highly meaningful when the discrimination parameter is below 0.5. In these instances, we focus our discussion on discrimination parameter differences.

To review, the discrimination parameter reflects the non-linear relation between the probability of the observed response and theta (level of the latent trait), as a function of varying levels of the latent trait. Threshold parameters provide information regarding how much of the latent trait is required before 50% of the sample endorse the item or response option. With the exception of the Anxious / Depressed scale, most CBCL items that showed significant DIF were less discriminating for African-Americans compared with European-Americans, suggesting that DIF items within these scales were less indicative of the psychopathology assessed by the scale. It perhaps is not surprising that in general CBCL items functioned better for the European-Americans, given that the large majority of subjects in the CBCL development samples were European-American.

For half of the CBCL scales analyzed (Anxious / Depressed, Attention Problems, Withdrawn / Depressed), item thresholds were higher for the African-Americans,

whereas for items from the Aggressive Behavior scale, African-Americans had lower as well as higher thresholds than the European-American children. Because the Social Problems and Somatic Complaints scales exhibited very poor discrimination for the African Americans or European Americans, it was not meaningful to compare the thresholds across groups.

When person level item bias (PDIF) was computed, African-Americans generally were under-estimated (when European-American parameters were used) on their scores for items from the Aggressive Behavior scale (particularly for physical aggression items), and the Withdrawn / Depressed and Social Problems scales. In contrast, true scores for items from the Attention Problems and Anxious / Depressed scales were over-estimated for African-Americans when they were estimated on European-American parameters. Because theta (the level of the latent trait) equals the weighted sum of the item true scores, if an item is under- or over-estimated, then this item will contribute to an under- or over-estimate (respectively) of theta, although other items with bias in the opposite direction may counteract this effect at the scale level (e.g., as in the Somatic Complaints scale).

This PDIF bias was significantly related to a variety of contextual factors such as family socio-economic status, neighborhood characteristics, family patterns for expressing feelings, and parental values. Parenting practices, in particular punitive discipline approaches and parents' beliefs about the effectiveness of their parenting, were most strongly related to the bias found in the CBCL scales. Thus, to sum in a single sentence, this study found that: (a) substantial African- versus European-American measurement bias exists in the CBCL at both the item and scale level, and (b) the bias is

due in part to socio-cultural differences between the African-American and European-American families.

Patterns of DIF

Below we provide a discussion regarding the specific patterns of items showing DIF in each scale. These patterns provide a description of the African- versus European-American differences in the latent constructs defined by the CBCL scales When appropriate, we suggest (but do not test) possible reasons for these differences.

In these interpretations, the discrimination parameter is taken to represent the magnitude of the relation between the item and the latent construct represented by the CBCL scale, using Baker's (2001) guidelines regarding subjective interpretation of discrimination parameters' magnitude. When discrimination parameters are above .5, threshold parameters are taken to indicate the extent to which the particular item is viewed as a serious sign of the latent trait represented by the CBCL factor, with higher thresholds indicating that the item is seen as a more serious sign or indicator of the latent construct. When discrimination parameters are below .5, threshold parameters are not interpreted, as the low discrimination indicates that the item is unrelated to the latent construct.

Anxious / Depressed. This scale showed the largest magnitude of differential functioning at the scale level, and 4 of the 14 items showed significant NCDIF. When African-Americans were calibrated on European-American parameters, the true scores for Items #31, #89, and #103 were over-estimated, ranging from .65 to .80 of a point – a substantial difference given that the item range is only 2. At a scale level, however, the

overall bias (2 points on a total scale range of 28 points) was less impressive because a number of other items tended to cancel this effect out.

It is important to note that although some researchers (e.g., Potenza & Dorans, 1995) have suggested that substantial item level differences (DIF) are not problematic as long as scale level differences (DTF) are not substantial (i.e., as long as biasing effects of items run in opposite directions, tending to cancel each other out at the scale level), we do not agree with this perspective. The content and meaning of the latent construct is defined by the relation of the items to theta, by the totality of the item parameters, and thus substantial item differences indicate that the latent construct is defined differently for the groups. Thus, substantial DIF with low DTF means that subjects with the same level of different latent traits are likely to receive the same observed score on the scale; this obviously is problematic.

For the most part, DIF for the Anxious / Depressed scale was a function of the difference between the discrimination parameters for the two groups (as opposed to being a function of differences in the threshold parameters). However, in contrast to items from the other CBCL scales analyzed in this study (discussed below), items with significant NCDIF on the Anxious / Depressed scale were more discriminating for African-Americans than for European-Americans (except for the item "Cries a lot"), indicating that the items provide more information about "anxiety / depression" (as operationalized by this scale) for the African-Americans. Two of three DIF items (item #31, "Fears doing bad things" and item #89 "Suspicious") appear to have in a common a reflection of a hyper-vigilant type of fear, suggesting that this may be more central to defining anxiety / depression for African-Americans than for European-Americans. In examining the

trend of the items that did not show significant NCDIF on this scale, we found that items with lower discrimination for African-Americans tended to be related to emotional insecurity (e.g., "Feels lonely," "Cries a lot," "Feels that no one loves them," "Fearful," and "Self-conscious"). Overall, these findings suggest that for African-Americans, anxiety / depression is defined more by a hyper-vigilant type of fear and less by emotional insecurity, relative to European-Americans. One might speculate that this reflects the reality of their cultural context, of the history of discrimination and poverty, and the need for vigilance in the face of these challenges.

The finding that the use of European-American parameters leads to an over-estimation for African-Americans of anxious / depressed items and the latent construct is somewhat inconsistent with the existing literature that suggests that African-Americans, particularly those living under conditions of "concentrated neighborhood disadvantage" (e.g., higher percentage of female headed households; greater percentage of residents unemployed and living below the poverty line) experience higher levels of anxiety than European-Americans (Sampson et al, 1997). In the PHDCN study, for instance, concentrated disadvantage was associated with more internalizing problems and a higher number of children in the clinical range for internalizing problems, even after accounting for family demographic characteristics, maternal depression, and earlier child mental health scores (Xue et al., 2005). However, although findings indicating higher symptom levels for African-Americans have been relatively consistent for symptoms directly involving fear (e.g., simple phobia; post traumatic stress disorder; Last & Perrin, 1993; Neal & Turner, 1991), they have been less consistent for more diffuse forms of anxiety such as generalized anxiety disorder (e.g., Austin & Chorpita, 2004; Lambert et al.,

2004). Given that the CBCL Anxious / Depressed scale emphasizes more GAD-type symptoms, our findings may not be that inconsistent with the literature. On the other hand, the fact that in the present sample raw scores on the Anxious / Depressed scale were significantly lower for the African-Americans than for the European-American subjects suggests that there may be sample specific characteristics, such as the young age of our subjects that influenced these finding.

Withdrawn / Depressed. Half of the items (#69, #80, #103, #111) on the Withdrawn / Depressed scale exhibited significant DIF, with 2 items (Item #69 "Secretive" and #80 "Stares blankly") contributing to scale level DIF (PDIF = -.65 and -.73, respectively). DIF for these two items as well as for #111 ("Withdrawn") was primarily a function of lower discrimination parameters for African-Americans, suggesting that these three items were relatively unrelated (relative to European-Americans) to "withdrawal / depression" as defined by the items on this scale. These items appear to have in common that they reflect a lack of interpersonal engagement. The fourth item showing significant DIF (#103, "Unhappy, sad, or depressed") had a higher discrimination parameter for the African-Americans (and in fact had the highest discrimination parameter of any item on the scale for the African-Americans), although it still showed acceptable discrimination for the European-Americans. Thus, overall, the Withdrawn / Depressed construct appears to focus more on the "depressed" and less on the "withdrawn" for the African-Americans as compared to the European-Americans.

Somatic Complaints. This scale produced two items (#56a "Aches and pains" and #56d " Problems with eyes") showing large DIF (PDIF = -.80 and .83, respectively) in opposing directions. Item #56a did not discriminate for African-American but did

discriminate well for European-American children whereas the opposite was found for Item #56d. This scale was mostly defined by two other items ("Headaches" and "Stomachaches"), both of which exhibited high discrimination (from 1.42 to 3.32). The remaining items ("Dizzy," "Overtired," "Nausea," and "Skin problems") showed poor functioning for both groups. The common defining factor for somatic complaints, across African-American and European-American children, was specific aches and pains (headaches, stomachaches), with somatic complaints among European-American children also more broadly reflecting aches and pains (item #56a, "Aches and pains") in general. For African-Americans, as well as headaches and stomachaches the specific aches and pains also included "Problems with eyes." It is possible that specific somatic complaints such as these receive more attention from African-American parents because these problems are perceived to be more valid than generic complaints about dizziness, being tired, or having general aches and pains, which could appear less serious to parents. It is important to note that in general, though, this scale did not function well for either group, and scale psychometric properties also were poor for both groups, with kurtosis highly elevated for African-Americans (kurtosis = 11.73).

Overall, these findings involving CBCL internalizing problems suggest that parents of African-American children may not notice or be less alarmed by depressive symptoms (compared to European-American parents), as evidenced by higher thresholds for depressive symptoms (on both the Anxious / Depressed and Withdrawn / Depressed scales). On the other hand, African-American parents may be more sensitive to fear-based behaviors, and this may be adaptive given the high-risk environments in which many of the African-American families in the study live, wherein a response to fear /

anxiety requires more immediate attention in contrast to emotional insecurity and sadness,.

Aggressive Behaviors. The Aggressive Behavior scale showed a high percentage (45%) of items exhibiting significant NCDIF, which suggest that aggression is defined differently for African-Americans. The two items exhibiting the largest DIF ("Cruelty," and "Physically attacks", PDIF=-.56 and -.83, respectively) had very low discrimination parameters for the African-Americans whereas these two items were highly discriminating for the European-Americans. This suggests that behaviors involving deliberate harm to others, such as cruelty and physically attacking other people, are seen as less apart of the broader aggression syndrome by African-American parents. This does not imply, however, that African-American parents are less concerned about these behaviors but that, unlike European-American parents, they do not see them as part of the broader CBCL Aggressive Behavior syndrome.

DIF for other items showing significant DIF was mostly a function of differences in thresholds. A number of items involving verbal behavior (argues, screams, stubborn, temper tantrums items) had higher thresholds, which suggests that African-American parents may more concerned about issues involving verbal behavior (i.e., African-American parents may see their children who argue or scream as having more "aggression" than European-American parents whose children argue or scream); on the other hand, African-American parents may be less concerned about obedience outside the home, as Item #23 ("Disobedient at school") had a lower threshold.

Attention Problems. This scale showed numerous items (45%) that functioned in the low or very low level. Item #61 ("Poor school work") exhibited the largest magnitude

of DIF, with an over-estimation of the item true score (and overall attention problems) for African-Americans (PDIF = .69) when the European-American parameters were used. This was primarily due to the very low discrimination of this item for European-American children. Although only moderately discriminating for African-Americans, relative to European-Americans problems with school work were more strongly related to the broader attention problems syndrome for young African-American children. This suggests that at this relatively early school age, African-American children may be more sensitive to the effects of attentional problems on their academic performance, hence the closer connection between academic and the broader attentional problems syndrome. For item #61, both thresholds were higher for the African-Americans (Fb1-Rb1 = 3.42; Fb2-Rb2 = 2.06), which suggests that African-American parents see academic problems as being more serious vis-à-vis attentional problems than European-American parents.

Social Problems. Like the Attention Problems scale, this scale showed poor functioning for both groups, with 63% and 38% of discrimination parameters in the low to very low range for African-Americans and European-Americans respectively. The one item showing significant DIF was #48 ("Not liked by other kids") with Fa=.18 and Ra=2.48. This reflects the fact that for the European-Americans, the latent construct appears to reflect social unpopularity whereas for the African-Americans it appears to more reflect immaturity. However, overall, this appears to not be a good scale for 6 year old African-American children.

Relations between Contextual Factors and PDIF

The previous discussion of the DIF analyses provides information regarding differences in <u>how</u> the latent constructs underlying the CBCL scales are differently

defined for African-American and European-American parents. In the next sections, we discuss contextual factors that may be related to why these definitions may vary. In the contextual analyses, we related item "bias" to a variety of contextual factors. In this context, between-group "bias" means that an item has a different relation to the latent trait for the different groups. This in turns mean that the item is related to something other than the latent trait, some exogenous factor, more or less for the African-Americans than for the European-Americans. If the discrimination parameter is lower, then for the African-Americans the item is influenced more by the unknown exogenous factor(s) than it is for the European-Americans. If the discrimination parameter is higher, then for the African-Americans the item is influenced less by the unknown exogenous factor(s). The purpose of our contextual analyses is to try to identify what some of these exogenous factors might be.

As noted previously, the Fast Track study selected its research participants (who also are the source of the present study's data) based on their current levels aggressive behavior as well as based on their risk for future development of more severe conduct problems. Fast Track thus chose contextual factors theoretically and empirically linked to aggression. In our discussion of how contextual factors may influence differences in African-Americans and European-Americans definitions of child psychopathology, we therefore focus primarily on the CBCL Aggressive Behavior scale, as it is most closely linked theoretically to the contextual factors available. As a contrast, however, we also discuss relations between contextual factors and the Anxious / Depressed scale because (a) this scale, assesses internalizing problems, provides a strong contrast to the

externalizing problems assessed by the Aggressive Behavior scale, and (b) this scale showed the highest DTF (scale-level bias),

Aggressive Behavior

Because they represent the largest source of bias, we focus our discussion on the two CBCL items that showed the greatest magnitude of PDIF on the Aggressive Behavior scale (#16, "Cruelty" and #57, "Physically attacks", PDIF=-.56 and -.83, respectively). For both of these items, all six of the contextual factor regression models were significant, indicating that PDIF bias was broadly related to a range of distal (e.g., SES) and more proximal (e.g., parenting practices) factors. However, as Tables 23 and 24 indicate, bias was most strongly related to the parenting practices variables, with R^2 of .21 and .15 for items #16 and #57 respectively, whereas R^2 for other contextual factors models for these two CBCL items ranged from .01 to .03. A similar pattern is seen for the first-order correlations in Table 17. Overall, this suggests that the more distal neighborhood influences may be partly mediated through the more proximal parental attitudes and behaviors.

Considering the specific contextual factors, (a) perceptions of neighborhood safety, (b) parental beliefs regarding the value for self-defense for their children, (c) the proportion of neighborhood residents who were African-American, (d) parental openess to expression of negative feelings and (e) parental openess to expression of positive feelings, (f) parental beliefs about the effectiveness of their parenting, and (g) parental beliefs about the extent to which their discipline was punitive all showed significant beta in relation to PDIF bias in CBCL items focusing on deliberate harm to others (see Tables

23 and 24). This can be viewed from two perspectives. From a psychometric, IRT perspective, this suggests that these various factors may underlie the item bias between African- and European-Americans. From a substantive perspective, this suggests that parental perceptions regarding the extent to which they see deliberate harm to others as aggression (as defined by the CBCL Aggressive Behavior scale) varies as a function of the parents' perceptions of neighborhood safety, etc.

Some of these relations have intuitive interpretations. One might speculate, for instance, that the reason parental perceptions regarding neighborhood safety are linked to bias (i.e., differential relations between intentional harm to others and "aggression") is because for parents who view their neighborhoods as unsafe, intentional harm to others may not be aggressive or pathological or undesirable, but rather a protective mechanism for establishing ones reputation in the neighborhood. Parental beliefs regarding the value for self-defense for their children might be similarly linked to the bias, in that parents who value self-defense may be less inclined to see intentional aggression as undesirable.

Other of the contextual factor findings might be interpreted similarly, though less intuitively. The fact that higher self-ratings of the punitive nature of parents' discipline are associated with higher PDIF values might reflect parental perceptions that both punitiveness and intentional aggression are, under certain circumstances, justified and hence their child's intentional aggression might be seen less as "aggressive," less linked to the CBCL Aggressive Behavior scale. A similar case might be made for the fact that parental openness to the expression of negative feelings was positively associated with PDIF. Parental openness to the expression of positive feelings was negatively associated with PDIF perhaps because parents who see expression of positive feelings with their

children as something desirable may be more inclined to see cruelty and physically attacking others (the PDIF items) as part of the aggression syndrome because they are more focused on warmth in relationships. These interpretations are obviously post hoc and speculative, and would need to be tested in an a priori manner. They do, however, provide interesting hypotheses and demonstrate the utility of the PDIF approach.

Anxious / Depressed

As with the Aggressive Behavior scale, we focus our discussion of the Anxious / Depressed scale on the CBCL items that showed the greatest magnitude of PDIF. These include #31 (Fears s/he might think or do something bad), #89 (Suspicious), and #103 (Unhappy, sad, or depressed). As with the Aggressive Behavior scale PDIF items, all contextual factor models were significant for all three PDIF items on the Anxious / Depressed scale and also as with the Aggressive Behavior scale, the parenting factors models showed the largest R^2. However, whereas the parenting factor models for the Aggressive Behavior scale had a mean R^2 of .18, for the Anxious / Depressed PDIF items these models had a mean R^2 of slightly less than .10. It is perhaps not surprising that the models were stronger for the Aggressive Behavior scale, as the contextual factors were selected for their theoretical links to aggressive behavior, rather than to anxiety and depression.

Considering the specific contextual factors, for all three PDIF items from the Anxious / Depressed scale, (a) the proportion of neighborhood residents who were African-American, (b) parental openness to expression of positive feelings (but not, in contrast to the Aggressive Behavior scale, parental openness to expression of negative

feelings), (c) parental beliefs regarding the value for self-defense for their children, (d) parental perceptions of public services being available in the neighborhood (which was not related to PDIF in the Aggressive Behavior scale) and (e) parental perceptions regarding the punitiveness and (f) consistency of their discipline (also not related to PDIF in the Aggressive Behavior scale) showed significant beta in relation to PDIF bias (see Tables 20, 21, and 22). In addition, PDIF for items #31 and #89 showed significant relations, and item #103 showed a marginally significant relation, to parental perceptions of social connectedness (again, in contrast to findings for the Aggressive Behavior scale) and PDIF for item #89 showed a significant relation to parental perceptions regarding the effectiveness of their discipline. Thus, overall, there were some similarities and some differences in PDIF / contextual factor relations across the Aggressive Behavior and Anxious / Depressed scales.

In part because the contextual factors were not selected for their links to anxiety and depression, interpretations of these results are less obvious, but still possible. For instance, the fact that PDIF for the two "hyper-vigilant" PDIF items (#31 and #89) was negatively related to parental perceptions of social connectedness in the neighborhood might be a function of more socially connected parents tending to see "hyper-vigilance" as less necessary, thus resulting in a decreased tendency for these items to be biased (relative to the European-Americans) in their relation to anxiety and depression.

Limitations

There are a number of limitations that should be considered when interpreting the results of the present study. First, study participants were all 6 years old. This is a

strength in that it provided for a homogeneous sample without age variability, and the focus on a relatively young population allowed us to identify biasing factors as they may begin to impact at the critical transition from home to school. A drawback to the narrow age-range is, however, that it prevented us from testing for developmental trends. Also, because behavioral and emotional problems increase in prevalence and severity up to some point in middle childhood or adolescence, depending on the particular domain of child psychopathology (low prevalence and severity is why we were unable to analyze the CBCL Delinquent Behavior and Thought Problems scales), our results very well may not generalize to other developmental levels. For instance, we found that although all of the CBCL scales analyzed contained items showing significant DIF, a number of the CBCL scales performed poorly for both African-American and European-American children. This perhaps reflects the fact that in general, our age range may represent the lower developmental boundary for the types of child psychopathology assessed by the CBCL.

A second limitation is that this study was based only on parent reports. Although parent reports are accepted as valid informants, they represent only one perspective on childhood problems and research often indicates that parent, teacher, and child report are not entirely consistent. Another limitation is that our indicators of contextual factors also may be biased. That is, we applied our IRT analyses to determine the extent to which African-American and European-American parents responded differentially to the CBCL, but it is also possible that they responded differentially to the measures of the contextual factors. Whether and how this may have influenced our results is unclear.

In this study, the 1991 rather than the most current version of the CBCL was used, because collection of these data began prior to the release of the ASEBA. There are a number of reasons why research on the 1991 version is still important. First, there is extensive overlap between the 1991 CBCL version and the 2001 ASEBA, and thus, much of the findings for the 1991 version are probably applicable to the 2001 version. Second, there is a widespread extant empirical literature base that has used the 1991 syndromes, and research findings from this existing literature and those from studies published in the future from these data might be interpreted within the context of findings from the present study. Finally, there are a number of ongoing longitudinal research projects (e.g., Fast Track; PHDCN) that continue to use the 1991 version of the CBCL.

Another limitation of the present study is that because our participants were selected based on their current and projected levels of aggressive behavior and conduct problems, our contextual factors were selected for the theoretical and empirical links to these externalizing problems. Thus, our primary focus on contextual factors' relations to bias was restricted to the CBCL Aggressive Behavior scale. Other datasets with other contextual factors will provide interesting tests for bias assessment.

Finally, we should note that although use of our PDIF approach appears to be a promising approach for identifying contextual factors related to bias, it needs replication before it can more generally be applied. To the best of our knowledge this is the first time that PDIF, as defined here, has been used and although many of our findings do fit with the literature relating race differences and aggressive behavior, it will be important to try the approach under different circumstances.

Implications

To the best of our knowledge, this is the first study that has used an Item Response Theory framework to compare the relative validity of the Child Behavior Checklist for African-Americans and Euro-Americans, and few studies using other approaches have assessed the validity of the Child Behavior Checklist for African-Americans. Therefore, our finding that significant differential functioning exists in the CBCL based on an IRT approach provides the first step in understanding the difference in how psychopathology may be defined by parents for these two groups, as well as the extent to which the CBCL.

This reinforces the perspective that these psychopathology labels or factors are social constructs that are at least to some extent culturally defined (Kitayama & Markus, 1992; Marsella, 2000), in contrast to Achenbach's perspective that the childhood psychopathology syndromes measured by the CBCL are universal, in that the items that define each syndrome do not carry the same meaning for two of the groups we examined. Although we did not examine second order structural differences in child psychopathology, differences in the discrimination parameters (similar to item loadings in factor analysis) also suggest that we should not assume that the higher order factors that make up child behaviors are the same for African-Americans. These cross-cultural difference in the definition and consequently expression of psychopathology support existing research in the adult literature suggesting cross cultural differences of psychological disorders (Kuo, 1984). Although there is less research in the child literature that has examined cross-cultural differences in psychological constructs, this

study contributes to the growing body of evidence that highlights the importance of cultural context.

The DFIT framework provides useful DIF indices that allow for interpretation of how DIF may impact both items (square root of NCDIF) and scales (square root of DTF), both in terms of magnitude and direction. This is especially useful in interpreting mean level differences, so often used to infer differences in prevalence rates or levels of psychopathology. For example, in the Fast Track sample, a significant difference was found in Anxious / Depressed measure, with African Americans exhibiting, on average, 0.71 points lower (F value = -3.10, $p < .01$). than their European American counterparts. With an adjustment (based on the DTF), African American children would likely score higher (i.e., use of European-American parameters results in an under-estimation for African-Americans, reducing or eliminating the significant greater difference between African American and European American children in their raw scores. Thus, the detection of bias has implications for not only understanding how the constructs differ, but also for interpreting the existing literature on race differences. Furthermore, evidence regarding the differential thresholds and discrimination for items from other scales also suggests that that norms and clinical cut-offs based on Euro-American samples may not be appropriate for younger African-American children, as they are over-estimated on social problems, anxiety, and attention and under-estimated on aggression and withdrawal.

Identification of contextual factors underlying bias also provides information about the processes by which factors associated with race may impact on how and why people from different groups define certain behaviors as normative or pathological. The

disentangling of these processes may provide clarification for racial disparities in mental health outcomes (particularly for African-American) often noted in this country (McLoyd & Steinberg, 1998). Also, findings regarding which specific contextual factors matter most in understanding differences in the meaning of problem behaviors ultimately will help us design interventions that consider *how* parenting behaviors may impact parental notions of pathology, as well as, *why* they may have certain beliefs about what is normative or pathological behavior in their children. This sort of research is necessary for identifying mechanisms of change and formulating interventions that address parental beliefs that may have detrimental consequences (e.g., high school drop outs, incarceration, suicide, etc.) for their children in varying context.

Conclusion

Given the wide usage of the CBCL with diverse populations, it is critical to conduct analysis of differential item and scale functioning before conclusions about group differences are drawn, as these differences may reflect measurement bias rather than true difference between groups. Differential functioning analysis using Item Response Theory is a useful methodology to study statistical bias in psychological test and to determine the nature of detected DIF. In the examination of the various scales of the CBCL, DIF analysis allowed for a more precise examination of how various properties of an item differed, as well as the implications for these differences. Detection of DIF also allows for the reformulation of existing scales by providing indicators that explicitly reveal how each item contributes to the overall bias of the test, to possibly

eliminate these items and create scales that more consistently measure psychological constructs of interest for various subgroups.

The findings of this study provide additional information regarding the validity of the CBCL for African-American children. Specifically, the findings suggest that contextual factors related to race impact on the response behavior of the African-Americans. The difference in the relation of scores in the underlying psychopathology dimension indicates that the scores do not mean the same thing across African-Americans and Euro-Americans, calling into question the cross-cultural construct validity of the CBCL. Our finding that DIF exists and is explained by cultural context, provides evidence suggesting that ideas about psychopathology are likely more culturally specific rather than universal as assumed by our measurement approach.

This provides an interesting but not easy to resolve dilemma for assessment instruments such as the CBCL. In the first version of the CBCL (Achenbach & Edelbrock, 1983), separate factors were derived for different age by sex groups, reflecting the empirical reality of different factor structures for the different groups. The present results suggest a similar empirical reality for African-Americans and European-Americans. Yet the distinct syndromes for different ages and for males and females produced difficulties for the use of the CBCL in both clinical work and research, making comparisons of males and females, older and younger children, or even the same children across time difficult or impossible. As a consequence, the 1991 version of the CBCL developed a common set of factors across the age and gender groups, reflecting the practical reality of the need for syndromes that could be consistent across gender and age, albeit at the expense of some increased error in the accuracy of assessment. How to fit

African-American versus European-American differences into such a scheme is unclear, but certainly understanding the nature of the differences between groups is a necessary step in the process.

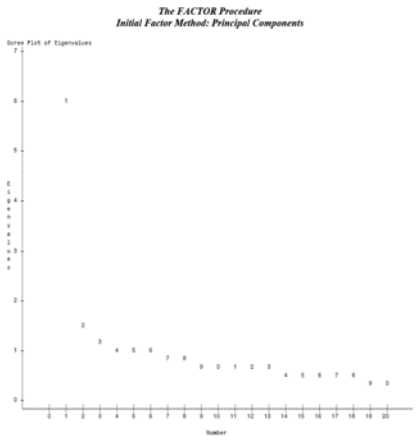
APPENDICES

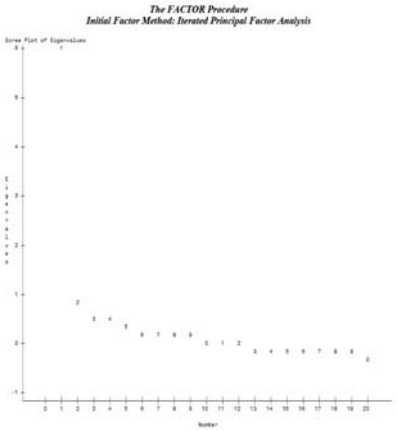
A. Scree Plots of CBCL Scales

A-A

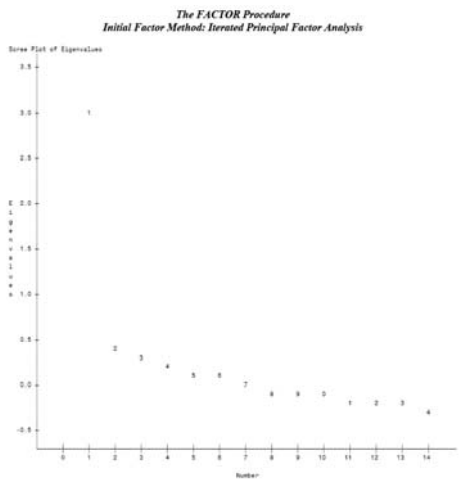
E-A

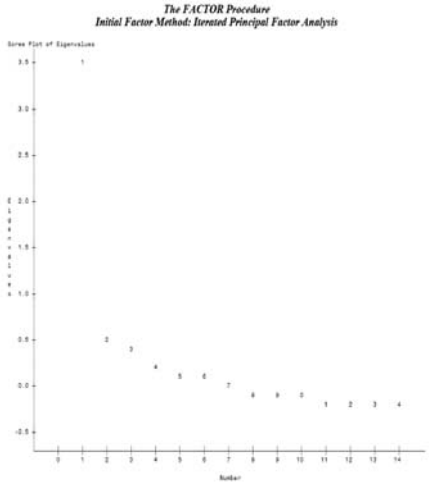
Aggressive Behavior





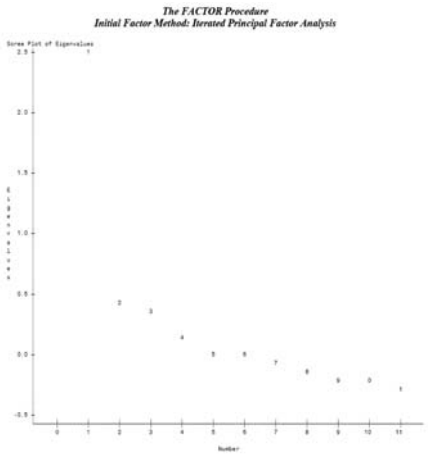
Anxious / Depressed

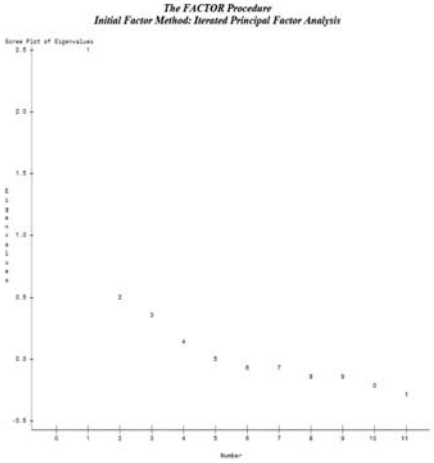




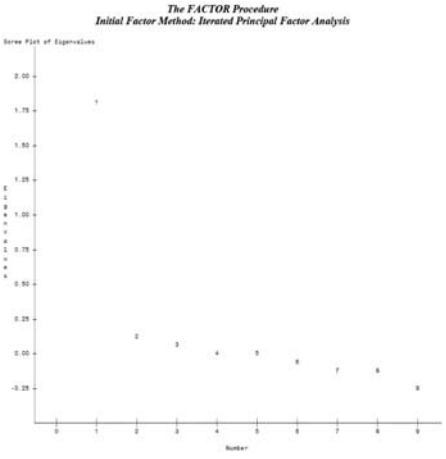
AA EA

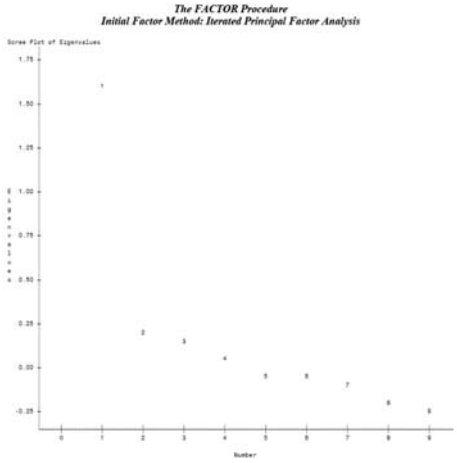
Attention Problems





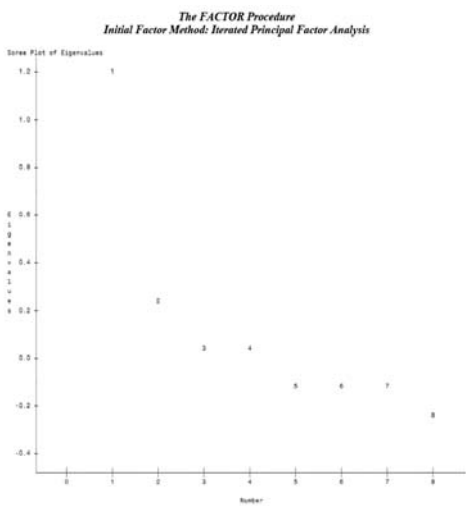
Withdrawn / Depressed

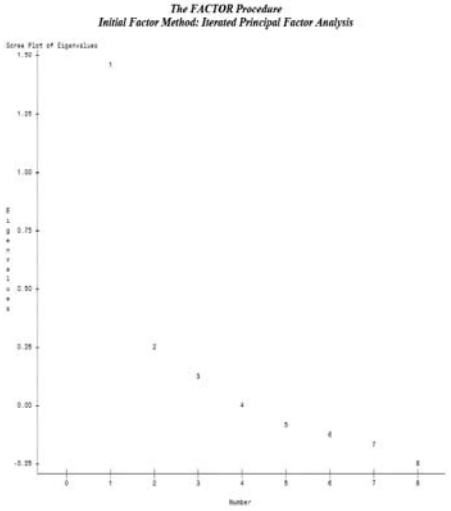




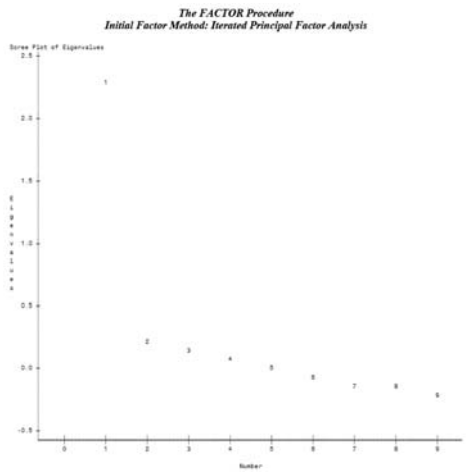
AA EA

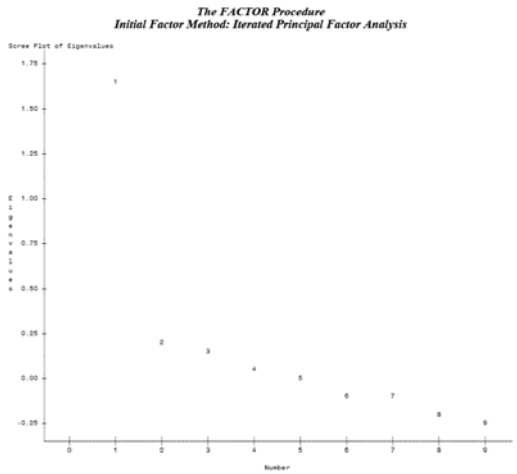
Social Problems





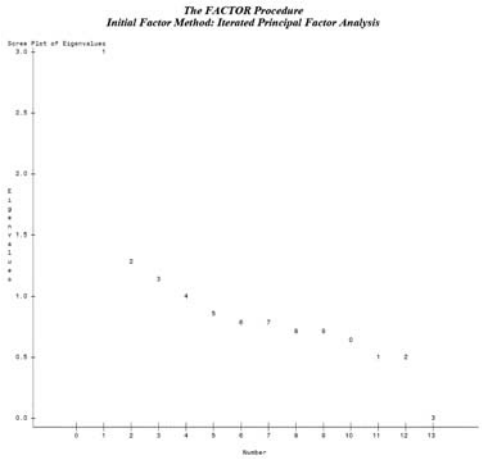
Somatic Complaints

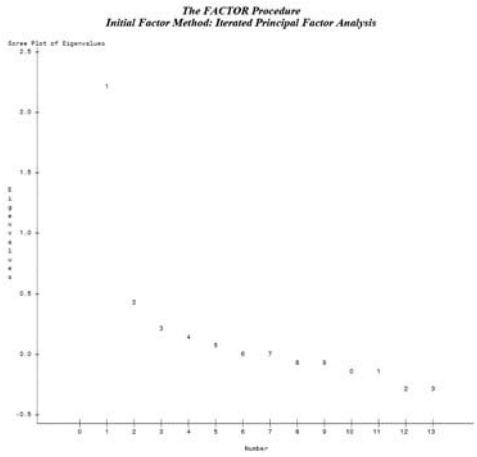




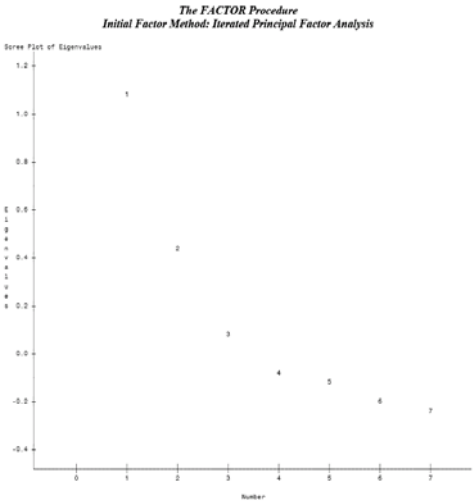
AA EA

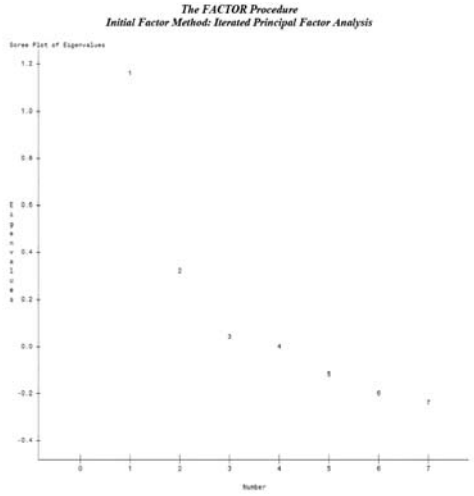
Delinquent Behavior





Thought Problems





B. Regression Models for Contextual Factors and PDIF

B1. Regression Models for Contextual Factors and D14 (Anxious / Depressed Scale) (N=569)

CBCL	Item		R2	Beta	F(df), p
D14	Cries a lot	Model	.01		1.87 (2, 566)
		SES		-.05	-.38
		FAMOC		-.03	-.27
D31	Fears he/she might think or do something bad	Model	.01		2.01 (2, 566)
		SES		.05	.69
		FAMOC		.03	.77
D89	Suspicious	Model	.01		2.13(2, 566)
		SES		.05	.39
		FAMOC		.04	.32
D103	Unhappy, sad, or depressed	Model	.01		1.71(2, 566)
		SES		.05	.41
		FAMOC		.03	.22

Notes: *=p<.05; **=p<.01; ***=p<.001

B2. Regression Models for Neighborhood Census Data and Anxious / Depressed (N=568)

CBCL	Item		R2	Beta	F(df), p
D14	Cries a lot	Model	.03		2.36(6, 561)*
		POVERTY		-.15	-.96
		PUBLIC		.02	.22
		FEMALE		.21	1.59
		UNEMPLOY		.18	2.06*
		BLACK		-.26	-3.01**
		NO MOVES		.00	.15
D31	Fears he/she might think or do something bad	Model	.02		2.37 (6, 561)*
		POVERTY		.14	.94
		PUBLIC		-.12	-.21
		FEMALE		-.21	-1.60
		UNEMPLOY		-.18	-2.0*
		BLACK		.26	3.0**
		NO MOVES		-.00	-.04
D89	Suspicious	Model	.02		2.34 (6, 561)*
		POVERTY		.12	.78
		PUBLIC		-.01	-.11
		FEMALE		-.22	-1.64
		UNEMPLOY		-.15	-1.74
		BLACK		.25	2.92**
		NO MOVES		.01	.24
D103	Unhappy, sad, or depressed	Model	.03		2.61 (6, 561)*
		POVERTY		.14	.91
		PUBLIC		-.01	-.12
		FEMALE		-.23	-1.74
		UNEMPLOY		-.18	-2.03*
		BLACK		.27	3.17***
		NO MOVES		-.00	-.09

Notes: *=p<.05; **=p<.01; ***=p<.001

B3. Regression Models for Neighborhood Factors and Anxious / Depressed (N=568)

CBCL	Item		R2	Beta	F(df), p
D14	Cries a lot	Model	.02		3.56 (3, 564)**
		SOC		.09	2.15*
		PUB		-.08	-1.80
		SAFE		-.06	-1.32
D31	Fears he/she might think or do something bad	Model	.02		.3.92(3, 564)**
		SOC		-.09	-2.15*
		PUB		.08	1.96*
		SAFE		.06	1.44
D89	Suspicious	Model	.03		4.89(3, 564)**
		SOC		-.08	-1.94*
		PUB		.10	2.40*
		SAFE		.08	1.82
D103	Unhappy, sad, or depressed	Model	.02		3.78(3, 564)**
		SOC		-.08	-1.89
		PUB		.08	1.99*
		SAFE		.07	1.56

Notes: *=p<.05; **=p<.01; ***=p<.001

B4. Regression Models for Family Expression and Anxious / Depressed (N=429)

CBCL	Item		R2	Beta	F(df), p
D14	Cries a lot	Model	.01		2.31 (2, 426)
		NEG		.04	.69
		POS		-.12	-2.04*
D31	Fears he/she might think or do something bad	Model	.01		2.89(2, 426)
		NEG		-.04	-.70
		POS		.14	2.26*
D89	Suspicious	Model	.02		4.81 (2, 426)**
		NEG		-.04	-.67
		POS		.17	2.82**
D103	Unhappy, sad, or depressed	Model	.01		2.75(2, 426)
		NEG		-.04	-.60
		POS		.13	2.17*

Notes: *=p<.05; **=p<.01; ***=p<.001

B5. Regression Models for Parenting Practices and Anxious / Depressed (N=427)

CBCL	Item		R2	Beta	F(df), *p* t Value, *p*
D14	Cries a lot	Model	.07		10.59(3, 423)***
		CON		.13	2.40*
		EFF		.07	1.42
		PUN		.14	2.73**
D31	Fears he/she might think or do something bad	Model	.08		12.70(3, 423)***
		CON		-.13	-2.46**
		EFF		-.08	-1.66
		PUN		-.16	-3.08**
D89	Suspicious	Model	.12		19.91(3,423)***
		CON		-.13	-.58**
		EFF		-.12	-2.43*
		PUN		-.21	-4.05***
D103	Unhappy, sad, or depressed	Model	.09		13.40(3, 423)***
		CON		-.13	-2.46**
		EFF		-.09	-1.80
		PUN		-.16	-3.15**

Notes: *=p<.05; **=p<.01; ***=p<.001

B**6. Regression Models for Parenting Values and Anxious / Depressed (N=414)**

CBCL	Item		R2	Beta	F(df), p
D14	Cries a lot	Model	.02		3.51 (2, 411)*
		DEF		.13	2.63**
		SOC		-.01	-.29
D31	Fears he/she might think or do something bad	Model	.02		4.00(2, 411)*
		DEF		-.14	2.81**
		SOC		-.01	.29
D89	Suspicious	Model	.02		5.01(2, 411)**
		DEF		-.15	-3.14***
		SOC		.02	.37
D103	Unhappy, sad, or depressed	Model	.02		3.74(2, 411)*
		DEF		-.13	-2.69**
		SOC		.02	.43

Notes: *=p<.05; **=p<.01; ***=p<.001

B7. Regression Models for SES and Withdrawn / Depressed (N=569)

CBCL	Item		R2	Beta	F(df), p
D69	Secretive, keeps things to self	Model	.05		14.96 (2, 566)***
		SES		-.22	-1.76
		FAMOC		-.01	-0.07
D80	Stares blankly				14.93 (2,
		Model	.05		566)***
		SES		.21	-1.75
		FAMOC		.01	0.01
D103	Unhappy, sad, or depressed	Model	.00		0.16 (2, 566)
		SES		-.06	-0.49
		FAMOC		.07	.55
D111	Withdrawn, doesn't get involved with others	Model	.05		14.60 (2, 566)***
		SES		-.23	-1.85
		FAMOC		.01	.05

Notes: *=p<.05; **=p<.01; ***=p<.001

B8. Regression Models for Neighborhood Disadvantage and Withdrawn (N=568)

CBCL	Item		R2	Beta	F(df), p
D69	Secretive, keeps things to self	Model	.02		1.98 (6, 561)*
		POVERTY		-.06	-0.41
		PUBLIC		.08	0.89
		FEMALE		.16	1.18
		UNEMPLOY		.05	0.59
		BLACK		-.11	-1.27
		NO MOVES		-.01	-0.26
D80	Stares blankly	Model	.02		1.92 (6, 561)*
		POVERTY		-.06	-0.39
		PUBLIC		.08	.87
		FEMALE		.16	1.18
		UNEMPLOY		.05	.54
		BLACK		-.11	-1.25
		NO MOVES		-.01	-0.26
D103	Unhappy, sad, or depressed	Model	.01		0.90 (6, 561)
		POVERTY		.12	.78
		PUBLIC		-.15	-1.58
		FEMALE		.10	.75
		UNEMPLOY		-.12	-1.34
		BLACK		-.03	-.40
		NO MOVES		.04	.70
D111	Withdrawn, doesn't get involved with others	Model	.02		2.13 (6, 561)*
		POVERTY		-.05	-0.32
		PUBLIC		.06	.67
		FEMALE		.19	1.44
		UNEMPLOY		.04	.50
		BLACK		-.13	-1.51
		NO MOVES		.00	.06

Notes: *=p<.05; **=p<.01; ***=p<.001

B9. Regression Models for Neighborhood Factors and Withdrawn (N=568)

CBCL	Item		R2	Beta	F(df), p
D69	Secretive, keeps things to self	Model	.02		3.83 (3, 564)**
		SOC		.06	1.51
		PUB		-.04	-0.83
		SAFE		-.11	-2.66**
D80	Stares blankly	Model	.02		3.72 (3, 564)**
		SOC		.06	1.50
		PUB		-.03	-0.82
		SAFE		-.11	-2.61**
D103	Unhappy, sad, or depressed	Model	.00		0.85 (3, 564)
		SOC		-.06	-1.53
		PUB		-.01	-0.33
		SAFE		-.01	-0.28
D111	Withdrawn, doesn't get involved with others	Model	.02		4.52 (3, 564)**
		SOC		.05	1.17
		PUB		-.04	-0.99
		SAFE		-.13	-3.08**

Notes: *=p<.05; **=p<.01; ***=p<.001

B10. Regression Models for Family Expression and Withdrawn (N=429)

CBCL	Item		R2	Beta	F(df), p
D69	Secretive, keeps things to self	Model	.03		6.54 (2, 426)**
		NEG		-.04	-.67
		POS		-.15	-2.43*
D80	Stares blankly	Model	.03		6.50 (2, 426)**
		NEG		-.04	-.71
		POS		-.14	-2.40*
D103	Unhappy, sad, or depressed	Model	.01		1.24 (2, 426)
		NEG		-.09	-1.37
		POS		.01	.21
D111	Withdrawn, doesn't get involved with others	Model	.03		6.87 (2, 426)***
		NEG		-.05	-0.82
		POS		-.14	-2.39*

Notes: *=p<.05; **=p<.01; ***=p<.001

B11. Regression Models for Parenting Practices and Withdrawn (N=427)

CBCL	Item		R2	Beta	F(df), *p* t Value, *p*
D69	Secretive, keeps things to self	Model	.10		14.85 (3, 423)***
		CON		.19	3.72***
		EFF		.06	1.26
		PUN		.13	2.59**
D80	Stares blankly	Model	.10		14.97 (3, 423)***
		CON		.20	3.65***
		EFF		.07	-1.30
		PUN		.14	2.68**
D103	Unhappy, sad, or depressed	Model	.02		2.92 (3,423)*
		CON		-.09	-.1.65
		EFF		.09	1.75
		PUN		.11	1.99*
D111	Withdrawn, doesn't get involved with others	Model	.09		13.93(3, 423)***
		CON		.19	3.58***
		EFF		.08	1.58
		PUN		.12	2.27*

Notes: *=p<.05; **=p<.01; ***=p<.001

B12. Regression Models for Parenting Values and Withdrawn (N=414)

CBCL	Item		R2	Beta	F(df), p
D69	Secretive, keeps things to self	Model	.03		5.71 (2, 411)**
		DEF		.13	2.58**
		SOC		-.10	-2.14*
D80	Stares blankly	Model	.03		5.65 (2, 411)**
		DEF		.12	2.56**
		SOC		-.10	-2.14*
D103	Unhappy, sad, or depressed	Model	.00		.11 (2, 411)
		DEF		-.00	-0.08
		SOC		-.02	-0.47
D111	Withdrawn, doesn't get involved with others	Model	.03		6.10 (2, 411)**
		DEF		.13	2.62**
		SOC		-.11	-2.27*

Notes: *=p<.05; **=p<.01; ***=p<.001

B13. Regression Models for SES and Aggressive Behavior (N=578)

CBCL	Item		R2	Beta	F(df), p
	Argues a lot				4.51 (2, 566)**
D3		Model	.02		
		SES		-.02	-.17
		FAMOC		.14	1.16
D16	Cruelty, bullying, or meanness to others	Model	.03		7.65 (2, 566)***
		SES		-.10	-.83
		FAMOC		-.06	-.50
D20	Destroys his/her own things	Model	.01		2.89(2, 566)
		SES		-.05	-.40
		FAMOC		.15	1.17
D23	Disobedient at home	Model	.01		4.30(2, 566)**
		SES		.01	.05
		FAMOC		.12	.94
D57	Physically attacks people	Model	.02		5.74(2, 566)**
		SES		-.16	-1.26
		FAMOC		.02	.14
D68	Screams a lot	Model	.02		6.37(2, 566)***
		SES		-.04	-.31
		FAMOC		-.11	-.89
D86	Stubborn, sullen, or irritable	Model	.02		4.62(2, 566)**
		SES		-.02	-.14
		FAMOC		.14	1.15
D93	Talks too much	Model	.02		4.70(2, 566)**
		SES		.00	.02
		FAMOC		.13	1.00
D95	Temper tantrums or hot temper	Model	.00		.90(2, 566)
		SES		.15	1.16
		FAMOC		-.11	-.87

Notes: *=p<.05; **=p<.01; ***=p<.005; ****=p<.001;

B13. Regression Models for Neighborhood Disadvantage and Aggressive Behaviors (N=568)

CBCL	Item		R2	Beta	F(df), p
D3	Argues a lot	Model	.02		1.71 (6, 561)
		POVERTY		-.08	-.54
		PUBLIC		.12	1.32
		FEMALE		-.21	-1.57
		UNEMPLOY		-.03	-.29
		BLACK		.14	1.58
		NO MOVES		-.02	-.38
D16	Cruelty, bullying, or meanness to others	Model	.03		2.60 (6, 561)*
		POVERTY		.13	.85
		PUBLIC		-.08	-.89
		FEMALE		.25	1.86
		UNEMPLOY		-.03	-.37
		BLACK		-.18	-2.13*
		NO MOVES		.04	.76
D20	Destroys his/her own things	Model	.02		1.43 (6, 561)
		POVERTY		-.16	-1.00
		PUBLIC		.16	1.67
		FEMALE		-.20	-1.47
		UNEMPLOY		.03	.31
		BLACK		.16	1.84
		NO MOVES		-.05	-.86
D23	Disobedient at home	Model	.02		2.13 (6, 561)*
		POVERTY		-.07	-.45
		PUBLIC		.14	1.51
		FEMALE		-.25	-1.91
		UNEMPLOY		-.05	-.59
		BLACK		.18	2.09*
		NO MOVES		-.03	-.62
D57	Physically attacks people	Model	.02		1.77 (6, 561)
		POVERTY		.03	.18
		PUBLIC		-.02	-.20
		FEMALE		.21	1.53
		UNEMPLOY		.02	.16
		BLACK		-.14	.11
		NO MOVES		.02	.32
D68	Screams a lot	Model	.03		2.50 (6, 561)*

		POVERTY	.17	1.09
		PUBLIC	-.13	-1.44
		FEMALE	.26	1.92
		UNEMPLOY	-.04	-.44
		BLACK	-.20	-2.30
		NO MOVES	.05	.98
D86	Stubborn, sullen, or irritable	Model	.01	1.27 (6, 561)
		POVERTY	-.20	-1.24
		PUBLIC	.06	.69
		FEMALE	-.10	-.76
		UNEMPLOY	.11	1.28
		BLACK	.07	.77
		NO MOVES	-.02	-.40
D93	Talks too much	Model	.02	2.05 (6, 561)
		POVERTY	-.18	-1.13
		PUBLIC	.14	1.52
		FEMALE	-.23	-1.72
		UNEMPLOY	.04	.49
		BLACK	.19	2.18
		NO MOVES	-.06	-.98
D95	Temper tantrums or hot temper	Model	.01	.47 (6, 561)
		POVERTY	.00	.05
		PUBLIC	-.09	-.95
		FEMALE	.00	.07
		UNEMPLOY	.05	.55
		BLACK	-.02	-.24
		NO MOVES	.00	.07

Notes: *=p<.05; **=p<.01; ***=p<.001

B14. Regression Models for Neighborhood Factors and Aggressive Behaviors (N=578)

CBCL	Item		R2	Beta	F(df), p
	Argues a lot				5.42 (3, 564)***
D3		Model	.03		
		SOC		.02	.60
		PUB		.08	1.83
		SAFE		.14	3.24***
D16	Cruelty, bullying, or meanness to others	Model	.03		6.01(3, 564)***
		SOC		.05	1.18
		PUB		-.06	-1.48
		SAFE		-.15	-3.45***
D20	Destroys his/her own things	Model	.02		4.69(3, 564)**
		SOC		.03	.76
		PUB		.07	1.73
		SAFE		.12	2.95**
D23	Disobedient at home	Model	.02		4.81(3, 564)**
		SOC		.02	.47
		PUB		.08	1.86
		SAFE		.13	2.96**
D57	Physically attacks people	Model	.02		3.12(2, 564)*
		SOC		.06	1.34
		PUB		-.04	-.90
		SAFE		-.10	-2.34*
D68	Screams a lot	Model	.03		6.43(3, 564)***
		SOC		.02	.39
		PUB		-.07	-1.78
		SAFE		-.15	-3.64***
D86	Stubborn, sullen, or irritable	Model	.03		5.50(3, 564)***
		SOC		-.03	-.66
		PUB		.05	1.18
		SAFE		.15	3.55***
D93	Talks too much	Model	.03		5.42(3, 564)***
		SOC		-.00	-.01
		PUB		.07	1.74
		SAFE		.14	3.31***

D95	Temper tantrums or hot temper	Model	.01		1.56(3, 564)
		SOC		-.08	-2.01*
		PUB		-.03	-.69
		SAFE		-.00	.04

Notes: *=p<.05; **=p<.01; ***=p<.001

B16. Regression Models for Family Expression and Aggressive Behaviors (N=429)

CBCL	Item		R2	Beta	F(df), p
D3	Argues a lot	Model	.02		4.51 (2, 426)**
		NEG		-.03	-.56
		POS		.16	2.69**
D16	Cruelty, bullying, or meanness to others	Model	.03		7.67(2, 426)***
		NEG		.16	2.66**
		POS		-.23	-3.90***
D20	Destroys his/her own things	Model	.02		4.36(2, 426)*
		NEG		-.06	-.95
		POS		.17	2.80**
D23	Disobedient at home	Model	.03		4.30(2, 426)**
		NEG		-.06	-.92
		POS		.19	3.10**
D57	Physically attacks people	Model	.02		5.00(2, 426)**
		NEG		.14	2.37*
		POS		-.19	-3.10**
D68	Screams a lot	Model	.03		7.25(2, 426)***
		NEG		.13	2.13*
		POS		-.23	-3.80***
D86	Stubborn, sullen, or irritable	Model	.02		2.25(2, 426)
		NEG		-.07	-1.23
		POS		.13	2.12*
D93	Talks too much	Model	.03		6.17(2, 426)**
		NEG		-.11	-1.85
		POS		.21	3.50***
D95	Temper tantrums or hot temper	Model	.01		1.22(2, 426)
		NEG		-.08	-1.27
		POS		.00	.04

Notes: *=p<.05; **=p<.01; ***=p<.005; ****=p<.001;

B17. Regression Models for Parenting Practices and Aggressive Behaviors (N=427)

CBCL	Item		R2	Beta	F(df), *p* T Value, *p*
D3	Argues a lot	Model	.18		30.30 (3, 423)***
		CON		.044	.71
		EFF		-.28	-5.81***
		PUN		-.26	-5.29***
D16	Cruelty, bullying, or meanness to others	Model	.21		37.56(3, 423)***
		CON		-.05	-1.07
		EFF		.25	5.43***
		PUN		.34	6.99***
D20	Destroys his/her own things	Model	.16		26.37(3,423)***
		CON		-.00	-.17
		EFF		-.24	-4.95***
		PUN		-.25	4.94***
D23	Disobedient at home	Model	.21		37.55(3, 423)***
		CON		.04	.77
		EFF		-.30	-6.30***
		PUN		-.30	-6.05***
D57	Physically attacks people	Model	.15		23.29(3, 423)***
		CON		-.10	-2.04
		EFF		.22	4.46***
		PUN		.30	5.79***
D68	Screams a lot	Model	.23		41.15(3, 423)***
		CON		-.02	-.43
		EFF		.27	5.84***
		PUN		.33	6.85***
D86	Stubborn, sullen, or irritable	Model	.07		10.01(3, 423)***
		CON		-.02	-.37
		EFF		-.13	-2.49**
		PUN		-.18	-3.38***
D93	Talks too much	Model	.03		34.97(3, 423)***
		CON		.00	.04
		EFF		-.25	-5.40***
		PUN		-.30	-6.11***
D95	Temper tantrums or hot temper	Model	.01		1.79(3, 423)
		CON		.04	.71
		EFF		.09	1.74
		PUN		.00	.07

Notes: *=p<.05; **=p<.01; ***=p<.001

B18. Regression Models for Parenting Values and Aggressive Behaviors (N=414)

CBCL	Item		R2	Beta	F(df), p
D3	Argues a lot	Model	.01		1.16 (2, 411)
		DEF		-.07	-1.46
		SOC		.02	.39
D16	Cruelty, bullying, or meanness to others	Model	.02		4.29(2, 411)*
		DEF		.14	2.93**
		SOC		-.00	-.06
D20	Destroys his/her own things	Model	.01		1.46(2, 411)
		DEF		-.08	-1.60
		SOC		.03	.59
D23	Disobedient at home	Model	.00		.90(2, 411)
		DEF		-.06	-1.21
		SOC		.03	.56
D57	Physically attacks people	Model	.01		2.09(2, 411)
		DEF		.10	2.02*
		SOC		.02	.34
D68	Screams a lot	Model	.02		3.53(2, 411)*
		DEF		.13	2.62**
		SOC		-.02	-.39
D86	Stubborn, sullen, or irritable	Model	.03		5.46(2, 411)**
		DEF		-.16	-3.30***
		SOC		.00	.06
D93	Talks too much	Model	.01		2.66(2, 411)
		DEF		-.11	-2.25*
		SOC		.02	.47
D95	Temper tantrums or hot temper	Model	.01		1.04(2, 411)
		DEF		-.06	-1.31
		SOC		-.03	-.61

Notes: *=p<.05; **=p<.01; ***=p<.001

REFERENCES

Abe Kim, J. & Takeuchi, D (1996). Cultural competence and quality of care: Issues for mental health service delivery in managed care. *Clinical Psychology: Science and Practice, 3*(4): 273-295.

Achenbach, T.M. (1991). *Manual for the Child Behavior Checklist/4-18 and 1991 Profile*. Burlington: University of Vermont, Department of Psychiatry.

Achenbach, T.M. (1995). Empirically based assessment and taxonomy: Applications to clinical research. *Psychological Assessment, 3,* 261-274.

Achenbach, T.M. & Edelbrock, C. (1983). *Manual for the Child Behavior Checklist and Revised Child Behavior Profile*. Burlington: University of Vermont, Department of Psychiatry.

Achenbach, T.M. & Edelbrock, C. (1981). Behavioral problems and competencies reported by parents of normal and disturbed children aged four to sixteen. *Monographs of the Society for Research in Child Development, 46* (Serial No. 188).

Achenbach, T.M. Verhulst, F.C., Baron, G.D. & Althaus, M. (1987). A comparison of syndromes derived from the Child Behavior Checklist for American and Dutch boys aged 6-11 and 12-16. *Journal of Child Psychology and Psychiatry, 28*, 437-453.

Allen, L. & Mitchell, C. (1998). Racial and ethnic differences in patterns of problematic and adaptive development: An epidemiological review. In V. McLoyd, V. & L. Steinberg, L (Eds.), *Studying minority adolescents : conceptual, methodological, and theoretical issue* (pp.29-54). Mahwah: Lawrence Erlbaum Associates.

Allen, J. & Walsh, James (2000). A construct-based approach to equivalence: Methodologies for cross-cultural/multicultural personality assessment research. In R.H. Dana (Ed.), *Handbook of Cross-Cultural and Multicultural Personality Assessment. Personality and clinical psychology series* (pp. 63-85). Mahwah: Lawrence Erlbaum Associates.

Angoff, W.H. (1982). Use of difficulty and discrimination indices for detecting item bias. In R.A. Berk (Ed.), *Handbook of methods for detecting test bias* (pp. 96-116). Baltimore: John Hopkins University Press.

Attar, B.K., Guerra, N.G. & Tolan, P.H. (1994). Neighborhood disadvantage, stressful life events, and adjustment in urban elementary-school children, *Journal of Clinical Child Psychology, 23*(4), 391-400.

Auerbach, J.G. & Lerner, Y. (1991). Syndromes derived from the Child Behavior Checklist for clinically referred Israeli boys aged 6-11: A research note. *Journal of Child Psychiatry, 32*(6), 1017-1024.

Austin, A.A. & Chorpita, B.F. Temperament, anxiety, and depression: Comparisons across five ethnic groups of children. Journal of Clinical Child and Adolescent Psychology, 33(2), 216-226.

Baker, Frank (2001). *The Basics of Item Response Theory.* ERIC Clearinghouse on Assessment and Evaluation, University of Maryland, College Park, MD.

Baker, F.B. (1995). *EQUATE 2.1: Computer program for equating two metrics in item response theory* [Computer program]. Madison: University of Wisconsin, Laboratory of Experimental Design.

Ben-Porath, Y. (1990). Cross-cultural assessment of personality: The case for replicatory factor analysis. In J.N. Butcher & C.D. Spielberger (Eds.), *Advances in personality assessment, Advances in personality assessment,* (Vol. 8, pp. 27-48). Hillsdale: Lawrence Erlbaum Associates.

Berk, R.A. (1982). *Handbook of Methods for Detecting Test Bias*. Baltimore: The John Hopkins University Press.

Betancourt, H. & Lopez, S. (1993). The study of culture, ethnicity, and race in American Psychology. *American Psychologist, 48*, 629-637.

Beyers, J.M., Bates, J.E., Pettit, G.S., Dodge, K.A. (2003). Neighborhood structure, parenting processes, and the development of youths' externalizing behaviors: A multilevel analysis. *American Journal of Community Psychology, 31*, 35-50.

Bradley, R., Corwyn, R., McAdoo, H., Garcia-Coll, C. (2001). The home environments of children in the United States Part I: Variations by age, ethnicity, and poverty status. *Child Development, 72*, 1844-1867.

Brody, G.H. et al. (2003). Neighborhood disadvantage moderates associations of parenting and older siblings problem attitudes and behavior with conduct disorders in African-American children. *Journal of Consulting and Clinical Psychology, 71* (a): 2111-222.

Brooks-Gunn, J., Duncan, G.J., & Aber, J.L. (1997). *Neighborhood poverty: Policy implications in studying neighborhoods*. New York: Russell Sage Foundation.

Bryk, A.S. & Raudenbush, S. W. (1992). Hierarchical linear models: Applications and data analysis methods. Newbury Park, CA: Sage.

Camilli, G. (1993). The case against item bias detection techniques based on internal criteria: do item bias procedures obscure test fairness issues? In P.W. Holland & H. Wainer (Eds.). *Differential Item Functioning*, (pp. 397-413). Hillsdale: Lawrence Erlbaum Associates.

Camilli, G.& Shepard, L.A. (1994). *Methods for Identifying Biased Test Items* (Vol. 4). Thousand Oaks: Sage Publications

Candell, G.L., & Drasgow, F. (1988). An iterative procedure for linking metrics and assessing item bias in item response theory. *Applied Psychological Measurement*, 12, 253-260.

Chatters, L.M., Taylor, R.J., Lincoln, K.D., & Schroefpfer, T. (2002). Patterns of informal support from family and church members among African-Americans. *Journal of Black Studies, 33*, 66-85.

Chernyshenko, O., Stark, S., Chan, K., Drasgow, F., & Williams, B. (2001). Fitting Item Response Theory Models to Two Personality Inventories: Issues and Insights. *Multivariate Behavioral Research, 36*(4), 523-562.

Chicchetti, D., & Toth, S.L., (1992). The role of developmental theory in prevention and intervention. *Development and Psychopathology, 4*, 489-493.

Children's Defense Fund. (1990). *Children 1990: A report card, briefing book, and action primer*. Washington, DC: Author.

Christenson, S.L. (1992). Review of the Child Behavior Checklist. In J.J. Kramer & J.C. Conoley (Eds.), *The Eleventh Mental Measurement Yearbook* (pp. 164-165). Lincoln: University of Nebraska.

Clark, L.A. & Watson, D. (1995). Constructing validity: Basic issues in objective scale development. *Psychological Assessment, 7*(3), 309-319.

Conduct Problems Prevention Research Group. (1992). A developmental and clinical model for the prevention of conduct disorders: The FAST Track Program. *Development and Psychopathology, 4*, 509-527.

Conduct Problems Prevention Research Group. (2002). The implementation of the Fast Track Program: An example of a large-scale prevention science efficacy trial. *Journal of Abnormal Child Psychology, 30*, 1-17.

Cronbach, L.J. & Meehl, P.E. (1955). Construct validity in psychological test. *Psychological Bulletin, 54,* 281-302.

Dana, R.H. (Ed.). (2000). *Handbook of Cross-Cultural and Multicultural Personality Assessment.* Mahwah: Lawrence Erlbaum Associates.

Deater-Deckard, K., Dodge, K.A., Bates, J.E., & Pettit, G.S. (1996). Physical discipline among African-American and European-American mothers: Links to children's externalizing behaviors. *Developmental Psychology, 32*, 1065-1-72.

Dedrick, R.F., Greenbaum, P.E., & Friedman, R.M. Wetherinton, C.M., & Knoff, H.M., (1997). Testing the structure of the Child Behavior Checklist/4-18 using confirmatory factor analysis. *Educational and Psychological Measurement, 57*, 306-313.

DeGroot, A., Koot, H.M.. & Verhulst (1994). Cross-cultural generalizability of the Child Behavior Checklist cross-informant syndromes. *Psychological Assessment, 6,* 225-230.

DeGroot, A., Koot, H.M.. & Verhulst (1995). Cross-cultural generalizability of the Youth Self Report and Teacher's Report Form cross-informant syndromes. *Journal of Abnormal Child Psychology, 24*, 651-664.

Dodge, K.A., Bates, J.E., & Pettit, G.S., (1990). Mechanisms in the cycle of violence. *Science, 250*, 1678-1683.

Drotar, D., Stein, R.E., Perrin, E.C. (1995). Methodological issues in using the Child Behavior Checklist and its related instruments in clinical child psychology research. *Journal of Clinical Child Psychology, 24*(2), 184-192.

Duncan, G. J. (1994). Families and neighbors as sources of disadvantage in the schooling decisions of White and Black adolescents. *American Journal of Education, 103*(1), 20-53.

Duncan, S., Stryker, L., Duncan, T., & Okut, H. (2002). A multilevel and contextual model of family conflict and deviance. *Journal of Psychopathology and Behavioral Assessment, 24*, 169-175.

DuRant, R. H., Cadenhead, C., Pendergrast, R. A., & Slavens, G. (1994). Factors associated with the use of violence among urban Black adolescents. *American Journal of Public Health, 84*(4), 612-617.

DuRant, R.H., Pendergrast, R.A., & Cadenhead., C. (1994). Exposure to violence and victimization and fighting behavior by urban black adolescents. *Journal of Adolescent Health, 15*, 311-318.

Embretson, S.E. (1996). The new rules of measurement. *Psychological Assessment, 8,* 341-349.

Embretson, S.E. & Herschenger, S. (1999). *The new rules of measurement: What every psychologist and educator should know.* Mahwah: Lawrence Erlbaum Associates.

Embretson, S.E. & Reise, S.P. (2000). *Item Response Theory for psychologists.* Mahwah: Lawrence Erlbaum Associates.

Emerson, E.N., Crowley, S.L., & Merrell, K.W. (1994). Convergent validity of the school social behavior scales with the Child Behavior Checklist and Teacher's Report Form. *Journal of Psychoeducational Assessment, 12*(4), 372-380.

Farley, R., & Frey, W. H. (1994). Changes in the segregation of Whites from Blacks during the 1980s: Small steps toward a more integrated society. *American Sociological Review, 59*(1), 23-45.

Fitzpatrick, K.M. & Boldizar, J.P. (1993). The prevalence and consequences of exposure to violence among African-American youth. Journal of the American Academy of Child and Adolescent Psychiatry, 32, 424-430.

Floyd, F.J. & Widaman, K.F. (1995). Factor analysis in the development and refinement of clinical assessment instruments. *Psychological Assessment, 7,* 286-299.

Foster, S.L. & Cone, J.D. (1995). *Validity issues in clinical assessment, Psychological Assessment, 7*(3), 248-260.

Gaines, S. O., Jr., Marelich, W. D., Bledsoe, K. L., & Steers, W. N. (1997). Links between race/ethnicity and cultural values as mediated by racial/ethnic identity and moderated by gender. *Journal of Personality and Social Psychology, 72*(6), 1460-1476.

Garcia Coll, C., & Garrido, M. (2000). Minorities in the United States: Sociocultural context for mental health and developmental psychopathology. In A. J. Sameroff (Ed.), *Handbook of developmental psychopathology* (2nd ed., pp. 177-195). Dordrecht, Netherlands: Kluwer Academic Publishers.

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

Gibbs, J.T., & Huang, L.N. (Eds.). (1998). *Children of color: Psychological interventions with culturally diverse youth.* San Francisco: Jossey-Bass Inc.

Giles Sims, J., Straus, M. A., & Sugarman, D. B. (1995). Child, maternal, and family characteristics associated with spanking. *Family Relations: Journal of Applied Family and Child Studies, 44*(2), 170-176.

Gorman-Smith, D., & Tolan, P. (1998). The role of exposure to community violence and developmental problems among inner-city youth. *Developmental Psychopathology, 10,* 101-116.

Graham, S. (1992). “Most of the subjects were White and middle class”: trends in published research on African-Americans in selected APA journals, 1970-1989. *American Psychologist, 47*(5): 629-639.

Greenbaum, P.E. & Dedrick, R.F. (1998). Hierarchical confirmatory factor analysis of the Child Behavior Checklist / 4-18. *Psychological Assessment, 10,* 149-155.

Greenberg, M & Lengua, L. (1995). Scale construction of the Neighborhood Questionnaire. (FAST Track Project Technical Reports). Seattle, WA: University of Washington.

Guerra, N.G., Huesmann, L.R., & Hanish, L. (1994). The role of normative beliefs in child’s social behavior. In N. Eisenberg (Ed.), *Social Development* (pp.140-158). Thousand Oaks: Sage.

Guerra, N.G., Huesmann, L.R., Tolan, P.H., Van Acker, R., & Leonard, D.E. (1995). Stressful events and individual beliefs as correlates of economic disadvantage and aggression among urban children, *Journal of Consulting and Clinical Psychology, 63*(4), 518-528.

Hall, N.G., Bansal, A., & Lopez, I.R. (1999). Ethnicity and psychopathology: A meta-analytic review of 31 years of comparative MMPI/MMPI-2 research. *Psychological-Assessment,* 11(2): 186-197.

Hambleton, R.K., Clauser, B.E., Mazor, K.M., & Jones, R.W (1993). Advances in the detection of differentially functioning test items. *European Journal of Psychological Assessment, 9*(1): 1-18.

Hambleton, R.K., Swaminathan, H., and Rogers, H.J. (1991). *Fundamentals of Item Response Theory*. Newbury Park: Sage.

Hammon, W.R. & Yung, B.R. (1991). Preventing violence in at-risk African-Americans youth. *Journal of Health Care for the Poor and Underserved, 2*, 1-16.

Hartman, C.A., Hox, J., Auerbach, J., Erol, N., Fonseca, A.C., Mellenbergh, G.J., Novik, T.S., Oosterlaan, J., Roussos, A.C., Shalev, R.S., & Ziber, N., & Sergeant, J.A. (1999). Syndrome dimensions of the child behavior checklist and the teacher report form: A critical empirical evaluation. *Journal of Child Psychology and Psychiatry, 40*(7) 1095-1116.

Heubeck, B. G. (2000). Cross-cultural generalizability of CBCL syndromes across three continents: From the U.S.A and Holland to Australia. *Journal of Abnormal Child Psychology, 28*, 439-450.

Hill, S.A. (1999). *African-American Children: Socialization and development in families.* Sage Publications: Thousand Oaks: CA.

Holland, P.W. & Wainer, H. (Eds.). (1993). *Differential Item Functioning*. Hillsdale: Lawrence Erlbaum Associates.

Hurd, E.P., Moore, C., & Rogers, R. (1995). Quiet success: Parenting strengths among African-Americans. *Families in Society, 76*, 434-443.

Iwata, N., Turner, R.J., & Lloyd, D.A. (2002). Race/ethnicity and depressive symptoms in community-dwelling young adults: a differential item functioning analysis. *Psychiatry Research, 110*: 281-289.

Jones, J. M. (1991). The concept of race in social psychology: From color to culture. In R. L. Jones (Ed.), *Black psychology* (3rd ed., pp. 441-467). Berkeley, CA: Cobb & Henry Publishers.

Jones, J.M. (1983). The concept of race in social psychology. In L. Wheeler & P. Shaver (Eds.), *Review of personality and social psychology* (Vol. 4, pp. 117-150). Beverly Hills, CA: Sage.

Kitayama, S. & Markus, H.R. (1992). Emotion and culture: Empirical studies of mutual influence. Washington, DC, US: American Psychological Association.

Kliewer, W., Lepore, S.J., Oskin, D., & Johnson, P.D. (1998). The role of social and cognitive processes in children's adjustment to community violence. *Journal of Consulting and Clinical Psychology, 66*(1), 199-209.

Krishnakumar, A., Buehler, C., & Barber, B. K. (2003). Youth perceptions of interpersonal conflict, ineffective parenting, and youth problem behaviors in European-American and African-American families. *Journal of Social and Personal Relationships, 20*(2), 239-260.

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

Lambert, M.C., Schmitt, N., Samms-Vaughan, M.E., An, J.S., Fairclough, M., Nutter, C.A. (2003). Is It Prudent to Administer All Items for Each Child Behavior Checklist Cross-Informant Syndrome? Evaluating the Psychometric Properties of the Youth Self-Report Dimensions With Confirmatory Factor Analysis and Item Response Theory. *Psychological Assessment, 15*(4): 550-568.

Lambert, M.C., Rowan, G.T., Lyubansky, M., Russ, C.M. (2002). Do problems of clinic-referred African-American children overlap with the Child Behavior Checklist? *Journal of Child and Family Studies, 11*(3), 271-285.

Latkovich, S.A. (1997). An investigation of the current, predictive, and construct validity of the Child Behavior Checklist for African-American children. *Dissertation Abstracts International: Section B: The Sciences and Engineering, 57*(11-B), 7271.

Lengua, L.J., Sadowski, C.A., Friedrich, W.N., & Fisher, J. (2001). Rationally and empirically derived dimensions of children's symptomatology: Expert ratings and confirmatory factor analyses of the CBCL. *Journal of Consulting and Clinical Psychology, 69*(4), 683-698.

Leventhal, T. & Brooks-Gunn, J. (2000). The neighborhoods they live in: The effects of neighborhood residence of child and adolescent outcomes. *Psychological Bulletin, 126*(2), 309-337.

Lord, F. & Novick, M. (1968). *Statistical theories of mental tests.* New York: Addison-Wesley.

Marsh, H., Hau, K.T, & Wen, Z. (2004). In Search of Golden Rules: Comment on Hypothesis-Testing Approaches to Setting Cutoff Values for Fit Indexes and Dangers in Overgeneralizing Hu and Bentler's (1999) Findings. *Structural Equation Modeling, 11*(3), 320-341.

Martinez, P. & Richters, J.E, (1993). The NIMH community violence project: Children's distress symptoms associated with violence exposure. *Psychiatry, 56*, 22-35.

Marsella, A.J. (2000). Culture: Culture and mental health. In A. E. Kazdin(Ed), Encyclopedia of psychology, Vol. 2. (pp. 400-404). Washington, DC, Oxford University Press.

McAdoo, H. P. (Ed.). (2002). *Black children: Social, educational, and parental environments* (2nd ed.). Thousand Oaks, CA: Sage Publications, Inc.

McDonald, R. (1999). *Test theory: A unified treatment*. Mahwah: Lawrence Erlbaum Associates.

McLeod, J.D. & Nonnemaker, J.M (2000). Poverty and child emotional and behavioral problems: Racial/Ethnic differences in processes and effects. *Journal of Health and Social Behavior, 41,* 137-161.

McLeod, J.D & Shanahan, M.J (1993). Poverty, parenting, and children's mental health. *American Sociological Review, 58,* 351-366.

McLeod, J.D. & Shanahan, M.J. (1997). Poverty, Parenting, and Children's Mental Health. *Sociological Review, 58,* 351-366.

McLoyd, V.C. (1990). The impact of economic hardship on Black families and children: Psychological distress, parenting, and socioeconomic development. *Child Development, 61,* 311-346.

McLoyd, V.C., Cauce, A.M., Takeuchi, D., & Wilson, L. (2000). Marital processes and parental socialization in families of color: A decade of review of research. *Journal of Marriage and the Family, 62*(4), 1070-1093.

McLoyd, V.C. & Steinberg, L. (1998). *Studying minority adolescents: Conceptual, methodological, and theoretical issues.* Mahwah: Lawrence Erlbaum Associates.

McLoyd, V.C. & Randolph, S.M. (1986). Secular trends in the study of African-American children: A review of Child Development, 1936-1980. *Monographs of the Society for Research in Child Development, 50*(4-5), 78-92.

Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist, 50*(9), 741-749.

Midgette, T.E. & Glenn, E. (1993). African-American male academies: A positive view. *Journal of Multicultural Counseling and Development, 21*(2), 69-78.

Muraki, E., & Bock, B.D. (2003). Parscale 4.1.: IRT based test scoring and item analysis for graded response models. [Computer program]. Chicago: Scientific Software International.

Murray, D.M. & Wolfinger, R.D. (1994). Analysis issues in the evaluation of community trials: Progress toward solutions in SAS/STAT MIXED. Journal of Community Psychology, 140-154.

Ngo, V. & Weiss, B. (2003, April). *Does the CBCL Measure the Same Construct for Blacks and Whites?* Poster session presented at the Society for Research on Child Development, Tampa, FL.

Ogbu, J. U. (1981). Origins of human competence: A cultural-ecological perspective. *Child Development, 52*(2), 413-429.

Okazaki, S. & Sue, S. (1995). Methodological issues in assessment research with ethnic minorities. *Psychological Assessment, 7*(3), 367-375.

Oshima, T.C., Raju, N.S., & Flowers, C.P. (1997). Development and demonstration of multidimensional IRT-based internal measures of differential function of items and tests. *Journal of Educational Measurement, 34,* 253-272.

Osofsky, J.D., Wevers, S., Hann, D.M.,& Fick, A.C., (1993). Chronic community violence: What is happening to our children? *Psychiatry, 56,* 36-45.

Park, D & Lautenschlager, G.L (1990). Improving IRT item bias detection with iterative linking and ability scale purification. *Applied Psychological Measurement, 14*(2), 163-173.

Perrin, E.C., Stein, R.E., & Drotar, D. (1991). Cautions in using the Child Behavior Checklist: Observations based on research about children with chronic illness. *Journal of Pediatric Psychology, 16*(4), 411-421.

Pettit, G.S., Bates, J.E., & Dodge, K.A., (1997). Supportive parenting, ecological context, and children's adjustment. Child Development, 68, 908-923.

Phinney, J.S., & Chavira, V. (1995). Parental ethnic socialization and adolescent coping with problems related ethnicity. Journal of Research on Adolescence, 5, 31-54.

Piquero, A.R., Macintosh, R., & Hickman, M. (2002). The validity of a self-reported delinquency scale: Comparison across gender, age, race, and place of residence. Sociological Methods & Research, 30(4): 492-529.

Ponterotto, J.G. (1988). Racial/ethnic minority research in the Journal of Counseling Psychology: A content analysis and methodological critique. *Journal of Counseling Psychology, 35*, 410-418.

Potenza, M.T. & Dorans, N.J. (1995). DIF assessment for polytomously scored items : A framework for classification and evaluation. Applied Psychological Measurement, 19(1), 23-37.

Quay, H.C., & Peterson, D.R.. (1987). *Manual for the Revised Behavior Problem Checklist*. Available from the senior author, University of Miami, Box 248074, Coral Gables, FL 33124.

Raju, N.S. (1988). The area between two item characteristic curves. *Psychometrika, 53*, 495-502.

Raju, N.S. (1990). Determining the significance of estimated signed and unsigned areas between two item response functions. *Applied Psychological Measurement, 14*, 197-207.

Raju, N.S. & Ellis, B.B. (2002). Differential item and test functioning. In F. Drasgow & N. Schmitt (Eds.), *Measuring and analyzing behavior in organizations: Advances in measurement and data analysis*. The Jossey-Bass business & management series, (pp. 156-188). San Francisco: Jossey-Bass Inc.

Raju, N.S., Laffitte, L.J., & Byrne, B.M. (2002). Measurement equivalence: A comparison of methods based on Confirmatory Factor Analysis and Item Response Theory. *Journal of Applied Psychology, 87*(3): 517-529.

Raju, N.S., van der Linden, W.J., & Fleer, P.F. (1995). IRT-based internal measures of differential functioning of items and tests. *Applied Psychological Measurement, 14*, 353-368.

Reise, S.P. & Waller, N.G. (2002). Item response theory for dichotomous assessment data. In F. Drasgow & N. Schmitt (Eds), *Measuring and analyzing behavior in organizations: Advances in measurement and data analysis. The Jossey-Bass business & management series* (pp. 88-122). S.F.: Jossey-Bass and Pfeiffer.

Richters, J.E., & Martinez, P (1993). The NIMH community violence project: Vol. 1. Children as victims of and witnesses to violence. *Psychiatry, 56*, 7-21.

Samejima, F. (1969). Estimation of latent ability using a response pattern of graded scores. *Psychometrika Monograph Supplement, 34* (4, Pt. 2).

Sampson, R. J. & Laub, J. H. (1993). *Crime in the making: Pathways and turning points through life.* Cambridge, MA: Harvard University Press.

Sampson, R.J., Morenoff, J.D., Raudenbush, S. (2005). Social anatomy of racial and ethnic disparities in violence. American Journal of Public Health, 95(2), 224-232.

Sampson, R.J., Radenbush, S.W., & Earls, F. (1997). Neighborhoods and Violent Crime: A multi-level study of collective efficacy. *Science, 277*, 918-924.

Schwab-Stone, M., Ayers, M.E., Kasprow, W., Voyce, C., Barone, C., Shriver, T., et al. (1995). No safe haven II: A study of violence exposure in urban communities. *Journal of American Academy of Child and Adolescent Psychiatry, 34,* 1343-1352.

Shakoor, B. & Chambers, D. (1991). Co-victimization of African-American children who witness violence: Effects on cognitive, emotional, and behavioral development. *Journal of the National Medical Association, 83,* 233-237.

Shealy, R.T. & Stout, W.F. (1993). In P.W Holland & H. Wainer (Eds.), *Differential Item Functioning* (pp.197-239). Hillsdale: Lawrence Erlbaum Associates.

Shuman, L., Vandell, D.L., Posner, J.K. (1998). Harsh, firm, and permissive parenting in low-income families: relations to children's academic achievement and behavior adjustment. *Journal of Family Issues, 19*, 483-502.

Simons, R.L., Johnson, C., Beaman, J., Conger, R.D., & Whitbeck, L.B., (1996). Parents and peer group as mediators of the effect of community structure on adolescent problem behavior. *American Journal of Community Psychology, 24*, 145-171.

Simons, R.L., Murry, V., McLoyd, V., Lin, K., Cutrona, C., & Conger, R.D., (2002). Discrimination, crime, ethnic identity, and parenting as correlates of depressive

symptoms among African-American children: A multilevel analysis. *Development and Psychopathology, 14*, 371-393.

Singer, J.D., & Willett, J.B. (2003). *Applied Longitudinal Data Analysis: Modeling Change and Event Occurrence*. New York: Oxford Press.

Steiger, J.H. (2000). Point estimation, hypothesis testing, and interval estimation using the RMSEA: Some comments and a reply to Hayduk and Glaser. *Structural Equation Modeling, 7,* 149–162.

Steinberg, L., Lamborn, S., Dornbusch, S., & Darling, N. (1992). Impact of parenting practices on adolescent achievement: Authoritative parenting, school involvement, and encouragement to succeed. *Child Development*, *63*, 1266–1281.

Steinberg, L., Mounts, N. S., Lamborn, S. D., & Dornbusch, S. M. (1991). Authoritative parenting and adolescent adjustment across varied ecological niches. *Journal of Research on Adolescence*, *1*, 19–36.

Strayhorn, J.M., & Weidman, C.S. (1988). A parenting practices scale and its relation to parent and child mental health. *Journal of American Academy of Child and Adolescent Psychiatry, 27*, 613-618.

Stocking, M.L. & Lord, F.M. (1983). Developing a common metric in item response theory. *Applied Psychological Measurement, 7,* 201-210.

Tolan, P.H., Gorman-Smith, D., Henry, D., Chung, K., & Hunt, M. (2002). The relation of patterns of coping of inner-city youth to psychopathology symptoms. *Journal of Research on Adolescence. 12*, 423-449.

Tolan, P.H., Gorman-Smith, D., Henry, D.B. (2002). Linking families violence to delinquency across generations. *Children's Services: Social Policy, Research, and Practice, 5*(4), 273-284.

Triandis, H.C. & Brislin, R.W. (1984). Cross-cultural psychology. *American Psychologist, 39*(9), 1006-1016.

U.S. Bureau of the Census. *Census of Population and Housing, 1990*. Summary Tape File 3 [CD-ROM]. U.S. Department of Commerce, Washington, D.C.

U.S. Census Bureau. (2001a). Profiles of general demographic characteristics: 2000 Census of Population and Housing, United States. http://www2.census.gov/census_2000/datasets/demographic_profile/0_National_Summary/.

U.S. Census Bureau. (2001b). The black population in the United States: March 2000 (Update) (Report NO. PPL-146). http://www.census.gov/population/www/socdemo/race/ppl-142. html.

Van de Vijver, F., & Hambleton, R. K. (1996). Translating tests: Some practical guidelines. *European Psychologist, 1*(2), 89-99.

Van de Vijver, F.J. & Poortinga, Y.H. (1997). Towards an integrated analysis of bias in cross-cultural assessment. *European Journal of Psychological Assessment, 13*(1),29-37.

Verhulst, F. C. (1989). The use of a standardized assessment instrument (the Achenbach Child Behavior Checklist) in longitudinal epidemiological research. In M.S. Brambring and L. Friedrich (Eds), *Children at risk: Assessment, longitudinal research, and intervention. Prevention and intervention in childhood and adolescence, 7.* (pp. 46-68): Walter de Gruyter, Berlin, Federal Republic of Germany

Verhulst, F. C., Achenbach, T. M., & Akkerhuis, G. W. (1989). Problems reported for clinically referred American and Dutch children. *Journal of the American Academy of Child & Adolescent Psychiatry, 28*, 516-524.

Verhulst, F. C., Achenbach, T. M., Althaus, M., & Akkerhuis, G. W. (1988). A comparison of syndromes derived from the Child Behavior Checklist for American and Dutch girls aged 6-11 and 12-16. *Journal of Child Psychology & Psychiatry & Allied Disciplines, 29*, 879-895.

Verhulst, F. C., & Althaus, M. (1988). Persistence and change in behavioral/emotional problems reported by parents of children aged 4-14: An epidemiological study. *Acta Psychiatrica Scandinavica, 77*(Suppl 339), 1-28

Waller, N.G., Thompson, J.S., & Wenk, E. (2000). Using IRT to separate measurement bias from true group differences on homogenous and heterogeneous scales: An illustration with the MMPI. *Psychological Methods, 5*(1), 125-146.

Weiss, B., Catron, T. & Harris, V. (2000). A two year follow-up of the effectiveness of traditional child psychotherapy. Journal of Consulting and Clinical Psychology, 68, 1094-1101.

Weiss, B., Catron, T., Harris, V. & Phung, T. (1999). The effectiveness of traditional child psychotherapy. Journal of Consulting and Clinical Psychology, 67, 82-94.

Weiss, B., Harris, V., Catron, T. & Han, S.S. (2003). Efficacy of the RECAP intervention program for children with concurrent internalizing and externalizing problems. Journal of Consulting and Clinical Psychology, 71, 364-374.

Weisz, J. R., Sigman, M., Weiss, B., & Mosk, J. (1993). Parent reports of behavioral and emotional problems among children in Kenya, Thailand, and the United States. *Child Development, 64*, 98-109

Weisz, J.R., Suwanlert, S., Chaiyasit, W., Weiss, B., Achenbach, T.M., & Trevanthon, D. (1989). Epidemiology of behavioral and emotional problems among Thai and American children: Teacher reports for ages 6-11. *Journal of Child Psychology and Psychiatry, 30*, 471-484.

White, L. & Rogers, S. (2000). Economic Circumstances and Family Outcomes: A review of the 1990s. *Journal of Marriage and Family, 62*, 1035-1051.

Whiteside Mansell, L., Bradley, R. H., Tresch Owen, M., Randolph, S. M., & Cauce, A. M. (2003). Parenting and children's behavior at 36 months: Equivalence between African-American and European-American mother - child dyads. *Parenting: Science and Practice, 3*, 197-234.

Williams, D.R., Lavizzo-Mourey, R., & Warren, R. (1994). The concept of race and health status in America. *Public Health Reports, 109*, 126-141.

Williams, D.R., Williams-Morris, R. (2000). Racism and mental health: The African-American experience. *Ethnicity & Health, 5,* 243-268.

Wilson, M. N. (1989). Child development in the context of the Black extended family. *American Psychologist, 44*(2), 380-385.

Xue, Y., Leventhal, T., Brooks-Gunn, J., Earls, F.J. (2005). Neighborhood residence and mental health problems of 5- to 11-year olds. *Archives of General Psychiatry, 62*, 554-563.

Yee, A.H., Fairchild, H.H., Weizman, F., Wyatt, G.E. (1993). Addressing psychology's problem with race. *American Psychologist, 48*, 1132-1140.White, L. & Rogers, S. (2000). Economic Circumstances and Family Outcomes: A review of the 1990s. *Journal of Marriage and Family, 62*, 1035-1051.