

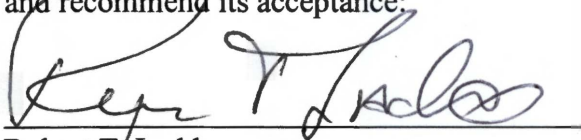
To the Graduate Council:

I am submitting herewith a dissertation written by Edgar H. Tyson entitled "Ethnic Differences Using Behavior Rating Scales to Assess the Mental Health of Children: A Measurement Equivalence Study." I have examined the final paper copy of this dissertation for form and content and recommend that it be accepted in partial fulfillment of the requirement for the degree of Doctor of Philosophy, with a major in Social Work.

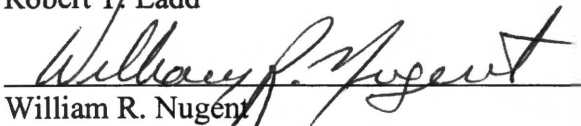


Charles Glisson, Major Professor

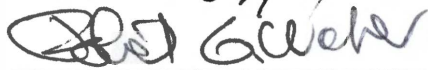
We have read this dissertation
and recommend its acceptance:



Robert T. Ladd

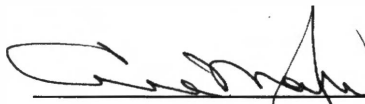


William R. Nugent



Robert G. Wahler

Accepted for the Council:



Vice Provost and Dean of
Graduate Studies

ETHNIC DIFFERENCES USING BEHAVIOR RATING SCALES
TO ASSESS THE MENTAL HEALTH OF CHILDREN:
A MEASUREMENT EQUIVALENCE STUDY

A Dissertation
Presented for the
Doctor of Philosophy Degree
The University of Tennessee, Knoxville

Edgar H. Tyson

December 2003

Thesis
2003b
• T97

DEDICATION

This dissertation is dedicated to my greatest critic, most rigorous editor
and most loyal supporter, my wife, Lorraine.

I love you Baby Girl.

ACKNOWLEDGEMENTS

All praises go to God, without whom nothing is possible. To the nameless participants in this study, I am grateful for their willingness to contribute to the advancement of knowledge. I also appreciate the hard work of the research assistants who collected these data. Thank you Cathy for finding Charles when I could not; Phillip for cleaning the data; Denny for always adding humor to whatever obstacle I confronted.

I have nothing but love and infinite gratitude for Melton Williams and Ron Simpson for helping me during the worst point in my life. Mel your guidance saved me. I am forever indebted to one of my mentors, Dr. Gil Raiford in helping me decide to pursue a MSW and then Ph.D. degree. Besides deciding to move to Miami where I became a man and met my love and soul-mate, Lorraine, it was the best decision I ever made. I am also grateful for the mentorship of Dr. Stan L. Bowie, it was his suggestion that I attend the great University of Tennessee. Thank you Stan, for some great advice.

To my cohorts, Juan, Ann, Elizabeth and Mona, I have tremendous respect and admiration for you all. Thank you for your friendship. Juan, that first year was made a whole lot easier because of you. Thank you Big Brother. Thank you Gary, for all of your wisdom and perspective.

I am profoundly grateful for the guidance and support of my committee, Dr. Tom Ladd, Dr. Bill Nugent, and Dr. Bob Wahler. Getting this group of people together consistently was essential in order for me to complete this work in a timely manner. Thank you for your direction and expertise.

I am especially indebted to my Chair, Dr. Charles Glisson. He is the most skillful researcher I have ever known. His commitment to help me complete this project as

scheduled, was surpassed only by his insistence that I complete a project that I would be proud of. Most important, he taught me that you do not have to sacrifice being a loving husband and dedicated father in order to be a successful scientist, scholar and educator.

I have infinite love and appreciation for the love and infinite support of my family. To my children, Tahir, Aminah, Wayne, Shatriya, Steven, and Christine, you are what motivates me to be the greatest father on the planet. It was often your laughter and simple joys in life that kept me going back to the office to work on this study (and most of the time just to get some peace and quiet). How you do it Baby Girl, I'll never know.

Mother, your presence has been a gift and a blessing. I love you. Thank you. To my Godmother Mae, it is from you and Mommy that I get my humility. No words can express the love and gratitude I have for you. Thank you. Willie (may you rest in peace), thank you for your support. Once again, I am sorry.

To my sisters and brothers, Afia, Wanda, Iveliesse, Abdul, and Chris, thanks for putting up with me when I was just a kid trying to be bad. Your love, support and direction helped me become a man who is finally going to do good. Iveliesse, I love you for saving my life and being there when I needed you most. Wanda, thank you for the laughter. Abdul, I love your strength. Afia, you will always have a special place in my heart because we have traveled similar paths in life. I am proud to be your brother. Chris, I have profound love, admiration and respect for you as a man, a father and a husband, but most of all as my big brother. I don't' now how God did it, but She now lives in you.

Finally to my parents I owe my life. Thank you both for bringing me into this world and teaching me that no one gets anywhere without some help and encouraging me to help others. My mother Alma "Theresa Birch" Neal, (Peace and Blessing Be Upon Your Soul)

was the most humble and beautiful person I have ever known. I see you smiling down on me Mommy and I hear your funny laughter. You gave me life twice, at birth and again when you put me on an airplane and sent me to live with daddy in Miami. Had I not gone to Miami, there is not doubt in my mind I could not possibly have lived past my 25th birthday. Mommy you were the best mother a son could ever have. Thank you Mommy for believing in me when no one else did. It is from you Mommy that I got my most powerful personal strength, the love and gratitude for God. I love you Mommy.

My "Pop", Harold C. Tyson, Sr., is the hardest working individual on the planet. It is from him that I got my most valuable resource, the love and thirst for knowledge and information. Pop, you are the brightest, most well rounded man I know. Thank you for the clips of the New York Times that you started sending me when I was eight years old living miles apart from you. More importantly, thank you for the clips of the Times that you continue to give me today. You taught me that being black in America means critically thinking about the contradictions in life and doing everything in my power to address and correct them. You also taught me that family is more important than anything. My greatest joy is that you were here to see this. I wish Mommy were here too. I love you Pop. Now, since you're a little aged to play, lets go watch the game!

ABSTRACT

Cultural competence in measurement has become increasingly important in child welfare and juvenile justice systems where disproportionate numbers of children and families from diverse ethnic groups are served. However, there is a present gap in the literature regarding the assessment of children from different ethnic groups using behavior-rating scales. This study examined the cross-ethnic measurement equivalence of the SAC (Glisson et al, 2001) using a sample that included 562 African American (AA) and 692 Caucasian-American (CA) children. These data were part of a three panel longitudinal survey of families referred to juvenile court in one state. Specifically, baseline and first time follow-up data (at least six months) on these children were used to conduct the analyses presented here. Caregivers completed several assessment instruments and a family interview questionnaire. Two types of reliability, as well as multiple sources of postdictive and predictive criterion-related validity were analyzed. Furthermore, the factorial equivalence of the SAC was analyzed using a structural equation modeling procedure. Alpha reliabilities for the full scale and for each sub-scale (i.e., externalizing and internalizing) were high for both groups. Test-retest reliabilities were also good for both groups. There were substantially more similarities than there were differences between the two groups in terms of the criterion validity evidence. The factorial equivalence evidence was mixed, and suggested that while the factor structure of the SAC was non-equivalent across the two groups, it worked similar and equally well for both groups. Practical implications of these results and directions for future research are forwarded.

TABLE OF CONTENTS

CHAPTER		PAGE
I.	INTRODUCTION	1
	Significance of the Problem.....	1
	Purpose of the Study.....	3
	Selection of the Shortform Assessment for Children (SAC).....	4
	Important Caveat of the Present Study.....	6
	Defining Ethnicity.....	7
II.	CHILD MENTAL HEALTH ASSESSMENT	11
	Historical Perspective on Classification Issues.....	11
	Categorical Diagnostic Classification	14
	Dimensional Classification	17
	Broadband Versus Narrow-band Classifications	20
III.	ETHNIC GROUP DIFFERENCES IN ASSESSMENT	25
	Ethnic Differences.....	25
	Defining and Assessing Test Bias	26
	Measurement Equivalence and Construct Validity	29
	Methods Used to Establish Measurement Equivalence.....	30
	Evidence of Measurement Equivalence Behavior Rating Scales.....	34
IV.	REMAINING GAPS IN THE LITERATURE.....	45
	Summary and Rationale for the Present Study	50
V.	THE PRESENT STUDY	53
	Research Questions	55
	General Research Hypothesis.....	56
VI.	METHOD	59
	Design.....	59
	Participants	60
	Data Collection.....	61
	Measures Used in the Study	61
	Data Analysis.....	64
VII.	RESULTS	71
	Descriptive Statistics	71
	Reliability Results	73
	Intercorrelations and Convergent-Divergent Validity	74
	Postdictive Criterion Validity.....	76
	Predictive Criterion Validity	79
	Separate Factor Analysis for AA and CA Children.....	82
	Simultaneous Factor Analysis (SFA) Across AA and CA Children.....	83

VIII. DISCUSSION.....	89
Summary and Implications of Major Findings.....	89
Summary and Implications of Unexpected Findings.....	93
Limitations of the Study.....	95
Conclusions and Directions for Future Research.....	97
REFERENCES	103
APPENDICES	127
Appendix A: Tables.....	129
Appendix B: FIQ-Postdictive Variables.....	152
Appendix C: FIQ-Predictive Variables.....	153
Appendix D: SAC.....	154
Appendix E: CIS.....	155
Appendix F: SACA.....	156
VITA.....	157

LIST OF TABLES

TABLE	PAGE
1. Demographic Profile, Reason For Referral, and Prior Custody of Participants	131
2. Means, Standard Errors and Standard Deviations of SAC Scores by Ethnicity.....	133
3. Internal Consistency and Test-Retest Reliability of the SAC for Both Groups.....	134
4. Intercorrelations of the SAC and Convergent-Divergent Validity With the Columbia Impairment Scale (CIS) for AA and CA.....	135
5. Frequency and Univariate Statistics of Postdictive-Validity Criterion Variables.....	136
6. Postdictive Criterion Validity Coefficients of Total SAC for AA and CA Children and Tests for Significant Differences.....	137
7. Postdictive Criterion Validity Coefficients of Internalizing Subscale for AA and CA Children and Tests for Significant Differences.....	138
8. Postdictive Criterion Validity Coefficients of Externalizing Subscale for AA and CA Children and Tests for Significant Differences.....	139
9. Frequency and Univariate Statistics of Predictive-Validity Criterion Variables.....	140
10. Predictive Criterion Validity Coefficients of Total SAC for AA and CA Children and Tests for Significant Differences.....	141
11. Predictive Criterion Validity Coefficients of Internalizing Subscale for AA and CA Children and Tests for Significant Differences.....	142
12. Predictive Criterion Validity Coefficients of Externalizing Subscale for AA and CA Children and Tests for Significant Differences.....	143
13. Initial Simultaneous Factor Analysis (SFA) of the SAC Across AA and CA Children.	144
14. Initial Chi-square Difference Tests for Parameter Free and Constrained Models.....	145
15. Ethnic Group Differences in Internalizing (INT) Standardized Regression Weights Within the All Free Estimates Model of the Simultaneous Factor Analysis.....	146

16. Ethnic Group Differences in Externalizing (EXT) Standardized Regression Weights Within the All Free Estimates Model of the Simultaneous Factor Analysis..... 148

17. Follow-Up^a Simultaneous Factor Analysis (SFA) of the SAC Across AA and CA Children..... 150

18. Follow-Up Chi-square Difference Tests for Parameter Free and Constrained Models..... 151

CHAPTER I

INTRODUCTION

Significance of the Problem

Currently, millions of children are overwhelming child welfare and juvenile justice systems nationwide (Snyder, 2000; U.S. Department of Health and Human Services, 1999; U.S. Department of Justice, 1999). The Federal Bureau of Investigation reported that there were approximately 2.5 million arrests of juveniles in 1999 (Snyder, 2000). According to a summary of state reports from 1999, an estimated 826, 000 children nationwide entered the child welfare system due to child maltreatment alone (U.S. Department of Health and Human Services, Administration on Children, Youth and Families, 2001). Depending on the study, the proportion of these children¹ with persistent and severe mental health and behavioral problems range from 68% to 80% (Bruce-Webb & Jones-Harden, 2003; Glisson, 1996; Glisson, Hemmelgarn, & Post, 2002; Martin, Peters & Glisson, 1998) and they are in need of appropriate, evidence-based interventions. We know that early and accurate assessment of children and adolescents is a fundamental prerequisite for appropriate and effective intervention (Malgady, 1996; Merrell, 1999a). Without timely and adequate assessment and intervention, children are at risk of chronic behavioral and mental health problems that will follow them into adulthood.

A recurrent theme in the children's services literature is that although a large percentage of youth entering the juvenile justice and child welfare systems have persistent serious mental health problems, few receive adequate assessment and treatment (Glisson, 1996; Martin, et al., 1998; Hoge, 2002; Justice for Juveniles Initiative, 1999).

¹ Children will be used to describe both children and adolescents throughout the remainder of this study.

Placement decisions and judgments about risk and service needs are often made through informal and unsystematic procedures (Glisson, 1996; Martin, et al, 1998; Hoge, 2002), such as labeling (Martin, et al, 1998). This body of research highlights the increasing importance of and need for standardized and valid assessment instruments in child welfare and juvenile justice systems across the United States.

Considerable attention has been given to the development of reliable and valid mental health assessment methods over the past two decades (see Elliot, Busse & Gresham, 1993; Ramsey, Reynolds, & Kampaas, 2002, for review). Researchers have been particularly interested in the use of problem behavior rating scales for children (Kampaas, Petoskey, & Rowe, 2000; Meyers & Winters, 2002). However, studies examining the validity of assessment tools with children from different ethnic groups within the United States have been rare and those that do exist have been less than optimal (Byrne, & Campbell, 1999; Epstein, March, Conners, & Jackson, 1998; Guerra & Jager, 1998; Knight & Hill, 1998). And few tools have been developed explicitly for child welfare and juvenile justice populations.

Past and recent editions of the Standards for Educational and Psychological Testing (American Psychological Association; APA, 1985, 1992, 1999) clearly state standards for assessment with special populations require psychologists to be familiar with the reliability, validation and standardization of the assessment methods they use. Nearly a decade ago, Nugent (1993) cited these standards to make the case that social workers should conduct rigorous evaluations of various types of validity of measurement tools for different groups. Moreover, some researchers have argued for increased volume and rigor in research on equivalence of measures used to assess children in this country (Okazaki & Sue, 1995). Despite these arguments, few of the assessment tools most often used in research and practice

with children in both psychology and social work have been rigorously tested for measurement equivalence across different ethnic groups here in the United States. And almost none of these measures have been tested for cross-ethnic equivalence with children referred to child welfare and juvenile justice systems (Tyson, in press).

Purpose of the Study

A major criticism of behavioral science research in general is that it has had limited focus on children of color, such as African-American (AA) youth (McLoyd, 1998). There is particular concern over the limited research on mental health assessments of AA children (Knight & Hill, 1998). With AA youth representing a significant proportion of youth referred to and assessed by child welfare, juvenile justice and educational systems, more information is needed about the generalizability and validity of methods that are developed with samples that are primarily Caucasian American (CA) to AA and other ethnic minority youth. This is particularly important in behavioral assessment research because ethnic or racial biases in any behavioral or psychological assessment measure have ethical, practical, and legal implications (Cooke, Kosson, & Michie, 2001; Gottfredson, 1994; Okazaki & Sue, 1995). Furthermore, these implications are especially problematic for children, given the importance of placement decisions in the child welfare and juvenile justice arenas that are based in part, on assessment outcomes. The use of unreliable and invalid assessments has been identified as a critical factor in understanding the gap between the mental health status of children and appropriate placements and services (cf. Douck, English, DePanfils, & Moote, 1993). Therefore, the purpose of this study is to examine the extent there is evidence that behavior-rating scales are equally reliable and valid for measuring the emotional and behavioral health

of both AA and CA youth referred to child welfare and juvenile justice systems in the United States.

Selection of the Shortform Assessment for Children (SAC)

An important first step in this effort to examine the cross-ethnic equivalence of behavior rating scales with samples of children in the child welfare and juvenile justice systems, is to identify an assessment tool that is appropriate for these systems of care. Careful selection of a measure for examination in this study is necessary because few instruments have been designed for child welfare and juvenile justice populations. Child welfare and juvenile justice case managers have particular needs that have rarely been considered in scale development research. These case managers are typically burdened with overwhelming caseloads, have limited training in mental health and inadequate knowledge of complex standardized assessment instruments (Duchnowski & Friedman, 1990; Samantrai, 1992). Demanding time constraints and limited resources exacerbate these issues. Taken together, these barriers limit the effectiveness of longer, more complex scales within child welfare and juvenile justice systems and tend to alienate workers and increase misinterpretation of the scales (Glisson, 1996; Glisson et al., 2002).

The Shortform Assessment for Children (SAC) was developed specifically to address the assessment issues inherent in child welfare and juvenile justice systems. Glisson and colleagues (Glisson et al., 2002; Hemmelgarn, Glisson and Sharp, 2003), with support from the National Institute of Mental Health (NIMH), designed the SAC to provide child welfare and juvenile justice workers with a short, reliable, easy-to-use and easy-to-interpret tool to make valid assessments of the behavioral and mental health needs of children.

In the initial study of the SAC, data collected from 3,790 children served by the child welfare and juvenile justice system in one state provided support for the internal reliability and factor structure stability of the SAC (Glisson et al., 2002). Using primarily case managers to obtain teacher and parent reports, several important outcomes have emerged from this research. The internalizing and externalizing items of the SAC loaded on their respective factors for four samples defined by adolescent and preadolescent children of each gender. Moreover, assessments by both teacher and parents were obtained and excellent model fit indices confirmed the factor stability of the SAC across each of the eight subgroups formed by age, gender and respondent (Glisson et al., 2002). Additionally, alpha reliabilities for all subgroups ranged from .94 to .96 for the externalizing scales and .86 to .90 for the internalizing scales. An important outcome of this study was that it documented the SAC can accurately assess child mental health using the two broadband syndromes of internalizing-externalizing without narrow-band sub-scales. Finally, the initial scale development study of the SAC reported data that were collected as part of another study involving children served by child welfare and juvenile justice systems and found that teacher and parent responses predicted time in custody and level of placement restrictions of children.

A second study was conducted to provide additional knowledge about the validity of the SAC (Hemmigarn et al., 2003). Essentially, this study compared the convergent, divergent and predictive-criterion validity of the SAC to that of the Child Behavior Checklist (CBCL) and Teacher Report Form (TRF) (Achenbach 1991a, 1991b) using parent and teacher reports of a moderate-sized sample of children as they entered custody. The importance of this study was that it showed SAC responses from parents and teachers provided validity evidence comparable to the longer scales of the CBCL and TRF, despite the

use of substantially fewer items. Specifically, both teacher and parent responses on the SAC converged and diverged with the CBCL and TRF in a manner consistent with theoretical knowledge. More important, both teacher and parent reports of externalizing problems predicted placement outcomes such as fighting, ejections, and length of stay in custody, as well as (and in some cases better than) the CBCL and TRF (Hemmelmarg et al., 2003).

The previous research on the SAC demonstrates its utility and feasibility, as well as its reliability and validity as a measure of child mental health for children in child welfare and juvenile justice systems (Glisson et al., 2002; Hemmelmarg et al., 2003). Moreover, the SAC has been recently identified as an important development in the assessment of childhood aggressive behaviors (Tyson, Wodarski, & Dulmus, 2002). Therefore, the SAC is ideal to test for cross-ethnic equivalence of measures used to assess the mental health problems of children referred to child welfare and juvenile justice systems.

Important Caveat of the Present Study

The present analysis is primarily concerned with the validity of behavior rating scales for one of the two largest ethnic/racial minority groups in the United States², which include AA and Hispanic-Americans (HA). These two groups are particularly important because of their relative numbers and risk factors such as poverty that characterize a large proportion of these children (Merrel, 1999a). The comparative research discussed here might be better understood as “cross-ethnic” research (Oomen, 1989; Yau-Fai Ho, 1984), as opposed to “cross-cultural” research. However, a clear distinction between these two types of research is not present in the literature (van de Vijver & Leung, 1997). It is generally considered cross-

² According to recent census data (US Census, 2000), the two main ethnic/racial minority groups in the United States are Hispanic-Americans and African-Americans.

cultural research when two or more cultures from different countries are compared and cross-ethnic when two or more groups are compared based on differences in race, language, region, color or some other attribute besides country of origin (Barber, Stolz, Olsen, & Maughan, manuscript submitted; Knight & Hill, 1999; van de Vijver & Leung, 1997). Barber et al., (manuscript submitted) suggested that making this distinction is important because comparative researchers in the past have argued that cultural group is not synonymous with ethnic group (Oomen, 1989; Yau-Fai Ho, 1984).

The lack of clarity notwithstanding, this study does not assess the cross-cultural validity of problem behavior rating scales. Many of the issues addressed in this study have been previously examined using cross-cultural samples. For example, there appears to be sufficient reviews and documentation of rigorous validity studies using the CBCL on samples from different cultures (Brown & Achenbach, 1996; Crijnen, Achenbach, & Verhulst, 1997; Heubeck, 2000; Weisz, & McCarty, 1999). Specifically, the CBCL has been translated into at least 51 languages and studies have been published from at least 36 different cultures (Brown & Achenbach, 1995). Surprisingly, less attention has been given to examinations of behavior rating scales used with various ethnic groups within the United States, particularly AA children. Therefore, the focus of this study is on cross-ethnic comparisons of behavior ratings of AA and CA youth.

Defining Ethnicity

Research involving issues of ethnicity and race must first address definitional assumptions and limitations. It is not surprising that there is no one definition of ethnicity, race, or culture that is agreed by all, and many researchers (possibly erroneously) use these terms interchangeably (Betancourt & Lopez, 1993). The rationale for using one approach

versus another is not always clearly explained by the researcher. One source of debate is whether race and ethnicity are social or biological constructs (Kaufman & Cooper, 2001). Ethnicity and race are not stagnant, but instead fluid, in that they primarily represent geographic, social and cultural forces. Therefore, they can be considered social constructs. On the other hand, some have argued that ethnic and racial groups differ genetically, and there are certain biological implications of these categories (cf. Cooper, Kaufman, & Ward, 2003). These issues continue to be vigorously debated and are highly charged with political and public health ramifications (Cooper et al, 2003; Burchard et al., 2003). What is certain is that race and ethnicity are concepts that will continue to have currency in our daily discourse and the more we know about these constructs the better we will be able to answer some of the hard questions related to serving the needs of children at risk.

The implicit reasoning behind using race, ethnicity or culture as grouping variables in research is based on the assumption that (a) individuals share some common psychological characteristic associated with these constructs and (b) these shared characteristics are related to personality and how individual behavior is expressed (Okazaki & Sue, 1995). These assumptions continue to underlie the vast majority of research involving diverse ethnic groups. However, within a given racial or ethnic group there remains a wide range of income, education and social capital that influence behavior and life outcomes (Furstenberg, Cook, Eccles, Elder, & Sameroff, 1999; Tatum, 1997). Additionally, the study of within group differences is becoming increasingly important in social research. Nonetheless, the use of ethnicity, race and culture in assessment and treatment research is necessary so that we might

better understand the prevalence, influence and responsiveness of factors related to these constructs.

To summarize, ethnicity is the primary grouping construct used in this study for two important reasons. First, some researchers argue that ethnic background is a broad construct that accounts for socialization, cultural tradition, common history, religion and other important group characteristics (Burchard et al., 2003; MacAdoo, 2002; Peters, 2002). Although ethnicity, race, culture and social class are very closely related (Eaton, 1980), ethnicity appears to encompass important aspects of all four of these grouping variables. Secondly, the shared social, cultural, and historical experience of AA not only shapes how they parent and interact with their children (MacAdoo, 2002; Peters, 2002), but is also likely to shape how they interpret and respond to questions (i.e., items from behavior rating scales) regarding their children's behavioral and mental health (Lambert, Markle, & Bellas, 2001; Lambert, Rowan, Lyubansky, Russ, 2002).

CHAPTER II

CHILD MENTAL HEALTH ASSESSMENT

Treatment and intervention with individuals begins with the assessment of a “target” behavior or emotion, or in the case of child mental health, a “behavior problem” or “childhood disorder.” Similarly, it has been suggested that before “it” can be changed or treated we have to know what “it” is, and that description and classification are two fundamental purposes of assessment (Cone, 1998; Evans, 1993; Merrell, 1999a). Moreover, there is a long history of the evolution of assessment procedures (Mash & Terdel, 1997; Merrell, 1999a). Theoretically, the improvement in assessment procedures has helped clinicians make similar improvements in how they classify or categorize a particular problem area that might be in need of intervention. Whether or not this has led to our ability to develop more effective treatments is still being debated. However, it does appear that the development of assessment instruments for children is invariably tied to a particular theoretical conceptualization of how behaviors are classified (Mash & Terdel, 1999). Therefore, it is important to review the prevailing methods of classification in order to identify which methods have been the most useful in guiding the development of assessment instruments of researchers and practitioners.

Historical Perspective on Classification Issues

A review of the literature reveals a rich history of empirical research that has contributed to the definition and classification of childhood behavior problems that extends back a half century (Hewitt & Jenkins, 1946; Wittenborn & Holzberg, 1951). These early researchers were well informed by the theoretical contributions of several important scholars (Horney, 1945; Riemer, 1942; Rosenzweig, 1945). For example, Riemer (1942) was a pioneer

criminologist who developed knowledge, mainly through theoretical explications, on how to define behavioral constructs. Riemer was very concerned about the need to employ better statistical control over constructs guided by theoretical speculation in order to make stronger predictions about future behavior. Specifically, Riemer stated,

“...a definition has to be chosen that is strictly verifiable...the relations between the ideal type [theoretical construct] and the operational definition is based upon assumptions as to the symptomatic value of these factors that enter into the statistically routine work. Objectivity in the sense of observable regularities can be established in relation to the operational definition only...” (p. 188-189).

This was one of the earliest arguments made in support of the use of factor analysis to define behavioral categories. Horney (1945) may have been the first to consider classifying individuals with emotional and behavioral problems into two broad dimensions when she postulated a theoretical model that identified two types of people, those who “move away from the world” and those that “move against the world”. A similar classification model was introduced by Rosenzweig (1945). Rosenzweig used the terms, “extrapunitive” and “intropunitive” to describe behavior types that correspond to the behavior patterns where people turn against others or turn against the self.

Hewitt and Jenkins (1946) were among the first researchers to empirically study the issue of classification when they examined the behavior problems of delinquent and non-delinquent youth referred to the Michigan Child Guidance Institute. One of the gaps that this seminal study attempted to address was that behavioral classifications were “...so broad that behavioral items with essentially different meanings and implications to the child have been thrown together, or single items of behavior have been so narrowly defined...” (p. 1).

Consequently, based on the theoretical work of Riemer (1942), Hewitt and Jenkins (1946) were the first to suggest the “structuralization of syndromes” of behavior. The underlying assumption that guided Hewitt and Jenkins thinking was the idea that “...problem behavior tends in time to become accumulative and in certain definite directions,...these correlated sets of behavior symptoms are not merely synonyms of each other but imply a similar way of responding to social situations...” (p. 25). Hewitt and Jenkins identified 94 symptomatic traits, which were then classified into three clusters. These three clusters were over-inhibited/repressed child, un-socialized aggressive child and socialized delinquent child. The first behavior cluster (i.e., over-inhibited/repressed) was the predecessor to what is now referred to as the primarily “internalized” dimension and the latter two clusters, predecessors to what is known as the “externalized” dimension.

Another of the earlier important studies was Wittenborn and Holzberg (1951). Although this study included adult psychiatric patients, it moved the literature on measurement of behavior forward in two very important ways. First, the work of Wittenborn and Holzberg provided additional evidence for the use of symptom-clustering through factor analysis as the basis for describing and identifying groups of patients in an effort to improve treatment for a particular disorder. This provided early support for the present day dimensional approach to classification discussed below. Second and more important to this discussion, these researchers were among the first to find evidence for the use of items such as “shouts, sings, and talks loudly”, “dramatically attention-demanding”, “initiates physical assaults”, “mood changes very frequent and abrupt”, and “characteristically oppositional” to describe a behavior pattern (p. 378). What is most significant about these findings is that these items all loaded high (.435 to .767) on either the “Excitement” Factor or “Manic-Depressed” Factor,

and are similar to items currently used to identify what we now refer to as externalizing behavior in children.

Before the field of behavior assessment was improved by the well-formulated classification schemes of modern day, the history of classifying child behaviors included many changes, partly in response to heavy criticism. For example, Hobbs (1975) discussed in detail the harmful and stigmatizing effects of early approaches of categorizing and labeling children, primarily because children were classified, rather than their behaviors. Moreover, Ross (1980) suggested that historically, classification mainly served administrative, rather than therapeutic purposes. However, the need for a common language in child mental health practice and research led the movement to gain greater acceptance of classification schemes and faith in their utility (Merrell, 1999a). Perhaps more than any other classification schemes, the categorical diagnosis and dimensional classification approaches have had the greatest impact on our current conceptualization of child mental health (Mash & Terdel, 1999; Sliverman & Serafini, 1998). These two approaches are the basis for the way in which we understand, assess, and classify nearly all child behavior problems and disorders (Mash & Terdel, 1999). The categorical diagnosis approach will be discussed first, followed by a review of the dimensional approach, which also has a rich history in the psychological literature. For the purposes of this study, it is important to note that this early body of research (as with many early scientific inquiries) did not address possible differences across ethnic groups.

Categorical Diagnostic Classification

Few would disagree with Merrells' (1999a) claim that the most widely used classification system of behavioral, social, and emotional problems in North America is the

Diagnostic and Statistical Manual for Mental Disorders (DSM; American Psychiatric Association; APA, 1980, 1987, 1994; 2001). The DSM was first introduced in 1952 and has had several substantial changes over the past two decades. The DSM is based on a medical-model of mental health from a pathological perspective and continues to have a significant impact on the mental health field (Merrell, 1999). It is the primary text used in clinical psychology for psychopathology courses and while it has been hotly debated whether or not the DSM manual *should be* used for teaching social work practice in mental health (Kutchins & Kirk, 1995; Williams & Spitzer, 1995), it *is* one of the main texts used in graduate training in clinical social work for psychopathology courses (Zide & Gray, 2001). In addition to its widespread use in educating social workers and other clinicians, the DSM-IV (APA, 1994) remains the primary reference manual for documentation for reimbursement purposes among all mental health professions (Merrell, 1999a). One of the major criticisms of the DSM classification system is that it was not derived through empirical analysis, but through a subjective process that relied on an expert panel of clinicians to determine the criteria of each disorder (Kasius, Ferdinand, van den Berg, & Verhulst, 1997). Specifically, expert panels developed a consensus through extensive negotiations to determine the criteria for a specific disorder.

The most recent DSM, the DSM-IV (APA, 1994) and its text revision form DSM-TR (APA, 2001), attempted to address this criticism by employing a standardized (i.e., structured and semi-structured) clinical interview to determine the diagnostic criteria of specific disorders (Mash & Terdel, 1999). However, the categorical approach, by definition, continues to assume that diagnostic entities are qualitative and discrete and that there are distinct boundaries between them (Moras & Barlow, 1992). Therefore, the core of

Achenbach's (1980) criticism of the categorical diagnostic classification scheme being a "yes/no" or "either/or" approach, as well as being a system compromised by diagnostic presupposition and clinical notions (Achenbach, 1988) continues to apply.

Nonetheless, it is important to note two critical developments. First, the DSM-IV (APA, 1994), included 11 National Institutes of Mental Health (NIMH) supported reliability and validity studies (Merrell, 1999a; Widiger, Frances, Pincus, Davis, & First, 1991). These research efforts were initiated for the distinct purpose of documenting the rationale and empirical support for the proposed changes of the DSM. Arguably, the empirical integrity of the DSM appears to have improved substantially through the inclusion of empirically derived data.

Secondly, the authors of the DSM recognized early on that a simple enumeration of symptoms may not be sufficient for diagnosis and treatment planning, and almost from the very beginning included a multi-axial classification scheme to help clinicians make comprehensive diagnoses (APA, 1980). The multi-axial scheme includes Axis I-focus of clinical attention (i.e., the primary disorder), Axis II-personality disorders, Axis III-general medical conditions, Axis IV-psychosocial and environmental problems, and Axis V- global assessment of functioning. Nonetheless, previous reviews of the multi-axial system concluded that Axes IV and V are of limited and undetermined reliability and validity (Goldman & Skodal, 1992). Thus, many practitioners and clinical researchers argue that "empirically-based assessment" approaches are more appropriate (Shapiro & Kratochwill, 2000).

Dimensional Classification

After Hewitt and Jenkins (1946), Achenbach (1966) published what may have been the next important study in the area of classification of childhood disorders by building on their empirical work and the work of others (e.g., Bennett, 1960; Geurtin, 1952; Phillips & Rabinovitch, 1958). Specifically, as part of a larger National Institutes of Mental Health study (under the direction of Norman Garnezy and Edward Zigler) Achenbach (1966) conducted a well-formulated factor analytic study of the mental health problems of children. Achenbach analyzed the “psychiatric symptoms” of 300 Caucasian male and 300 Caucasian female psychiatric patients, between the ages of 4 and 16 years, using data from their case histories. The procedure Achenbach used was exploratory factor analysis (EFA) and the significance of this method will be discussed later in this review.

There were two relevant outcomes of Achenbach’s (1966) early study that deserve particular attention here. First, Achenbach found support for the use of a symptom checklist to empirically identify problem behaviors that were “...constructed from items which regularly appeared in previous studies, which seemed to involve minimal inference, which could be considered mutually exclusive with regard to specific observations...” (p. 7). This is particularly important to note because this symptom checklist became the basis for the development of the widely used CBCL (Merrell, 1999), which is discussed in detail below.

The key difference between the categorical diagnostic approach and the dimensional approach is that the latter is fully rooted in empirical methods and statistical procedures for identifying clusters of highly inter-correlated behaviors (Kasius, et al., 1997; Merrell, 1999). Conversely, the categorical diagnostic approach continues to be significantly influenced by expert consensus about which disorders should be considered for classification and what the

criteria are for defining such disorders (Kasius, et al., 1997; Verhulst & Achenbach, 1995). Although the DSM-IV (APA, 1994) has incorporated much clearer guidelines for culturally relevant practices in making a diagnosis, confirmatory factor analytic procedures are much more precise methods to control for errors in making classification distinctions. Therefore, the dimensional approach appears to have the potential for stronger empirical support (Floyd & Widaman, 1995).

A second important outcome of Achenbach's (1966) early study was that it provided empirical evidence for the classification of the symptoms/items into two of the dimensions or constructs empirically derived by Hewitt and Jenkins (1946) and Wittenborn and Holzberg (1951). Taken together, these three studies form the foundation of subsequent research on the dimensional classification approach. Finally, another area in which Achenbach (1966) made a significant contribution was the further differentiation of the list of items (i.e., symptoms) into 8 and 11 specific or "narrow-band syndromes", for males and females respectively. Further discussion on broadband versus narrow-band classification schemes will be presented later.

More than a decade after Achenbach's (1966) initial study, two seminal reviews on dimensional classification by Achenbach and Edlebrock (1978) and Quay (1979) revealed that across a variety of studies with respect to assessment instrument, target population and type of informant, these two broadband dimensions were reliably obtained. More importantly, since Achenbach's (1966) early work, the broadband scheme of "internalizing and externalizing" behavior has been one of the most widely recognized classification schemes in the child behavior community. For example, the internalizing and externalizing dimensional scheme used to describe large classes of behavioral and emotional problems is

what has been most frequently published in research on child psychopathology (Cicchetti & Toth, 1991), particularly in longitudinal research (Ollendick & King, 1994; McMahon, 1994). Therefore, a detailed discussion of internalizing and externalizing behaviors is important to understanding the current state of behavior assessment and is presented in the next section.

Internalizing-Externalizing Symptoms or Behaviors. The internalizing dimension refers to a behavioral cluster of related symptoms such as anxiety, depression, social withdrawal, and somatic complaints (Cicchetti & Toth, 1991; Ollendick & King, 1994). While there is some debate surrounding the conceptual and nosologic foundation of the term, the label is commonly used in child psychiatry and psychology (e.g., Brady & Kendall, 1992, Kovacs & Devlin, 1998), as well as social work (e.g., Glisson, et al., 2002). Moreover, with the exception of externalizing problems (see McMahon, 1994), this cluster of internalizing problems is the most frequently reported set of behavior problems (Ollendick & King, 1994).

The externalizing dimension refers to a second cluster of related behaviors such as noncompliance, destructiveness, impulsivity, delinquent and aggressive behavior and hyperactivity (McMahon, 1994). These behaviors show a tendency toward coercion and acting out, thus the term “external.” While the internalizing behaviors relate to behaviors previously categorized as “overcontrolled”, the externalizing dimension relates to behaviors previously categorized as “undercontrolled” (Merrell, 1999). Some researchers further suggest that “internalizing disorders” have a central feature of disordered mood or emotion and “externalizing disorders” have a central feature of dysregulated behavior (Kovacs & Devlin, 1998). Finally, this distinction has led some researchers to categorize children with mostly internalizing symptoms as having an “emotional disorder” and children with primarily

externalizing symptoms as having a “behavioral disorder” (Barber, Neese, Coyne, Fultz, & Fonagy, 2002; Kovacs & Devlin, 1998). We do know that both clusters of behavior have stable and enduring etiologies that require accurate and early assessment (Ollendick & King, 1994; McMahon, 1994) for appropriate and timely treatment.

Broadband Versus Narrow-band Classifications

The CBCL is probably the best-known measure that incorporates narrow-band sub-scales and broadband syndromes to assess the mental health of children. The broadband scales of externalizing and internalizing of the CBCL represent the higher order factors and the narrow-bands are represented by the eight sub-scales, (1) withdrawn, (2) somatic complaints, (3) anxious/depressed, (4) social problems, (5) thought problems, (6) attention problems, (7) delinquent behavior, and (8) aggressive behavior (Achenbach & McConaughy, 1997). The first three narrow-band syndromes (i.e., withdrawn, somatic complaints, and anxious/depressed) are what define the internalizing dimension and the final two narrow-band syndromes (i.e., delinquent behavior and aggressive behavior) are what define the externalizing dimension on the CBCL. There are additional scales that also have both broadband and narrow-band sub-scales, such as the Conners Rating Scales-Revised (Conners, 1999).

There are several rating scales that assess behavior and mental health problems of children using the two broadband syndromes of internalizing-externalizing without the narrow-band sub-scales, such as the SAC (Glisson et al., 2002; Hemmelgarn et al, 2003). Similarly, there are numerous scales that are specifically tailored to measure narrowly defined behavior problems, such as the Children’s Depression Inventory (CDI; Kovacs, 1992), the Revised Children’s Manifest Anxiety Scale (Reynolds & Richmond, 1985), and

the ADHD Ratings Scales (Depaul, 1991). Breen and Fiedler (1996) and Ramsey, et al., (2002) provide further review of specific narrow band scales and broader measures.

Stability of Behaviors. There appears to be considerable debate as to whether or not broad measures are more appropriate in classifying children's mental health status than narrow-band measures. Research on the CBCL has been used to support the efficacy of narrow-band measures. For example, Achenbach and colleagues (Achenbach, Howell, Quay and Conners, 1991; McConaughy, Stanger, & Achenbach, 1992; Stanger, McConaughy, & Achenbach, 1992) studied a large ($N > 2,400$, depending on the study) nationally representative sample of children and found favorable results for the stability of the narrow-band syndromes. Specifically, these researchers found that the narrow-band stability coefficients (e.g., aggression, $r = .58$ and attention problems, $r = .57$) were comparable to the broadband stability coefficients (e.g., externalizing, $r = .59$). These stability coefficients were computed over a three-to-six year period, depending on the study. Moreover, from a practical perspective some researchers suggested that practitioners incorporate narrow-band measures when conducting school-based assessments (Eckert & Depaul, 1996) because these measures permit the investigation of specific aspects of behavior, which may not be possible through the use of broadband rating scales (Depaul, 1992).

In contrast to the data reported by Achenbach and colleagues (Achenbach et al., 1991; McConaughy et al., 1992; Stanger et al., 1992), Hartman, Hox, Auerbach, Erol, Fonseca, Mellenbergh et al., (1999) published recent evidence that suggest broadband classifications are more stable and more reliable than narrow-band syndromes, particularly as it relates to the cross-informant model of the CBCL and the TRF. In what appears to be one of the most rigorous evaluations on this issue to date, Hartman, et al., (1999) re-analyzed data

based on over 13,000 parent ratings and 8,800 teacher ratings from seven different countries. These researchers found that there was poor conceptual differentiation and little empirical evidence of the stability of the composition of narrow-band scales across countries and informants in both general population and clinical samples.

Another study (i.e., Mattison & Spitznagel, 1999) found important within-person differences between broadband and narrow-band classification schemes, which appear to highlight the advantage of using the broadband classification scheme. Mattison & Spitznagel (1999) found that both boys and girls identified as symptoms in one broad dimension on the CBCL at an early period in life can change narrow-band types (e.g., from anxious/depressed to withdrawn), while continuing to be classified within the same broadband syndrome (e.g., internalizing). This may in part be explained by the diffusion of item functioning with respect to the narrow-band syndromes of the CBCL found by Hartman, et al., (1999), which was presented above. Nonetheless, long-term stability and reliability of mental health profiles, however they are constructed, are important areas of assessment that practitioners and researchers must begin to address. More important to this study, research on ethnic differences in the stability of scales that measure internalizing and externalizing constructs is absent from the literature.

Overlap of Behaviors. It should be emphasized that many children exhibit both externalizing and internalizing problems and the combination of these problems is an important issue (Angold, Costello, & Erkanli, 1999). For example, Loeber and colleagues (Loeber, Stouthamer-Loeber, & White, 1999) re-analyzed longitudinal data across three samples of boys, ages 7 to 18 years-old and found that those with significant internalizing and externalizing symptoms were at greatest risk for persistent substance use. The limited

research on this issue represents a significant gap in knowledge. More important, research on ethnic differences in how the overlap of behaviors is expressed is completely absent from the literature.

CHAPTER III

ETHNIC GROUP DIFFERENCES IN ASSESSMENT

Ethnic Differences

One of the major underlying assumptions of this study is that ethnic differences between informants can differentially affect how they understand, interpret and respond to standardized questions on assessment instruments. There is a decade long program of research that suggest these differences exist and that one of the mechanisms that partially explains it is differential patterns in child rearing values, beliefs and practices (see Deater-Deckard & Dodge, 1997, for review). The argument is that parents from different cultural and ethnic groups understand and interpret their children's behavior differently and thus respond differently (i.e., use differential discipline practices) to behavior in a manner that affects their children's development.

A counter-argument to Deater-Deckard and Dodge (1997) is that ethnic-group differences in parenting and subsequent child development in this country (i.e., United States) are the exception and not the rule. Rowe (1997) suggests that studies that have found differences between groups from different countries/cultures on constructs are more common than the studies reporting cross-ethnic differences between children within the United States. Rowe also illustrates that many statistical differences across groups can reflect the "exquisite sensitivity of a statistic to minor statistical anomalies," such as those found in sample distributions, skewness, and kurtosis (p. 220). While this debate is far from settled, we do know that when ethnic-group differences in reporting child emotional and behavioral problems are found we need a method of determining how confident we can be that these differences are real and not artifacts of instrumentation or measurement bias. One possible

method is to examine the level of measurement equivalence of an instrument across ethnic groups (within the United States). The purpose of this study is to examine equivalence across AA and CA children who are similar in terms of family income levels and having been referred to juvenile court.

Defining and Assessing Test Bias

As with most complex discussions, defining the constructs of primary interest is a critical component because it helps to establish clarity. Although there are many goals of assessment research with different ethnic groups, this paper is primarily concerned with evidence of ethnic bias. For this reason, the construct of “test bias”, particularly in terms of test bias across different ethnic groups must be defined. However, disentangling the voluminous theoretical and practical debates on what exactly is meant by test bias is not a minor task (see van de Vijver & Poortinga, 1997, for a review). While there are several different approaches to understanding and interpreting bias, this study is primarily concerned with bias from a psychometric perspective.

What is Bias in Measurement? From a practical standpoint, bias is a term used to describe all nuisance factors that threaten the validity of our interpretation of scores or outcomes from a measurement tool (Goldstein, 1996; van de Vijver & Leung, 1997). Thus, measurement bias is a notion that is inextricably bound to questions of validity. Measurement validity is the extent to which an assessment tool measures what it claims to measure (Goldstein, 1996). Extending these concepts to our discussion on group differences, bias here is the extent to which behavior-rating scores used to compare children from different ethnic groups are not indicators of “real” differences. More specifically, bias in measurement suggests that differences in scores reflect differences in how the measurement tool is performing (i.e., how

the construct is measured) across groups and not differences in how groups are performing (Cooke, Kosson, & Michie, 2001). While there are many sources of potential bias, there are a few that are particularly relevant to this discussion and they are reviewed below.

Sources of Potential Bias. van de Vijver and Poortinga (1997) presented one of the most complete conceptual models of types of bias and their most common sources to date. These authors suggest that there are generally three major types of bias and each has specific sources that can be identified and addressed in research on (cross-ethnic) group differences. Specifically, there is construct bias (e.g., differential appropriateness of test across groups), method bias (e.g., error connected to the method used to obtain the responses), and item bias (e.g., incidental differences in the appropriateness of the item content) (see van de Vijver & Poortinga, 1997, for a complete review). While a complete review of these sources of bias is beyond the scope of this paper, it is necessary to describe some of the problems associated with each in order to gain further insight on the importance of examining instruments for their presence.

Construct bias is primarily a result of incomplete overlap of the definitions of the construct domain (van de Vijver & Poortinga, 1997), and this can take on several forms. For example, poor sampling of all relevant behavior across groups can result in incomplete overlap of the construct. Another problem is created when the instrument has differential appropriateness across groups. This can occur when certain behaviors do not belong to the repertoire of one or more of the groups being assessed. Method bias is typically caused by characteristics of the test or its administration. This can take the form of differential response styles such as acquiescence and extremity rating. For example, when Hispanic respondents tend to choose extremes on a 5-point rating scale more often than CA respondents it distorts

the results of cross-cultural analysis (Hui & Trandis, 1989). Finally, item bias refers specifically to measurement problems at the individual item level. Item bias is more commonly referred to as “differential item functioning” (dif) in contemporary literature. The term “item bias” is used here to describe how differential responses to items across ethnic groups jeopardizes the validity of cross-ethnic comparisons (van de Vijver & Poortinga, 1997).

It is relevant to note that construct bias is one of the most important types of bias in research on ethnic group differences because the presence of construct bias makes comparisons across groups inappropriate (Goldstein, 1996). “Cultural bias” is related to construct bias and needs further explanation. Cultural bias is the extent to which persons from different cultural backgrounds attach different meanings to the construct that a particular test is attempting to measure (Skiba, Knesting, & Bush, 2002; van de Vijver & Leung, 1997). Cultural bias is one of the main reasons for the need to conduct cross-cultural research and there are a few well-supported techniques used to detect and correct for cultural and related forms of bias (see Byrne & Campbell, 1999; van de Vijver & Leung, 1997). The most relevant of these techniques will be discussed in a later section.

Weisz and McCarty (1999) suggested that considerable caution should be used before trusting parent reports in research on ethnic differences in child psychopathology. Citing a decade-long series of cross-cultural comparisons (see Weisz, MacCarty, Eastman, Suwanlert, & Chaiyasit, 1997), these authors argued that parent reports of child problems may be influenced significantly by the cultural perspective the parent brings to the process, clouding interpretation of differences in parent-reported child behavior problems. While the research conducted by Weisz and his colleagues (Weisz & MacCarty, 1999; Weisz, et al, 1997)

primarily involved cross-cultural differences, they also argued that similar problems in interpretation could be evident in parent reports of children of different ethnic and cultural groups within the same country. For example, it might be that AA parents rate their children's behavior significantly different than CA parents would rate similar behavior in their children. This appears to be a problem that has not yet been closely examined.

Measurement Equivalence and Construct Validity

The question of measurement equivalence is a complex one. There are many methods that have been used to assess measurement equivalence, and these methods all have the common underlying purpose of establishing the validity of a measure across different groups. Knight and Hill (1998) argued that measurement equivalence is a question of the degree to which reliability and validity coefficients associated with a measure from ethnic minority samples are similar to those in the majority sample. Hui and Triandis (1985) categorized issues of measurement equivalence into several areas, which included item equivalence (i.e., items have the same meaning across ethnic groups), functional equivalence (i.e., scores generated by a measure have similar precursors, correlates across ethnic groups) and scalar equivalence (i.e., scores on a measure refer to the same intensity or magnitude across different groups). These three concepts are similar to item bias, construct bias and method bias respectively, which were discussed above. More importantly, Hui and Triandis (1985) suggested that multiple sources of evidence from various analytic procedures be examined before we form conclusions regarding the cross-ethnic or cross-racial equivalence of a measure.

Hui and Triandis (1985) were essentially concerned with the degree of differential validity associated with scores from a measure. Specifically, a measure would have to show

little or no evidence of differential validity in order for one to make the argument that it has measurement equivalence across ethnic or racial groups. Messick (1989) wrote one of the most comprehensive reviews on validity to date. Messick stated that, “Validity is an integrated evaluative judgment of the degree to which empirical evidence and theoretical rationales support the adequacy and appropriateness of inferences and actions based on test scores or other modes of assessment.” (p.13). Messick emphasized that it is the responses that a test produces that can claim reliability or validity and not the test itself. More important, validity is not only sample dependent, but also context-specific. In short, the compelling question surrounding validity is whether or not a measure can produce the same properties (e.g., psychometric properties) and patterns of relationships for different groups (e.g., ethnic groups) and under different conditions (Messick, 1989).

Methods Used to Establish Measurement Equivalence

Most writers discuss validity in categories. For example, the most common is criterion validity, which primarily consists of predictive, concurrent and postdictive types of validity. Criterion-related validity is established by the degree to which there are empirical relationships (in terms of correlations or regressions) between test scores and criterion scores (Messick, 1989). For example, a researcher might be interested in the relationships between externalizing scores and number or severity of juvenile offenses, using the juvenile offenses as the criterion scores. According to Messick, concurrent validity typically refers to the degree to which a set of scores correlate with or “estimates an individual’s present standing on the criterion” (p. 16). For example, a clinician or clinical researcher might want to know how well a child’s internalizing or externalizing score on the SAC correlate with the child’s diagnosis of depression or conduct disorder.

There are also predictive and postdictive types of criterion validity. Predictive validity is concerned with the degree to which scores can predict future criterion outcomes that are theoretically related to the construct that the test claims to measure. For example, in the case of juvenile offenses, a researcher might want to know how well externalizing scores predict future offenses and more specifically, future violent offenses. Conversely, postdictive validity is the degree to which scores can predict past criteria that are theoretically related to the construct that the test claims to measure. Therefore, using juvenile offenses as a criterion, a researcher might want to know how well externalizing scores predict past offenses and more specifically, past violent offenses.

Finally, there are two other types of validity that are important to understand when discussing construct-validation. Messick (1989) suggested that convergent and discriminant (or divergent) validity be established as part of a nomological net, or nomological validity (Cook & Campbell, 1979). This refers to how well the observed or empirical pattern of relationships between constructs fit a theoretical pattern. Essentially, measures of similar constructs should converge and measures of dissimilar constructs should diverge. For example, measures of externalizing behavior assessed by one scale should converge with another measure of externalizing (i.e., produce higher correlations), while measures of externalizing behavior should diverge from measures of unrelated constructs (i.e., provide lower correlations). It appears that researchers rarely test for ethnic differences in the extent to which measures demonstrate convergent and divergent validity.

IRT and SEM. Several analytic techniques are becoming increasingly popular among research scientists to assess the extent to which a measure can claim group equivalence (i.e., measurement equivalence across groups). IRT (Cooke, et al., 2001; van de Vijver & Leung,

1997; van de Vijver & Poortinga, 1997) and factorial invariance (i.e., SFA) procedures (Floyd & Widaman, 1995; Reise, Widaman, & Pugh, 1993; van de Vijver & Poortinga, 1997) are two prominent approaches currently available to assess measurement equivalence of items and scales. IRT is concerned with item-bias and can determine whether differences in scale equivalence is present. Some suggest that ruling out item bias is a necessary, but not sufficient line of evidence to make the claim of measurement equivalence (van de Vijver & Poortinga, 1997). Furthermore, IRT models (Harvey & Hammer, 1999) are among the most common methods of determining measurement equivalence in cross-cultural research (Drasgow & Hulin, 1987) and detecting item bias in psychological research (Harvey & Hammer, 1999), but have seen limited use in mutli-cultural research on behavior ratings scales.

Factor analysis has been improved by the use of new and advanced statistical software that has the ability to perform CFA procedures using structural equation modeling (SEM) (Floyd & Widaman, 1995; Merrell, 1999). Although early scale development efforts typically begin with exploratory factor analysis (EFA), subsequent efforts to address validity issues move into more precise procedures such as CFA. The distinction in practice between EFA and CFA, rather than being conceptualized as a strict dichotomy, can be thought of as that of an ordered progression (Anderson & Gerbing, 1988). The original factor model developed by Achenbach (1966) was derived by EFA, which basically is a procedure that allowed the items he used to freely load (i.e., correlate with) on an unspecified number of factors. After conducting EFA, the next step in this ordered progression would typically be to specify the number of factors that are expected to represent the data, but continue to allow each item to freely load on the specified number of factors. This procedure adds a constraint on the data

and has been referred to as “restricted” factor analysis (Anderson & Gerbing, 1988). CFA is a more stringent procedure than EFA. When using CFA the researcher must specify the number of factors, as well as the factor each item must load on. This allows the goodness of fit of the resulting solution to be tested (Anderson & Gerbing, 1988). SEM is a significant improvement in this area because it is particularly suited for controlling for error in latent constructs such as the broadband syndromes of externalizing and internalizing behaviors (Floyd & Widaman, 1995; Loehlin, 1998). More important to this review, SEM is also well-suited for establishing the degree of equivalence of measures across various groups (Floyd & Widaman, 1995).

Factorial invariance (Bollen & Long, 1993), is concerned with construct bias that some suggest should be ruled out to make the claim of measurement equivalence across different subgroups (Cooke, et al., 2001). A SEM technique called “simultaneous factor analysis” can provide simultaneous tests of the extent to which relations between factors, item-loadings, and item error variances differ across groups. Simultaneous factor analysis (SFA) and a similar procedure, multi-sample confirmatory factor analysis (MSCFA) (Keith, Fugate, Degraff, Diamond, Sharach, & Stevens, 1995), are well-suited for examining the measurement equivalence of rating scales and might be the most stringent tests for this question. While there are fundamental differences between IRT and SFA in terms of the kinds of questions each approach can answer, both methods attempt to address issues of measurement invariance (Reise, et al, 1993).

The review below will show that, in terms of the major ethnic groups within the United States, few behavior rating scales have been systematically tested for measurement

equivalence using IRT models or well-recognized SEM procedures, such as SFA (Floyd & Widaman, 1995).

Evidence of Measurement Equivalence of Behavior Rating Scales

Child Behavior Checklist. No behavior rating scale has been used, evaluated and cited in published research more than the CBCL. Concern about the appropriateness of the CBCL for AA children has some empirical merit. Lambert et al., (2002) analyzed the clinical records of 1,605 AA children seen in outpatient and inpatient settings in one state to determine how well their problems matched items of the CBCL. These researchers reported that a small to large portion of the sample reported 20 problems that failed to match any of the CBCL items and caregivers of 16 or less (i.e., < 1%) of the sample reported problems that matched 57 of the 118 items on the CBCL. Lambert et al., (2002) concluded that the CBCL might not provide adequate coverage of clinically relevant problem behaviors of AA children. Therefore, it would be important to examine the extent to which measurement equivalence of the CBCL has been demonstrated.

Latkovich (1996) conducted an unpublished study that investigated the internal and external validity of the CBCL for AA children. Latkovich reported that no previous study had directly investigated the measurement equivalence of the factor structure of the CBCL on a sample of AA children. The initial norms and psychometric properties of the CBCL were published in 1983 (Achenbach & Edlebrock, 1983). To confirm Latkovic's (1996) finding, a review of studies that examined the validity of the CBCL published between 1983 and September 2002 was conducted. The only study found was conducted by Latkovich (1996), which was a doctoral dissertation. Some of the findings in this study reportedly supported the factorial equivalence of the CBCL for the AA children in her sample. However, the statistical

analysis (i.e., Kaiser ratio; a congruence coefficient) used in this study was not the best method currently available to test for factorial invariance (Keith, et al. 1995). To date, it appears that there is no published research in peer-reviewed journals that examines the factor structure, factorial invariance or item equivalence of the CBCL for either AA or HA children.

Two points must be made before reviewing studies involving other child behavior measures. First, many studies have been published that reported various reliability and validity coefficients associated with a measure using samples that included children from diverse ethnic backgrounds, such as AA and HA children. These kinds of studies, while offering very valuable and useful information, do not address the essential question of measurement equivalence across groups that advanced procedures, such as SFA and IRT are specifically designed for.

Secondly, there have been many studies that have examined the factor structure of the CBCL using samples that included appropriate population proportions of children from diverse ethnic groups. In most cases, these studies consistently reported statistically significant results, replicating the original factor structure of the CBCL using the chi-square statistic and other well-recognized fit measures. However, using this kind of evidence to support the claim that a behavioral rating scale is equally valid for children from various ethnic groups because they were represented in the normative data may be misleading (Reid, Dupaul, Power, Anastopulos, Rogers-Adkinson et al., 1998). Norms established in this manner might be substantially influenced by the majority group.

Researchers studying group differences in measurement properties of assessment instruments, such as Reid, et al., (1998) and others (Arnold & Matus, 2000; Dana, 2000; Okazaki & Sue, 1995), argue comparisons of separate analyses for each group are needed in

order to better estimate the degree of equivalence of measures. Others conducting measurement equivalence research with assessment tools used in psychology go further to suggest that the most appropriate tests for this question are SFA (Allen & Oshagan, 1995; Cooke, et al., 2001; Reise, et al, 1993) and item analysis (i.e., Cooke, et al, 2001; Reise, et al, 1993). This view would be strongly supported by statisticians who conduct research in the area of SEM (Bollen & Long, 1993; Raykov & Marcoulides, 2000). Achenbach and his colleagues (Achenbach & McConaughy, 1997; Crijnen et al., 1997) and others (see Brown & Achenbach, 1996, for complete list of studies) have published studies examining the separate factor structures of the CBCL based on scores of youth from numerous countries. No study to date has been found making similar comparisons for youth from different ethnic and cultural backgrounds here in the United States. This is unfortunate because of the vast numbers of AA and other ethnically diverse children in the United States that have been assessed using the CBCL, and it represents a significant gap in the literature.

Other Behavior Rating Scales. There are six studies that we know of which appeared to have tested for race differences in the factor structure of a behavior rating scale, using samples of children in the United States (i.e., Epstein, March, Conners, & Jackson, 1998; Jarvinen & Sprague, 1995; Mayfield & Reynolds, 1998; O'Donnell, Stein, Machabanski, & Cress, 1982; Reid, et al., 2001; 1998). These studies used varying methods to test the factor structure of a very small number of scales. We will review these studies and assess to what extent we can make firm conclusions about the measurement equivalence of the scales involved in each study, as well as identify additional gaps in the literature.

O'Donnell, et al., (1982) compared the factor structure of a modified Problem Behavior Checklist (BPC; Quay & Peterson, *available from senior author*) for a sample of

Mexican-American (MA) (n=208) and CA (n=208) children. The modified BPC had 55 items and through EFA these researchers found five factors for both groups, although there were differences in the items and the number of items that loaded on each factor. The methods used to determine the degree of factor structure equivalence were Tucker's (1951) congruence coefficient and Cattell's (1949) salient variable index (s).

Only one of the factors (i.e., Anxiety-Withdrawal) was found to have high and statistically significant cross-ethnic matching ($r_c = .80$; $p < .001$) between the two groups (O'Donnell, et al, 1982). The remaining factors, while producing matching indexes that were statistically significant, were low. In addition to being low, some of the factors had correlations of the off-diagonal elements that were of comparable magnitude to the diagonal elements (off-diagonal elements should have lower correlations). These researchers correctly concluded that it appears as though some of the responses in the study may have an ethnically relative component.

Jarvinen and Sprague (1995) conducted a study to determine whether bias against ethnic minority children exists for the ADD-H Comprehensive Teacher's Rating Scale (ACTeRS; Ulmann, Sleator, & Sprague, 1991). The ACTeRS is a 24-item, 5-point Likert-type scale, ranging from 1 = *almost never* to 5 = *almost always*. These researchers used teacher informants to collect data on a sample of 2,277 elementary school students (i.e., grades K thru 8) with an ethnic composition that included 12% AA (n = 264) and 17% HA (n = 396) children.

The data analysis used by Jarvinen and Sprague (1995) was a modified Mantel-Haenszel procedure, which essentially is a differential item analysis used for performance tasks (see Zwick, Donoghue, & Grima, 1993). In this analysis, Mantel chi-square statistics,

degrees of freedom and associated p-values are reported for every scale item. These researchers concluded that no systematic item differences were found that favored either group. For every scale item that favored one group, a similar scale item was found that favored a different group (Jarvinen & Sprague, 1995).

One of the limitations of this study is that it did not include IRT methods, which are more rigorous item analysis procedures (Cooke, et al, 2001; Reise, et al, 1993). Secondly, these researchers did not attempt to examine the equivalence of the factor structure of the ACTeRS across the different ethnic groups in their study (i.e., SFA). Therefore, while the research found no systematic item bias using the ACTeRS (Jarvinen & Sprague, 1995), more knowledge is needed on whether or not this scale has factorial invariance across groups, as well as various other types of cross-ethnic validity.

Epstein, et al., (1998) analyzed the factor structure of the Conners Teacher Rating Scale (CTRS; Conners, 1973) to determine whether the CTRS factors are similar for AA and CA school children. The CTRS is a 39-item behavior rating scale that measures four factors or constructs related to classroom behavior problems (i.e., conduct problems, hyperactivity, anxious/passive, and social problems). Epstein, et al. first conducted separate EFA for CA males (n=292), AA males (n=187), CA females (n=286), and AA females (n=198). These researchers then compared factor structures derived from the EFA analyses across groups (i.e., CA males vs. AA males and CA females vs. AA females) using two measures of factor congruence: congruence coefficients (Tucker, 1951) and the salient variable similarity indexes (Cattell, 1949).

According to Epstein, et al., (1998), their results supported the use of one set of norms for the CTRS for both AA and CA school children for two reasons. First, there were

high statistical estimates of congruence and item content across similar factors for males across ethnic groups. Secondly, for females, there were high statistical estimates on both tests, but only for factors related to conduct, social functioning and anxiety. However, there were important differences in the factor structure within each sex-comparisons. Specifically, an Antisocial factor emerged for AA males and an Inattention factor emerged for CA females.

Further examination of this study reveals a somewhat different conclusion regarding the similarities of the factor structures across ethnic groups. First, the statistical test conducted to determine factorial invariance of the CTRS for AA and CA children was less than optimal (Okazaki & Sue, 1995). Congruence coefficients (Tucker, 1951) are essentially correlation coefficients that compare two sets of factor loadings, which is an acceptable method, but some researchers have suggested this method does not meet the standards of and are far less direct than the more rigorous SFA (see Keith et al., 1995). Secondly, the emergence of a fifth factor for AA males (i.e., Antisocial) and CA females (i.e., Inattention) suggest that considerable caution should be used when drawing inferences regarding the interpretation of group differences in scores on the CTRS.

The decision to use EFA instead of confirmatory factor analysis (CFA) raises additional concerns about these results. It is considered more appropriate to use CFA when assessing factor models that have been well-established in prior analyses (Loehlin, 1998), as is the case with the CTRS (see Epstein, et al, 1998, for review). In sum, there is agreement with Epstein et al., that the study's results highlight the need for consideration of ethnic differences when using the CTRS and that further research is warranted to determine whether mean differences in ratings are real or artifacts of systematic bias.

Mayfield and Reynolds (1998) conducted an elaborate study to evaluate “Harrington’s Hypothesis” (Harrington, 1984) by testing race differences in the psychometric properties of the Parent Rating Scale (Adolescent form; PRS-A) from the Behavior Assessment System for Children (BASC; Reynolds & Kamphaus, 1992). Harrington (1984) stated that test items are selected based on characteristics of the total sample and are primarily affected by the numerical majority of the sample (most often Caucasians), therefore favor the majority group. Harrington hypothesized that the majority group would test better than any other group on these kinds of tests. While Harrington’s argument involved aptitude tests and was based on laboratory experiments with rats, Mayfield and Reynolds (1998) tested this hypothesis with a behavior rating scale.

The data reported by Mayfield and Reynolds (1998) was part of the second of three item tryouts for the BASC. As part of the second item tryout, to test Harrington’s hypothesis, three separate tests were developed, each with 100% representation of Black (n=267), Hispanic (n=64) and White (n=1,387) parents. Each test included 10 subscales (Leadership, Social Desirability, Social Skills, Anxiety, Aggression, Conduct Disorder, Depression, Hyperactivity, Atypical, Somatization) and had a total of 114 items. While each set of 114 items came from the same pool (double in number) of items, only the top 50% of items selected in the tryout were retained for each group. The first test was proportion of item overlap, because in theory, each test would look (face validity) significantly different for each ethnic group. In Harrington’s (1988) most recent study (of six strains of rats) approximately 25% items overlapped between tests. The proportion of item overlap, via pair-wise comparisons, in the three tests constructed by Mayfield and Reynolds (1998) averaged

69% (range = 38% - 87%). Therefore, the first set of results does not appear to support Harrington's hypothesis.

Mayfield and Reynolds (1998) administered the three tests to all three groups (accomplished through rescoring of data already collected) and examined the overall performance of each group on the alternative tests. The reliability coefficients were comparable among 89 of the 90 comparisons. Only the white-Hyperactivity scale had significantly higher internal consistency for whites than it did for blacks. Moreover, score distributions on all three tests were compared across groups and no distinctive patterns favoring one group over another were found. Therefore, these researchers concluded that the differences found were due to true differences in the distribution of the measured behaviors. The final analyses of these data were tests of mean differences of each group on the subscales of the alternative tests using one-way ANOVA's. The results did not show that different ethnic groups scored better on their subscales than either of the other two ethnic groups. Therefore, Harrington's hypothesis that the majority group would score better on the test that reflects their majority influence was rejected.

There are several reasons why these results must be interpreted with caution. First, while the statistical analyses employed by Mayfield and Reynolds (1998) were fairly elaborate, they do not meet the stringent requirements for testing factorial invariance found in SEM methods (i.e., SFA). It is of some concern that the three ethnic groups did not actually respond to three separate scales. Rather, the study reports that the three groups here were created "through rescoring of data already collected" (p. 323). Additionally, there were significant sample size differences between groups. In this study 64 Hispanics were compared to more than 1,387 Whites, and the black sample was not much larger (i.e.,

n=267). However, a counter argument would be that given the small sample size and the robustness of the results, the sample size issue is one of the study's strengths. Nonetheless, a sample size should, by liberal estimates, have five times the number of items when conducting psychometric analysis (Arbuckle & Wothke, 1999). This would have required a minimum sample size of 570 per group. While the efforts of Mayfield and Reynolds (1998) are impressive, the concerns raised suggest no firm conclusions can be made regarding the factor or item equivalence of the PRS-A.

Reid et al, (1998; 2001) conducted two studies examining two different scales that measure Attention-Deficit Hyperactivity Disorder (ADHD) in children. Reid et al., (1998) used SEM and multi-dimensional scaling (MDS) to test the factorial invariance of the ADHD-IV Rating Scale-School Version (Depaul, et al., 1997) based on teacher ratings of 381 AA and 1,259 CA males. The ADHD-IV consists of 18 items taken directly from the symptom list specified in the DSM-IV.

Reid et al, (1998) appears to be the first study reviewed here that used the appropriate test of factorial equivalence. These researchers reported degrees of freedom (df), χ^2 (as well as difference statistics for each), and several conventional fit indices (e.g., GFI = goodness of fit; RMSEA = root mean square error of approximation; SRMR = standardized root mean square residual). The simultaneous test of model fit across AA and CA male children in the study revealed that the GFI = .70, RMSEA = .101, and SRMR = .11, and suggested that there was some degree of difference in the model fit across groups. Although not stated by these researchers, another possibility is that neither sample fit the model. The MDS analysis revealed that of three dimensional solutions found for both groups, only two were found to have high congruence coefficients (i.e., AA= .989, .905, .760; CA= .991, .921, .824). And of

the two dimensions, only one (i.e., .986, .388) resulted in a high and significant congruence coefficient across groups. According to these researchers, the results of the MDS analysis also suggest that there are item dissimilarities for Dimension 2. Overall, these results highlight the possibility of differences in teacher's perceptions of child behavior between AA and CA youth because the same teachers rated both groups of children.

Reid et al, (2001) assessed the construct and normative equivalence of the IOWA Conners Rating Scale (IOWA; Pelham, Milich, Murphy, & Murphy, 1989) with a sample of AA (n= 2,124) and European-American (EA) (n=1,874) children, as well as within-group gender differences. The IOWA has two subscales (i.e., Inattention/Overactivity and Aggression), with five items per subscale. Only the test of construct equivalence was important for this discussion. To test construct equivalence these researchers used CFA to compare separate model fit indexes for each group and within-group analyses. The fit indexes were, GFI = .99 and above (i.e., EA = 1.00), AGFI (adjusted goodness of fit) = .99 for all groups, RMSEA < .69 for all groups, and CFI (confirmatory goodness of fit) = .99 and above (i.e., EA = 1.00). Therefore, these researchers concluded that the results of this study showed that there was construct equivalence of the IOWA across AA and EA groups for both boys and girls. However, these researchers did not directly test for factorial equivalence. They should have taken the next step and constrained the item loadings, factor correlations, and possibly uniqueness to be equal which is the appropriate, simultaneous test of model fit for both groups. Nonetheless, these data appear to be in the direction of adequate fit for both groups. Future measurement equivalence analysis with the IOWA should include multiple models, with increasing constraints in order to conduct a simultaneous test of model fit for different groups.

CHAPTER IV

REMAINING GAPS IN THE LITERATURE

A large majority of children who are referred to the child welfare and juvenile justice systems in the United States have mental health problems. Children served by child welfare and juvenile justice systems are serviced by social workers and other mental health professionals who need assessment tools to help them make proper referrals and placement decisions, as well as monitor their progress. Therefore, assessment tools that have practical utility and validity across different age ranges, ethnic groups, and genders are needed by these systems to improve the quality and outcomes of services. It is against this backdrop, that several important conclusions regarding the mental health assessment of children coming into contact with child welfare and juvenile justice systems can be made from this review.

One outcome of research in this area is that through the use of empirically-derived, standardized behavior rating scales, we are now better able to identify children who are in need of mental health services. However, more information is needed about the specific advantages and disadvantages of these types of scales and effective strategies for their application in actual service systems. Knowledge in this area could come from studies that compare the practical utility of categorical assessment tools, such as the newly developed Diagnostic Interview Schedule for Children (DISC; Shaffer et al., 2000) with dimensional assessment tools (e.g., CBCL; SAC). In what might be the only study to date that directly compared the DISC to the CBCL, Jensen et al., (1996) reported mixed results. They concluded that although these two assessment approaches were reasonably comparable, the diagnostic categories of the DISC showed fewer relationships with external validators than the CBCL. Without additional validation studies that compare categorical instruments with

dimensional instruments, there will be little agreement and less clarity regarding which behavioral assessment approach should be used and why. This represents a significant gap in the literature and the information that can be derived from these types of studies is needed.

More research is needed on the strengths and weaknesses of using broadband versus narrow-band rating scales and which of these models of assessment has greater utility in child welfare and juvenile justice systems. It might be that narrow-band scales are too narrow in scope and do not capture the wide-variety of known mental health problems exhibited by children who come into contact with child welfare and juvenile justice systems (Glisson, 1996; Teplin et al., 2002). More importantly, it is rare that children in these service systems are affected by a narrowly defined problem and more often the case that these children struggle with multiple symptoms. Nonetheless, future research designed to address these kinds of questions would be important in the advancement of this body of knowledge.

Within this debate there is also a need for further clarity on the advantages and disadvantages of simple structured broadband scales (e.g., SAC) versus more complex multi-problem scales that separate broad dimensions into narrow-band subscales (e.g., CBCL). Based on what we know about the nature of the work (e.g., overwhelming caseloads, stress levels, and time constraints) in child welfare and juvenile justice systems, assessment tools should meet certain criteria. Specifically, assessment tools used in these service systems should be brief, easy to use, easy to interpret and have wide applicability (e.g., used across many age groups, both genders, with a variety of informants, and with children from diverse ethnic and racial communities). Moreover, an assessment tool that has a simple structure, and yet does not forfeit comprehensiveness (i.e., can capture various types of emotional and behavioral problems), would be ideal for use in child welfare and juvenile justice systems.

The SAC is one such instrument, in that it has been shown to capture the broad internalizing and externalizing problems that children present using substantially fewer items than the “gold standard” of child behavior assessment, the CBCL (i.e., 48 vs. 118, respectively). This comprehensive, yet simple factor structure was replicated across boys and girls, and for different age groups, using teachers and parents as informants.

Scant research has directly compared the utility of broadband scales that have a simple structure with that of multi-problem broadband scales in child welfare and juvenile justice systems. In a rare exception, Hemmelgarn et al., (2003) conducted a validity study of the SAC by comparing it to the CBCL and TRF (Achenbach, 1991a; 1991b) using a sample of children in one state’s child welfare and juvenile justice systems. These researchers found that the brief, 48-item SAC performed as well as the CBCL and the TRF in providing convergent, divergent, and predictive validity evidence for the broadband constructs of internalizing and externalizing. Studies such as this one are important because it includes real world constructs and outcomes that are relevant to children at risk for long-term care and serious mental health problems. However, additional studies that directly compare the simple structure model to the complex structure model are needed for us to have greater confidence in these findings. Furthermore, this research must include efforts to examine whether certain types of assessment tools (i.e., SAC versus CBCL) are more appropriate (and for what reasons) for some ethnic groups while less useful with others.

Completely absent from the literature is cross-ethnic comparisons in how these assessment tools perform in systems that serve children who are more at risk. This represents a significant gap in child mental health assessment research and is particularly troublesome

given that child welfare and juvenile justice systems serve disproportionate numbers of AA and HA children.

Because many children are at risk for long-term, severe mental health problems and multiple placements in child welfare and juvenile justice systems, these assessment tools must be rigorously examined for their utility as monitoring tools. Although there is some disagreement, broadband scales appear to be more stable over time, in terms of consistently measuring emotional and behavioral (i.e., internalizing and externalizing) problems in the same child. Previous research has suggested that narrow-band scales are less stable than broadband scales, but it is not clear if simple narrow-band scales (e.g., CDI) are more stable than complex scales that combine narrow and broad constructs (e.g., CBCL). It is also less clear whether or not broadband scales that use the short and simple structure model (i.e., SAC) and multi-problem scales that are longer and combine broad and narrow-band subscales (i.e., CBCL) are more stable over time. The best research method to answer these kinds of questions is the longitudinal design. In order for us to move the literature forward, there is a need for longitudinal research that addresses these issues.

Research is also needed to disentangle internalizing and externalizing problems to understand the unique and shared contributions of each construct to treating child behavioral and mental health problems. Specifically, examining and comparing summed scores on the two constructs or separate scores might be misleading. A child that is clinical on the externalizing dimension and non-clinical on the internalizing dimension might have an overall total score on a combined behavioral problem scale that is comparable to a child that is clinical on the internalizing dimension and non-clinical on the externalizing dimension.

Yet, these two youth are substantively different in terms of how they display problems and thus, should receive very different mental health services.

With few exceptions (i.e., Farmer & Bierman, 2002; Loeber, et al., 1999), the study of the overlap of internalizing and externalizing problems is absent from the literature. Preliminary results from the limited research in this area suggest that it might be more appropriate to examine and compare various combinations of “aggressive-withdrawn” behavior profiles (Farmer & Bierman, 2002) because this approach may provide significantly more information than either combining subscale scores or isolating one or the other. More specifically, accurately assessing behavior profiles might enable practitioners to identify specific needs and tailor interventions to meet those needs, such as identifying which profiles are more likely to predict a substance abuse problem (Loeber et al., 1999). Finally, no study has explored how these behavior profiles might differ across various ethnic groups in the United States. These are important issues and deserve greater attention in the literature. The differences between externalizing and internalizing youth and youth who have either or both characteristics, and the possible age, gender and ethnic differences that exist need further discussion and explanation.

The final area of research on these assessment tools that has received little attention is measurement equivalence studies across different ethnic groups in this country. The precise information on measurement equivalence that can be obtained examining the factorial invariance of assessment tools can provide specific information about response patterns between ethnic groups. Yet, few have tested factorial equivalence of behavioral rating scales for AA children. Moreover, few studies have examined the criterion validity of behavioral rating scales for children in this country from diverse ethnic groups. The wide-spread use of

these types of scales suggests that this gap in knowledge represents a significant limitation of the overall body of literature in this area.

We do have examples of behavioral rating scales that have been tested vigorously (e.g., Reid et al., 1998) and some less vigorously, for the degree of factor equivalence with AA and HA children. Moreover, there are also important studies on factorial equivalence of intelligence test (e.g., Keith et al., 1995) found in the psychological literature that further highlights the importance of this type of validity evidence. Because some researchers have made attempts to address this issue, the validity and equivalence of measures with children from different ethnic groups within the United States is an important question and deserves considerably more attention than currently found in the literature.

Summary and Rationale for the Present Study

While several gaps remain in research on behavior rating scales in particular and child mental health assessment in general, the integrity of the psychometric properties of assessment measures across different ethnic groups is of paramount concern. Specifically, the absence of cross-ethnic studies on how well behavior-rating scales perform in systems that serve disproportionate numbers of children from diverse ethnic groups is unsettling. At the most fundamental level, there is a significant gap in the literature on behavior rating scales in terms of our knowledge of the degree of factorial and item equivalence of these measures across different ethnic groups. Secondly, it appears that most studies that have attempted to test for this important difference have been less than optimal. With limited exceptions (i.e., Reid et al., 1998), most studies have not appropriately used SEM or IRT models in their data analyses. This is unfortunate and represents a significant weakness in this area of research.

Disentangling these gaps in the literature is an important step in guiding future directions of research in this area.

In summary, along with comparing various reliability and validity coefficients, studies on the psychometric properties of new behavior rating scales should include the analyses of measurement equivalence of the constructs across the major ethnic groups in the United States. More important, it would be useful to re-examine existing scales that are frequently used in psychological research, as well as research and practice in child welfare, juvenile justice, educational systems.

CHAPTER V

THE PRESENT STUDY

The primary issue in the present study concerns differences in reporting of behavioral and emotional problems between CA and AA children. One theory is that these differences are real world differences and practitioners, service system providers, and child mental health policy makers should not be concerned with differential application and interpretation of assessment instruments. McLoyd (1990) suggests that the disparity in incidence and reporting of behavioral outcomes is partly explained by the numerous social, cultural, and educational disadvantages that disproportionately impact AA. Moreover, Reid et al, (1998) extends this argument by concluding that historical disadvantage has overexposed some groups to prenatal risk factors, psychosocial stressors and economic hardships, which in turn have negatively affected behavioral and mental health outcomes of children.

However, there is an alternative explanation for ethnic differences in outcomes from behavioral rating scales. It might be that these scales produce biased estimates of one group's behavioral and mental health problems. This might manifest in overreporting or underreporting of behavioral and mental health problems. There is a long history of research on test bias in the United States and studies detecting bias have led to substantial improvements in educational testing and to a lesser extent, psychological testing (see Berk, 1982 for review).

One of the explanations for past evidence of bias in measurement tools is that developers of these measures did not adequately include measurement equivalence analyses in the early stages of development.

It is necessary to establish conceptual and normative equivalence across groups during the development of an instrument if it is to provide valid information for members of each ethnic group (Marsella & Kameoka, 1989). Conceptual equivalence is established when constructs with similar conceptual meaning across groups are included in a scale. Normative equivalence is established when normative standards are developed that are appropriate for each group. Lambert et al., (2002) suggested that developers of rating scales appear to have neglected important ethnic and cultural differences between AA and CA in terms of their values and family practices. This has led to inadequate representation of those behaviors that are most troubling to AA parents. Lambert states that “uncooperative behavior” was the most prevalent problem reported by AA parents when they referred their children for services, and yet, this item is absent from the most widely used behavioral rating scales in the field. Inappropriate behavior within the context of the family and the community might be less tolerated in the AA community than that of the CA community (see Lambert et al., 2001 for review).

The above arguments suggest that explaining differences in reporting of behavioral and emotional problems between CA and AA children remains an important issue. It is clear that rigorous psychometric studies would help to advance knowledge of the differential validity of behavior-rating scales used in research and practice. This current gap in knowledge limits our confidence that widely used scales in research on child mental health are providing information that is accurate across different age groups, genders and children from diverse ethnic groups.

The increasing numbers of ethnic minority children referred to child welfare and juvenile justice systems suggest that these systems are logical starting points for conducting

evidenced-based research on cross-ethnic equivalence of rating scales. Research on the equivalence of measures using samples of children referred to child welfare and juvenile justice systems would further our knowledge regarding the utility of behavior-rating scales for service providers in these systems. The purpose of the present study is to meet this critical need by comparing responses from a behavior rating scale from samples of AA and CA children referred to juvenile court in one state. In this study, the measurement equivalence of the SAC (Glisson et al., 2002; Hemmelgarn et al., 2003) will be tested using a comprehensive set of statistical analyses.

The SAC was selected for analysis because it was developed specifically to meet the needs of providers in child welfare and juvenile justice systems. According to providers, the SAC is easy to use and easy to interpret (Glisson et al., 2002). Moreover, previous studies have provided extensive evidence documenting the reliability and factor, criterion, convergent and divergent validity of the SAC (Glisson et al., 2002; Hemmelgarn et al., 2003). Knowledge gained in this study will add to the existing evidence of the utility of the SAC, particularly in terms of its reliability and validity across different ethnic groups within child welfare and juvenile justice systems.

Research Questions

The following research questions will be examined in this study to forward knowledge on ethnic differences in using behavior rating scales to measure the emotional and behavioral health of children referred to child welfare and juvenile justice systems.

Research Question 1. Are behavior rating scales equally reliable and valid for/in measuring the emotional and behavioral health of both CA and AA children referred to child welfare and juvenile justice systems?

Research Question 2. Is the internal-consistency reliability of one behavior rating scale, the SAC, similar for both CA and AA children referred to child welfare and juvenile justice systems?

Research Question 3. Is the test-retest reliability of one behavior rating scale, the SAC, consistent across CA and AA children referred to child welfare and juvenile justice systems?

Research Question 4. Are the intercorrelations of the SAC and the convergent-divergent validity of the SAC with the CIS similar for both AA and CA children referred to child welfare and juvenile justice systems?

Research Question 5. Is criterion validity of one behavior rating scale, the SAC, equivalent across CA and AA children referred to child welfare and juvenile justice systems?

Research Question 6. Is the factor structure of the SAC equivalent across CA and AA children referred to child welfare and juvenile justice systems?

General Research Hypothesis

This is an exploratory study, but few differences in the psychometric properties of the SAC across AA and CA children are expected for two important reasons. First, the behavioral profiles of children of different ethnic groups referred to juvenile justice and child welfare systems are likely to be more similar than dissimilar. Nonetheless, the critical issue here is whether or not the informants (i.e., primary caretakers) providing the data on these children *perceive and report* the children's behavior in similar ways. Secondly, the ability of the SAC to successfully identify broadband behavior problems of children should allow for small, individual differences (e. g., expressing different types of externalizing behaviors) while accurately measuring overall mental and behavioral health.

Therefore, a general hypothesis will be tested in this study. It is expected that the pattern of results from the analyses described below will be more consistent with the notion that there is measurement equivalence for the SAC across AA and CA children sampled in this study.

CHAPTER VI

METHOD

Design

This study examines data from a large NIMH funded study that includes the first two waves of a longitudinal, three-panel survey. A prospective research design was used to identify and follow a sample of children referred to juvenile court in 20 Tennessee counties. The design includes a purposive sample (i.e., children referred to juvenile court who were at risk of state custody) and the initial data collection period extended one full year. The criteria used to select children for the larger study were that they (a) were not in state custody at the time of the initial interview, (b) were at risk for state custody according to the juvenile court youth service officer (YSO), and (c) could not have a sibling in the study. Children who were in court for reasons that did not place them at risk of state custody (e.g., child support issue) were excluded from the study. Roughly 30% of the children referred to juvenile court did not meet the criteria for the study. Three of the 20 counties represent the state's major urban communities (i.e., Hamilton, Knox, and Sevier), and the remaining 17 counties ranged from rural to semi-rural.

The data collection period and type (rural and urban) and number of counties included in the study suggest that this sample is representative of all children referred to juvenile court in Tennessee. The design provides information about the emotional and behavioral health needs of children both prior to their court referral at the time of the baseline interviews and during the first follow-up period (of at least six months).

Participants

The participants in this study include the African-American (AA) (n=562) and Caucasian-American (CA) (n=692) children who met the criteria listed above. Children of other race and ethnic backgrounds were excluded from these analyses (3.8%). There were slightly more males (65.3%) than females (34.7%). The ages of the children range from five to 18 years ($M = 14.5$, $SD = 2.4$) and they are primarily from low-income families, with a large proportion of these children living in families with incomes that are below the poverty level. Complete demographic data are presented in Table 1 (all tables are found in Appendix A).

Approximately 65% of those children and families, who met the criteria for the study and were asked to participate, agreed to be interviewed. Although no data was collected from those who declined to participate, the characteristics of the larger sample are similar to those children referred to juvenile courts in the state of Tennessee (Tennessee Council of Juvenile and Family Court Judges, 2002). For example, 81 % of the children referred to juvenile court in Tennessee were referred for reasons that place them at risk for custody, 60% were male, 61 % were CA, 42% lived in urban areas and the mean age was 15 years according to the data for 2001. In this sample, slightly more than 70% of those referred to juvenile court were identified as being at risk for custody, 65% were male, 55% were CA, 60% lived in urban areas and their mean age was 14.5 years.

Human Subjects Review and Confidentiality. The study was approved by the University of Tennessee Institutional Review Board (IRB #: 4395B) and participation in the study was voluntary. Subsequently, this study was approved by the Human Subjects Committee of the College of Social Work at the University of Tennessee.

Data Collection

Research assistants (RA's; including the author) were trained by the Children's Mental Health Services Research Center (CMHRC) and placed on site in each of 20 juvenile courts to collect the data used in this study. With the help of juvenile court case managers, these RA's identified potential participants and obtained written consent from the children's caretakers (and from those children over 10 years of age) for involvement in this study following protocols approved by The University of Tennessee IRB. After obtaining informed consent, the baseline (panel 1) interviews were conducted using the child's primary caretakers as informants and took place at the time the child was initially referred to juvenile court. The second wave of data (panel 2) was collected at least six-months following the initial interview to conduct predictive criterion validity analyses and test-retest analyses.

Although few researchers report the ethnicity and gender of the RA's collecting data, McLoyd (1998) argues that this kind of information can be important. McLoyd suggested that researchers conducting studies that compare groups based on race or ethnicity, or examines issues of race or ethnicity should report the race and gender of the RA's conducting interviews or collecting data. In this study there were 14 RA's, and all but one were female. Three RA's were African-American, one was Hispanic-American and one identified herself as biracial.

Measures

The RA's gathered data during the initial interviews and follow-up using a variety of instruments. Only data collected from four instruments are the focus of this study. These four instruments are described below. The first instrument discussed below was developed by the CMHSRC primarily for the purposes of obtaining demographic and background information

on the children included in the study. The rest of the instruments are standardized measures that have been used in a number of previous studies and were all developed with funding from NIMH.

Family Interview Questionnaire (FIQ). The FIQ is a non-standardized interview form that was developed by the CMHSRC for the purposes of gathering demographic information on the children (e.g., age, race, and gender) and families (e.g., monthly income and family structure). There were two versions of this form, an initial interview version and a follow-up version. The initial interview version of the FIQ provides descriptive data for this study, and background information used in assessing the postdictive-criterion validity of the SAC. The follow-up version will be used to assess the predictive-criterion validity of SAC scores. Most relevant to this study were items describing the child's prior offenses, substance-abuse history, previous custody and reason for referral. Operational definitions of each postdictive variable included in this study that were taken from the FIQ are listed in Appendix B and operational definitions of each predictive variable included in this study that were taken from the FIQ are listed in Appendix C.

Shortform Assessment for Children (SAC). The SAC is a 48-item standardized measure used to assess the overall mental health of children that has been validated in previous studies (Glisson, Hemmelgarn & Post, 2002; Hemmelgarn, Glisson, and Sharp, 2003) and is the primary focus of this study. The SAC was developed by the CMHSRC with the support of the NIMH for the specific purpose of assessing children referred to child welfare and juvenile justice systems. This measure has a three-point (*0 = Never, 1 = Sometimes, and 2 = Often*) behavioral rating scale format. The SAC includes two broadband

scales. One scale has 24 items that describe internalizing behavior and the other has 24 items that describe externalizing behavior (see Appendix D for complete scale).

Separate norms were established for the SAC for girls and boys, pre-adolescents and adolescents, and for using parents and teachers as informants. The SAC takes 10 minutes to complete. Its brevity, ease-of-use, and little training requirements to administer, score, and interpret makes the SAC particularly useful in service systems, such as juvenile justice and child welfare (Glisson et al, 2002).

Columbia Impairment Scale (CIS). The CIS is a 13-item standardized measure of children's overall mental health that has been validated in previous research (Bird, 1999). This measure has been particularly useful in identifying children with clinical levels of mental health and behavioral problems that require mental health services and will be used to assess the degree of convergent-divergent validity of the SAC scores for both AA and CA children in this sample. Similar to the SAC, the CIS provides an overall measure of child mental health, but unlike the SAC the CIS does not include internalizing and externalizing problem scales. The response format is *0 = No problem, 1 – 3 = Some problem, and 4 = A very big problem* (see Appendix E for complete scale).

Services Assessment for Children and Adolescents (SACA).

The SACA is a multi-component measure that is used to describe and monitor a variety of social, behavioral and mental health services received by a child or adolescent and has been validated in previous studies (Hoagwood, et al., 2000; Stiffman, et al., 2000). The SACA has been used to obtain information from caretakers about the types of services received by the children and adolescents in their care. In previous studies, this measure has provided accurate and valid descriptions of mental health services using caretakers as

informants that are similar to those used in this study (Hoagwood, et al., 2000; Stiffman, et al., 2000). The length of this instrument, in terms of number of items varies from the 25-item initial services form to the 5-item final services module. Not all items from this measure were used in this study. The items that were selected were those that would provide the best information regarding the postdictive (baseline/initial interview) and predictive (6-month follow-up) criterion validity of the SAC. Operational definitions of variables used in this study that were adopted from this measure and their respective purposes (i.e., postdictive or predictive criterion), are presented in Appendix F.

Data Analyses

Descriptive Data Analyses. Descriptive statistics for the total sample and each group (i.e., AA and CA children) including frequency distributions of all relevant child and family variables, as well as mean total and sub-scale scores on all relevant measures used in this study are reported. Additionally, because the SAC is the measure of primary interest in this study, the means, standard deviations, standard errors, and test of significant differences across groups are reported.

Reliability Analyses. Two types of reliability analyses were conducted in this study. First, internal consistency-reliability analyses (Chronbach's alpha) were performed on SAC data for the total sample and separately, for both the AA and CA children in this sample. This reliability analyses were conducted using the alpha reliability procedure in the SPSS statistical software package.

Secondly, test-retest reliability analyses were performed on the SAC scores for the total sample and for both the AA and CA children separately to determine the stability of the SAC across groups. Specifically, the bivariate correlations (i.e., Pearson's r) procedure in the

SPSS software was used to describe the correlations between the SAC scores obtained during the initial interview and the SAC scores obtained at follow-up.

Intercorrelations and Convergent-Divergent Validity. It is a common practice for researchers to examine intercorrelations of scales and subscales (or factors) when assessing the overall validity of a measure. Subscales should be more highly (i.e., strongly) correlated with total scale scores and less (i.e., moderately) correlated with other subscales. In this study it was expected that total SAC scores would be more highly correlated with externalizing and internalizing scores than externalizing and internalizing scores would be correlated with each other, and that this pattern of relationships would be consistent across groups. Therefore, the AA and CA groups' intercorrelations of the SAC and its two factors were compared. These correlations were interpreted using Cohen's (1992) standards for effect sizes (i.e., small = .10, moderate = .30, large = .50).

Additionally, the CIS scores of both groups of children were correlated with the SAC scores of both groups and these correlation coefficients were compared across groups. These correlations are the equivalent of convergent-divergent validity coefficients because the CIS was treated as another measure of child emotional and behavioral problems. Note that the majority of the items on the CIS (see Appendix C) reflect externalizing or aggressive behavior problems. Therefore, it was expected that the CIS would be more highly correlated or converge with both the overall (i.e., total) SAC scores and externalizing SAC scores, and only moderately correlated or diverge with internalizing SAC scores.

Criterion-Validity Analyses. For the subsequent analyses, three sets of SAC scores for each subgroup and the total sample were used as independent variables. Specifically,

SAC total scores, SAC internalizing scores and SAC externalizing scores were used to conduct criterion validity analyses.

The correlation between a measurement outcome (i.e., score) and a criterion variable, whether measured concurrently, postdictively (i.e., occurred prior to obtaining the score), or predictively (i.e., occurred sometime after obtaining the score), is typically referred to as a validity coefficient (Reynolds, 1982). A validity coefficient (i.e., r_{xy}) is a direct measure of the strength and magnitude of the relationship between two variables. The two observed validity coefficients were compared to determine whether or not they were statistically different. Cohen & Cohen (2003) suggests that when comparing two (independent) Pearson product-moment correlation coefficients, an additional test should be conducted to test whether or not the observed coefficients are statistically different. Cohen and Cohen suggested that each correlation be converted to a Fisher's z by using an r -to- z table. After computing the standard error for the difference between two independent Fisher's z (Φ_{dz}), the following equation can be used to compute a standard score, $z = z_{r1} - z_{r2} / \Phi_{dz}$ to test the significance of this difference.

Postdictive–criterion coefficients were calculated by conducting bivariate correlations (i.e., Pearson's r) between SAC scores and data from several variables. First, SAC scores were correlated with various types of offenses (e.g., status offenses, violent offenses, etc.). This information came from data collected using the FIQ. Appendix A presents each type of offense and its definition. Also taken from the FIQ are educational and mental health categories that denote behavioral problems (e.g., ADHD, SED, etc.), as well as any mental health medication that a child might have taken. These are also listed in Appendix A. Secondly, SAC scores were correlated with items on the SACA that denote spells (i.e., stays)

in any type of custody (i.e., residential placement, group home, and juvenile detention) and various types of mental health services that took place prior to the initial assessment. The variable labeled “custody,” indicated whether a child had a spell in any of these placements with a value of “1.” If a child was not placed in custody, the child received a value of “0” for this variable. The various types of mental health services that were used as criterion variables in this study are listed in Appendix D. It is expected that SAC scores will be positively correlated with custody, offenses, mental health services and medication use. Additionally, the Fisher’s z was used as described above, to test whether or not the observed postdictive-criterion validity coefficients of both subgroups are statistically different from each other.

Predictive-criterion validity coefficients were calculated by conducting bivariate correlations (i.e., Pearson’s r) between SAC scores and data from several variables. Specifically, all variables used in the above postdictive-criterion validity analyses (e.g., custody and various types of offenses) were assessed again during follow-up (at least six months later) and used as predictive-criterion variables. Because of some minor adjustments (i.e., collapsing of categories), these predictive variables are listed and defined separately in Appendix A. Additionally, the Fisher’s z was used, as described above, to test whether or not the observed predictive-criterion validity coefficients of both subgroups are statistically different from each other.

One final note must be mentioned regarding the criterion validity analyses mentioned above. When performing multiple comparisons using the same data, the probability of making a Type I error is inflated. Thus, finding a significant difference in one or more of these comparisons might occur simply due to chance. For this reason, many researchers incorporate a correction procedure such as the Bonferroni correction (Neter & Wasserman,

1987), which divides the number of comparisons by an overall alpha level (e.g., $p = .10$) used to control for Type I error whenever multiple comparisons are made.

Using a correction procedure is appropriate in most cases because it is typically the case that researchers are testing the null hypothesis that there are “no differences” between two or more groups. However, when researchers are attempting to demonstrate measurement equivalence among groups (i.e., “there are no differences”), a Bonferroni correction would result in researchers having an advantage in finding results that support their objectives. It was expected that there would be no differences in how the SAC performs between the two groups in this study. Therefore, no Bonferroni correction was applied to the multiple comparisons of criterion validity coefficients.

Simultaneous Factor Analyses (SFA). Although SFA is the primary method of examining the factorial invariance of the structure of the SAC across groups, independent CFA of the SAC for each group were first performed. To carry out the SFA, several preliminary model management procedures were taken. First, a structural equation model (SEM) using the AMOS software program (AMOS IV; Arbuckle, 1999) that can be used to compare the factor structure of the SAC across the AA and CA groups in this study was created. The SFA method described below is also called multi-group (or multi-sample) confirmatory factor analysis (MGCFA) (Keith et al., 1995).

Bollen and Long (1993) and Byrne (2001) suggested that each parameter in the separate models must be identified and labeled. Each group is then labeled (i.e., AA and CA) and the labeled parameter estimates of the groups are then manipulated as needed. In AMOS, the model management window was used to set up (in this case) all four models. The first model is called the “all parameters free” model, because all parameter estimates are free to

load as they would if the two models were compared individually. In this model there are no constraints between the two groups. It is the simplest and least stringent comparison and is created by simply selecting the groups to be compared.

The second model was created in the model management window by arranging the phi parameter estimate to be constrained as “equal” across the two groups. The phi parameter is an estimate of the correlation between the two factors (i.e., INT = Internalizing and EXT = Externalizing) in the model. This second model in the SFA is called the “constrained phi” model.

The third model was created by constraining the “factor pattern” of each group to be equal, using the same procedure described above. The factor patterns are the patterns of the individual item parameters or weights associated with each factor. This model is called the “constrained factor pattern” model. In the final model of these SFA analyses, both the phi parameters and the factor patterns were constrained to be equal. This fourth model is called the “constrained phi and pattern” model.

After creating all four models from the least to most restricted (i.e., constrained), the SFA analyses was conducted using the Maximum Likelihood (ML) estimation procedure with an oblique (i.e., orthogonal or correlated) rotation. Essentially, this test is concerned with the degree to which the measurement and structural model of a measure “varies” across groups. It has been suggested that factorial invariance of a measure must be established prior to making the claim of measurement equivalence across different subgroups (Cooke, et al., 2001).

For interpretation, several fit indices were examined to determine how well the data fit the models under analysis. Chi-square statistics were reported and examined, but because

the chi-square statistic can be misleading, several other fit indicators were reported and examined. Specifically, fit indices included the standardized “root mean square residual” (RMR), “root mean square estimation of approximation” (RMSEA), “goodness of fit” (GFI), and the “comparative fit index” (CFI), which is an incremental fit index.

Multiple fit indices are used to examine fit in this study because different fit indices provide different information and are sensitive to different aspects of a model (Bentler & Bonnett, 1980; Hu & Bentler, 1999). While there is likely some disagreement on rules of thumb for evaluating model fit, conventional wisdom in the social sciences suggest the practice of specifying GFI and CFI $> .90$ as acceptable model fit (Byrne, 1998). Close model fit for RMSEA $< .05$ is an applied rule of thumb for this index (Browne and Cudeck, 1993). For the RMR index, it is recommended that a cutoff value of .08 be used along with a cutoff value of .06 for RMSEA (Hu & Bentler, 1999).

CHAPTER VII

RESULTS

Descriptive Statistics

There were a total of 1,254 children and families that participated in this study, 562 (45%) of the target children were AA children and 692 (55%) were CA children. Complete demographic profiles of these children are presented in Table 1 (all tables are located in Appendix A). For the purposes of this study, the demographic profiles are presented separately for each ethnic group. As shown in the table, there was not a significant difference between the ages of the AA ($M = 14.4$ years, $SD = 2.4$) and CA ($M = 14.5$ years, $SD = 1.6$) children in the study. Similarly, although the family size of the AA ($M = 5.5$ members, $SD = 2.4$) children in this study appeared to be slightly larger than the CA ($M = 4.3$ members, $SD = 1.6$) children, this difference was not statistically significant. There were also no significant differences between AA and CA children in terms of reason for referral to juvenile court. Similar numbers of AA and CA children were referred to juvenile court for dependency and neglect (60.7% and 54.2%, respectively), status offenses (27.2% and 33.3%, respectively), and delinquent offenses (9.4% and 9.8%, respectively). Finally, AA (14.6%) and CA (13.3%) children were similar in terms of percentage of each group that was previously in state custody.

However, there were significant differences in the income and gender of these children. Although both groups had families that could be considered low income, AA families had lower monthly incomes ($M = \$1414$, $SD = 1015$) than CA families ($M = \$1773$, $SD = 1372$), $t(1252) = -5.16$, $p < .001$. Additionally, the AA sample was 70.1% ($n = 392$) male, while the CA sample was 62.0% ($n = 427$) male, $\chi^2(1) = 9.14$, $p < .01$. Furthermore,

most AA children lived in urban counties (93.4%), while most of the CA children lived in rural counties (66.8%), $\chi^2(1) = 468.80, p < .001$.

There were also significant differences in the primary care givers that provided the data for these children. While biological mothers were the top primary care givers for both groups of children, CA children (75.6%) had slightly more biological mothers as primary care givers than AA children (68.0%), $\chi^2(10) = 46.89, p < .001$. For AA children, the order of the next highest primary care giver was grandmother (12.5%), other relative (8.2%), and biological father (6.0%), but for CA children, it was biological father (8.7%), grandmother (6.9%), and other relative (2.5%). Finally, there were more CA children (21.4%) than AA children (6.4%) in this sample that had parents who lived together, $\chi^2(1) = 54.37, p < .001$.

Although not presented here, statistical analyses comparing the demographic profiles of youth that were present at baseline ($N = 1,254$), to those youth that data were provided for at follow-up ($N = 969$), found no significant differences between the two groups on any of the variables listed in Table 1.

SAC Scores for AA and CA Children. Before presenting a comparison of the reliability results on the SAC, the scores obtained by each group are noted. Scores are presented using the total SAC and each of the two dimensions (i.e., INT and EXT). Ethnic group comparisons in baseline and follow-up SAC scores are shown in Table 2.

These data indicate ethnic differences in baseline SAC outcomes were found. Specifically, mean total SAC scores were significantly higher for CA children ($M = 36.34, SD = 18.05$) than they were for AA children ($M = 33.32, SD = 19.00$), $t(1252) = -2.88, p < .01$. CA children also had significantly higher mean INT scores ($M = 14.19, SD = 9.46$), than did AA children ($M = 1.98, SD = 8.97$), $t(1252) = -4.21, p < .001$.

In terms of the overall pattern of differences, the question is whether these differences reveal actual differences or are they significantly influenced by differences in the measurement properties of the SAC for each group. Indeed, the primary purpose of this study is to provide empirical data that would allow us to address this question. The potential problem here is that measurement artifacts of the SAC caused AA respondents to significantly under-report their children's emotional and behavioral problems. The remainder of these results will help us determine how confident we can be that these scores reveal true differences between AA and CA children in this sample.

Reliability Results

Internal Consistency. Internal consistency reliability of the SAC and its EXT and INT scales were high for both groups of children. These results are shown in Table 3. Total SAC internal consistency was $\alpha = .95$ for AA children and $\alpha = .94$ for CA children. INT internal consistency was $\alpha = .90$ for AA children and $\alpha = .90$ for CA children. Finally, EXT internal consistency was $\alpha = .94$ for AA children and $\alpha = .94$ for CA children. These results indicate that the scores from the SAC and its INT and EXT scales were equally reliable and in the case of total SAC, slightly more reliable for AA children than for CA children in this sample.

Test-Retest Reliability. Test-retest reliability is essentially a measure of stability of an instrument over time (Corcoran & Fisher, 2000). In this study test-retest coefficients represent the correlation between baseline and follow-up assessments on the SAC. An important consideration must be noted regarding the test-retest reliability analyses in this study. The length of time between the baseline and follow-up assessments on the SAC was substantial (i.e., at least six months). While there are no hard, fast rules for determining adequate strength of test-retest reliability coefficients, Chronbach suggested that correlations

of .69 or better for a one-month period between administrations of an instrument is considered a “reasonable degree of stability” (cf. Corcoran & Fischer, 2000, p. 144). Given that the correlations between assessment periods in this study was at least six-months, reliability coefficients close to .69 would appear to indicate more than “reasonable” stability.

In addition to the internal consistency coefficients, test-retest coefficients of the total SAC and its INT and EXT scales for AA and CA children are also shown in Table 3. Total SAC, INT, and EXT stability coefficients for AA children were $r = .66$, $r = .60$, and $r = .68$, ($p < .001$) respectively. Total SAC, INT, and EXT stability coefficients for CA children were $r = .63$, $r = .63$, and $r = .62$, ($p < .001$) respectively. These stability coefficients appear to indicate adequate stability of the SAC and its two scales for both groups of children. Moreover, it appears that the stability of total SAC and EXT scores of AA children was stronger than they were for CA children.

Intercorrelations and Convergent-Divergent Validity

The intercorrelations (IC's) of the SAC, as well as its convergent-divergent validity with the CIS are presented in Table 4. The IC's range from $r = .57$ to $r = .92$ for AA children and from $r = .40$ to $r = .88$ for CA children, and were significant at $p < .001$. It is evident from these results that the pattern of IC's of the AA sample was similar to that of the CA sample in this study. The IC's between total SAC scores and scores from its INT and EXT dimensions appear slightly stronger for AA children than they were for CA children. Specifically, for AA children the IC between total SAC and INT $r = .84$, but for CA children this IC was $r = .79$. Similarly, for AA children the IC between total SAC and EXT $r = .92$, but for CA children this IC was $r = .88$.

Consistent with theoretical knowledge of the constructs underlying the SAC (i.e., INT and EXT behaviors) (Glisson et al., 2002), the lowest IC's were between INT and EXT scores. The IC between the INT and EXT for AA children was $r = .57$, and for CA children was $r = .40$, $p < .001$. This difference was statistically significant, z -statistic = 5.99, $p < .001$. Theoretically, it is optimal to have IC's between different, but related constructs to be $r \leq .50$, which suggests a moderate correlation. The logic behind this optimal IC of INT-EXT is that one interpretation of its observed coefficient is that it represents a proxy for divergent validity of the SAC and its constructs (Hemmigarn et al., 2003). The somewhat high INT-EXT IC for AA children raises some concern that these constructs are perceived to have more overlap for AA informants than we would like. However, the IC between EXT and INT for AA children found in this study was lower than similar IC's found in previous research on both the SAC and the CBCL (Hemmigarn et al., 2003). Specifically, Hemmigarn et al., found that the IC between INT and EXT based on parent-report SAC scores was $r = .62$ and IC between INT and EXT based on parent-report CBCL scores was $r = .57$. Therefore, while the IC between INT and EXT appeared to have been more moderate (i.e., better) for the CA children than it was for AA children, both groups had IC's that were similar to or better than INT-EXT IC's of the SAC and the CBCL found in previous research. Therefore, it appears that the IC's (or convergent-divergent validity) for both groups was consistent with previous research on the SAC.

The convergent-divergent validity of the SAC with the CIS (Bird, 1999), another measure of child mental health, is also shown in Table 4. As expected the total SAC scores converged well with the CIS for both groups, with CA children ($r = .65$, $p < .001$) having a slightly higher validity than AA children ($r = .63$, $p < .001$), but this difference was not

statistically significant. Because the items on the CIS (see Appendix C) are more indicative of externalizing behavior, the scores on the CIS were expected to have higher convergent validities with EXT scores than INT scores for both groups. EXT scores converged well with the CIS scores for both groups. CA children ($r = .63, p < .001$) had a slightly higher convergent validity than AA children ($r = .62, p < .001$), but this difference was not statistically significant. Finally, the validities between INT and the CIS diverged (i.e., were moderate), as expected, for both groups. Although CA children ($r = .43, p < .001$) had a slightly lower divergent validity with the CIS than AA children ($r = .48, p < .001$), this difference was not statistically significant. Overall, the SAC scores of the two groups appeared to have very similar convergent-divergent validity with the CIS.

Postdictive Criterion Validity

Frequency of Postdictive Criteria. The univariate statistics and frequencies of all postdictive criterion variables are presented in Table 5. The operational definitions of these variables are listed in Appendices A and D. These statistics provide an overall view of the similarities and differences in baseline levels (i.e., base-rates) of offenses, violent behavior, and mental health service history between the two groups of children in this study.

In terms of offenses it appears that AA children had higher rates of offenses than CA children in all three categories. AA children had higher status offenses ($M = 5.3, SD = 19.0$ vs. $M = 2.8, SD = 13.7, t[1252] = 2.65, p < .05$), nonviolent offenses ($M = 3.1, SD = 5.4$ vs. $M = 1.9, SD = 1.9, t[1252] = 3.74, p < .001$), and violent offenses ($M = 1.1, SD = 1.4$ vs. $M = .7, SD = 2.6, t[1252] = 3.06, p < .01$) than CA children. There were two additional postdictive criteria in which AA children had significantly higher base-rates than CA children. AA children had higher rates of custody of any kind (86.5% vs. 55.5%, $\chi^2[1] = 30.51, p < .001$)

and more than twice the rate of custody in detention, prison, or jail (83.8% vs. 41.8%, $\chi^2[1] = 229.63, p < .001$) than CA children in this study. Although the differences in offense rates would suggest that AA children would have higher rates of detention, these latter two statistics appear to indicate that AA children experienced more restrictive pathways to services than did CA children.

There were also several mental health criteria in which CA children had higher base-rates than AA children. CA children had higher base-rates of ADHD (30.5% vs. 19.4%, $\chi^2[1] = 20.10, p < .001$), community mental health services (42.5% vs. 32.0%, $\chi^2[1] = 14.24, p < .001$), services from mental health professionals (29.2% vs. 20.6%, $\chi^2[1] = 11.98, p < .01$), and psychiatric hospitalizations (17.3% vs. 10.7%, $\chi^2[1] = 11.21, p < .01$) than AA children. Medications for psychiatric, emotional and behavioral issues were also used by CA children at a higher rate (29.2% vs. 13.9%, $\chi^2[1] = 41.93, p < .001$) than AA children in this study.

Previous research has shown good predictive validity of the SAC on several of the criterion used in this study in a mixed sample of children that had higher percentages of CA children (Hemmelgarn et al., 2003). This study will build past research by demonstrating the criterion validity of the SAC for AA and CA children separately.

Postdictive Validity Coefficients for Total SAC. The total SAC postdictive validity coefficients are presented in Table 6. As expected, caregiver SAC responses (i.e., as a total score) provided criterion validities that were similar. AA caregivers provided validities that were higher than those provided by CA caregivers on five criteria and one of these was significantly different. Specifically, although CA children had higher base-rates than AA children in terms of using the services of mental health professionals (29.2% vs. 20.6%), AA children had a significantly higher validity coefficient ($r = .25$ vs. $r = .18$, z -statistic = -3.02,

$p < .001$) than CA children. Conversely, while CA children appeared to have higher validities on six postdictive criteria, none of these were significantly different. Therefore, the data presented in Table 6, suggest that when using total SAC scores, AA caregivers provided postdictive validities that were as good and in one case, better, than those provided by CA caregivers.

Postdictive Validity Coefficients for INT. The INT postdictive validity coefficients are presented in Table 7. As expected, caregiver INT responses on the SAC provided criterion validities that were very similar to those of AA caregivers. Interestingly, AA caregivers provided validities that were slightly higher than those provided by CA caregivers on four of the eight criteria and CA caregivers also provided slightly higher validities on four of the eight criteria. However, there were no statistical differences between the two groups on any of these criteria. Overall, the validity data presented in Table 7, suggest that when using INT SAC scores, AA caregivers provided postdictive validities that were as good as those provided by CA caregivers.

Postdictive Validity Coefficients for EXT. The EXT postdictive validity coefficients are presented in Table 8. As was the case with total SAC and INT scores, caregiver EXT responses provided criterion validities that were markedly similar to those of AA caregivers. For example, AA caregivers provided six validities that were slightly higher than those provided by CA caregivers and CA caregivers also provided six validities that were slightly higher than those provided by AA caregivers. There were two validities that were exactly the same for both groups. Of those that appeared different only one was statistically significant. Specifically, although CA children had higher base-rates than AA children in terms of using the services of mental health professionals (29.2% vs. 20.6%), AA children had a

significantly higher EXT validity coefficient ($r = .26$ vs. $r = .18$, z -statistic = -3.42 , $p < .001$) than CA children. Overall, the validity data presented in Table 8, appear to suggest that when using EXT SAC scores, AA caregivers provided postdictive validities that were as good and in one case, better, than those provided by CA caregivers. In sum, when taken as a whole, the postdictive criterion validities found in this study seem to support the contention that the SAC has appropriate measurement equivalence with AA caregivers.

Predictive Criterion Validity

Frequency of Predictive Criterion. The univariate statistics and frequencies of all predictive criterion variables are presented in Table 9. The operational definitions of these variables are listed in Appendix A. These statistics provide an overall view of the similarities and differences in base-rates of offenses, violent behavior, and mental health service history between the two groups of children at follow-up (at least six months). It is important to note that these base-rates are rates of occurrences since base-line, so for example, if a child reported to have been in detention at baseline, this child would have been released and then re-detained in order to meet the criteria for having had this service. Similarly, any mental health service represents “new” services that the child was receiving at follow-up. In this sense, the SAC total, INT and Ext scores are considered “true” predictors of the variables listed in Table 9.

At follow-up, it appears that AA children had higher rates of SED (9.3% vs. 5.8%, $\chi^2[1] = 4.41$, $p < .05$) and a higher rate of custody in detention, prison, or jail (31.2% vs. 18.7%, $\chi^2[1] = 20.10$, $p < .001$) than CA children in this study. In contrast to baseline rates of offenses, offense rates of AA and CA children were not significantly different. Yet, AA children continued to experience higher rates of various forms of detention, which suggest

the AA children's pathway to service was once again more restrictive than the service pathways of CA children.

Additionally, at follow-up, there were several mental health criteria in which CA children had higher base-rates than AA children. CA children had higher base-rates of ADHD (29.4% vs. 17.5%, $\chi^2[1] = 20.10, p < .001$), any mental health service (41.0% vs. 34.0%, $\chi^2[1] = 5.05, p < .05$), community mental health service (28.9% vs. 15.1%, $\chi^2[1] = 25.97, p < .001$), unruly adjudication (21.1% vs. 14.3%, $\chi^2[1] = 7.32, p < .01$), and alcohol use (18.6% vs. 10.3%, $\chi^2[1] = 13.16, p < .001$) than AA children. As with the postdictive criteria, baseline levels (i.e., base-rates) of the predictive variables are important to consider when interpreting the criterion validity analyses. Specifically, it was expected to find a validity coefficient for AA children that was as high or higher than the CA group corresponding to those variables where AA children had higher base-rates

Predictive Validity Coefficients for Total SAC. The total SAC predictive validity coefficients for both AA and CA children, along with tests of significant differences are presented in Table 10. Caregiver SAC responses (i.e., as a total score) provided predictive criterion validities that were similar. AA caregivers provided validities that appeared slightly higher than those provided by CA caregivers on three criteria. However, none were significantly different. CA children appeared to have somewhat higher validities on seven predictive criteria, and three were significantly different. Therefore, of the 11 validities presented in Table 6, only three were significantly different. Moreover, it was expected that CA children would have a higher validity for community mental health service (i.e., $r = .24$ vs. $r = .16$, z -statistic = $-3.06, p < .01$) than AA children because they had higher base rates on this particular criterion (i.e., 28.9% vs. 15.1%, respectively). The finding that CA children

had higher predictive validities for any offense ($r = .20$ vs. $r = .09$, z -statistic = -4.74 , $p < .001$) and violent behavior ($r = .23$ vs. $r = .17$, z -statistic = -2.30 , $p < .05$) than AA children was not expected and possible reasons for this difference will be addressed in the discussion section. Nonetheless, overall, the predictive validities of the total SAC scores of AA and CA children appeared to be similar, providing further evidence of measurement equivalence of the SAC across these two groups.

Predictive Validity Coefficients for INT. The INT predictive validity coefficients for both AA and CA children, along with tests of significant differences are presented in Table 11. As expected, caregiver INT SAC responses of AA and CA children provided predictive criterion validities that were similar. AA caregivers provided validities that appeared slightly higher than those provided by CA caregivers on four criteria, however, none of these were significantly different.

CA children appeared to have somewhat higher validities on four predictive criteria, and one was significantly different. Specifically, CA children had a higher validity for community mental health service than AA children (i.e., $r = .24$ vs. $r = .16$, z -statistic = -3.06 , $p < .01$). In short, of the six validities presented in Table 11, only one was significantly different. Thus, the predictive validities of the INT SAC scores of AA and CA children appeared to provide further evidence of similar psychometric properties of the SAC across these two groups.

Predictive Validity Coefficients for EXT. The EXT predictive validity coefficients for both AA and CA children, along with tests of significant differences are presented in Table 12. AA caregivers provided validities that appeared slightly higher than those provided by CA caregivers on two criteria, neither of which was significantly different. In contrast, CA

children appeared to have somewhat higher validities on nine predictive criteria, but only four were significantly different. Of the four that were significantly different in favor of CA children, three were expected because of the significant difference in base rates. Specifically, CA children had a higher base rate of ADHD (29.4%) than AA children (17.5%) and it was expected that CA children would have a higher validity on this variable (i.e., $r = .23$ vs. $r = .16$, z -statistic = -2.70, $p < .01$) than AA children. Additionally, CA children also had a higher base rate of any mental health service (41.0%) than AA children (34.0%) and there was a similar expectation that CA children would have a higher validity on this variable (i.e., $r = .24$ vs. $r = .18$, z -statistic = -2.70, $p < .01$) than AA children. Finally, CA children had a higher base rate of community mental health service (28.9%) than AA children (15.1%), it was expected that CA children would have a higher validity on this variable (i.e., $r = .20$ vs. $r = .14$, z -statistic = -2.44, $p < .05$) than AA children.

Overall, where expected, AA children's EXT predictive validities were similar to those of CA children. The finding that CA children had a higher EXT predictive validity for any offense ($r = .23$ vs. $r = .13$, z -statistic = -3.97, $p < .001$) than AA children was not expected. Nonetheless, the predictive validities of the EXT SAC scores of AA and CA children were similar and appeared to provide further evidence of measurement equivalence of the SAC across these two groups.

Separate Factor Analysis for AA and CA Children

Before conducting simultaneous factor analyses (SFA), an individual factor analysis was conducted separately using the Maximum Likelihood estimation procedure (with oblique rotation) to assess the overall model fit of the model presented by Glisson et al, (2002) for each group of children. Fit indices were examined separately for AA and CA children. The

results of the fit indices for AA children were mixed, with two fit indices indicating close fit (i.e., RMR = .034, RMSEA = .042) and two others (i.e., GFI = .871, CFI = .852) indicating that the model approached, but did not reach the .90 good fit criterion. Similarly, the results of the fit indices for CA children were also mixed, with two fit indices indicating close fit (i.e., RMR = .033, RMSEA = .041) and two others (i.e., GFI = .867, CFI = .852) indicating that the model approached, but did not reach good GFI or CFI fit criterion of $\geq .90$.

Therefore, it appears that for this sample, some of the fit indices previously reported, which showed a good overall model fit of the factor structure of the SAC across parents and teachers for boys and girls and across all age ranges by Glisson et al, (2002) were replicated and others were not. With these initial factor analyses as a backdrop, SFA were pursued.

Simultaneous Factor Analysis (SFA) Across AA and CA Children

The initial SFA was conducted on the SAC by constructing the separate models of the two-factor scale with one-higher-order factor as presented by Glisson et al, (2002) for AA and CA children in this study. Specifically, AMOS was used to construct a structural equation model (SEM) of the SAC (i.e., 24 items loading on the INT factor and 24 items loading on the EXT factor) for both groups. After arranging the model management procedure as described in the methods section above, the ML estimation procedure was again used to estimate model fit between groups for all four models (i.e., all parameters free, constrained phi, constrained factor pattern, and constrained phi and factor pattern) simultaneously. The models were arranged from least to most restrictive. The results of this initial SFA analysis are shown in Table 13.

When examining SFA, interpretation of the chi-square (X^2), degrees of freedom (df), and appropriate p-values, as well as fit indices of each model are required. Based on an

examination of the X^2 , df , and appropriate p-values of each model in the initial SFA, it appears that all four models were statistically different, thus, indicating that the factor structure was non-equivalent across the AA and CA samples in this study. When examining the fit indices, two fit indices (i.e., RMR ranged from .039 to .044 and RMSEA ranged from .042 to .043) suggest that all four models had close fit. In contrast, when examining the other two fit indices, it appears that the models lacked good fit (i.e., GFI ranged from .779 to .782 and CFI ranged from .814 to .818). After examining the above fit indices, it appears that the all-parameters-free model had best fit and an adjustment to the model might show improvement in the measurement equivalence of the model between the two groups in this study. Before adjusting the model, the appropriate first step is to examine the “model comparison” results that are also provided by AMOS when conducting SFA analyses.

AMOS provides model comparison statistics in which the X^2 differences, change in df , and change in X^2 , along with appropriate p-values associated with each comparison. Because the all-parameters-free model has the least (i.e., none) constraints between each group’s original model, AMOS uses it as the comparison model. As shown in Table 14, the results again suggest that the models were not improved by adding constraints as indicated by the X^2 - df -ratio increasing as constraints were added. Specifically, for Comparison B, the ratio was 2.51 and for Comparison C, the ratio was 2.71. The model comparison results suggest that the X^2 - df -ratio of each model in the SFA might respond (i.e., improve) to minor adjustments in the overall model of the SAC. One logical step might be to discard items that appear to be problematic (i.e., show a large difference between the two groups) (Trendafilov, 2003).

Given that the all-parameters-free model appears to have had the best fit, the standardized regression weights of each item in this model were examined in order to identify problem items that, if discarded, might improve the overall fit between the two groups. Table 15 shows the standardized parameter estimates of the INT items for both groups and Table 16 shows the standardized parameter estimates of the EXT items for both groups. An initial examination of the INT and EXT parameter estimates reveals that the pattern of differences was fairly balanced, with equal numbers of items loading higher for AA children and higher for CA children. Moreover, all but a few items had high (i.e., strong) estimates for both groups of children, with most items having estimates higher than .40 on each factor for both groups. This initial examination of the INT and EXT parameter estimates suggests that, overall, there was no systematic error or bias in the SAC that favored one group over another.

However, further examination of the results in Tables 15 and 16, show that three INT items (i.e., items 12, 17, and 31) and two EXT items (i.e., items 32 and 38) had differences in loadings between the two groups that were larger than .10. Items that had differences in parameter estimates between the two groups that were larger than .10 were discarded and a follow-up SFA analysis was conducted. Note that in the follow-up analysis of the revised scale, four of the five items (i.e., INT items 12 and 31, and EXT items 32 and 38) loaded higher for AA children and one item (i.e., INT 17) loaded higher for CA children.

Although there were minor improvements in the measurement equivalence of the SAC after discarding the problem items, the results of the follow-up SFA were similar to the initial SFA results. These results are presented in Table 17. As with the initial SFA, the p-values associated with the X^2 of each model indicate statistical significance and suggest that the model remained non-equivalent between the AA and CA groups of children in this study.

Nonetheless, the fit indices appear to indicate that the model fit between the two groups of the least constrained model (i.e., all parameters free) did improve somewhat. The RMR (i.e., ranged from .039 to .044) and RMSEA (i.e., .044 for all four models) showed as close a fit as they did in the initial SFA results. The CFI increased from .782 to .793, and the GFI increased from .818 to .832. While both fit indices again failed to reach the conventional level of .90, the improvement in the model suggested that further examination of the model comparison results was appropriate.

An examination of the overall pattern in model comparison results of the follow-up SFA shown in Table 18, also appear to indicate that the model improved somewhat. For example, X^2 -*df*-ratio improved (i.e., decreased) for each model comparison after deleting the problem items from the model. Taken together, the fit indices and model comparison results of both the initial and follow-up SFA indicate that while the all-parameters-free model appeared to have had the best fit between groups, none of the models showed good cross-ethnic equivalence of the SAC. However, both the independent and simultaneous factor analyses seem to suggest that while the structural and measurement model of the SAC might work fairly well independently for the AA and CA youth in this study, the SAC lacks measurement equivalence across both groups.

It is important to note that factor equivalence is difficult to achieve among groups, because it indicates that all items work similarly for both groups. However, the lack of factor equivalence does not mean that the measure cannot be used successfully with both groups. One interpretation of the results of the factor structure analyses is that SAC has practical utility and substantive significance (Rubin & Babbie, 1997) for both groups but does not work equivalently for both groups. It might be that the SAC works well for AA informants

for one set of reasons and works well for CA informants for another set of reasons. This issue will be addressed more completely in the discussion section.

CHAPTER VIII

DISCUSSION

Summary and Implications of Major Findings

The primary purpose of this study was to undertake a comprehensive analytic approach to test the measurement equivalence of the SAC, which was developed to assess the mental health of children referred to child welfare and juvenile justice systems. This study provides several kinds of preliminary evidence to support the conclusion that the SAC lacks systematic bias and has measurement equivalence across both CA and AA children referred to juvenile court. These findings, although preliminary, are somewhat convincing because they were generated using a large sample of children with survey interviews that were completed over an extended period of time. There were several findings that were particularly noteworthy and warrant further discussion.

The reliability evidence supported the contention that the SAC is equally reliable for both groups. Both internal consistency reliability and test-retest reliability (i.e. stability) was markedly similar, showing good reliability for both AA and CA youth. Studies reporting the reliability of scales rarely present evidence of cross-ethnic reliability. These results demonstrate that the SAC not only provides estimates of INT and EXT behaviors that are internally stable for AA youth, but also provides estimates that are stable over time for both groups of children. It is important that practitioners and researchers facing critical decisions and issues regarding assessment outcomes can be fairly confident that these scores are reliable.

The IC's of the SAC showed that the relationships between the two factors were consistent with theoretical knowledge for AA and CA children. Furthermore, both the AA

and CA children in this study had IC that were more consistent with theoretical knowledge of the underlying constructs of the SAC than were found in a previous study of the SAC (Hemmelgarn, et al. 2003). This is further evidence of the adequacy of the SAC's measurement properties for AA children. Moreover, the SAC had convergent-divergent validities with the CIS that were equally good for both AA and CA children. In addition to providing further support of the validity of the SAC evidence in previous research (Glisson, et al. 2002; Hemmelgarn, et al. 2003), the results of this study provide new data supporting the cross-ethnic validity of SAC.

The postdictive validities of the total INT and EXT SAC scores are markedly similar and equally good for both AA and CA children in this study. In a similar fashion, the predictive validity of the total, INT and EXT SAC scores of the AA and CA children were comparable. The findings from the postdictive and predictive validity analyses are important in several ways. The practice implications of these findings are that the SAC appears useful for both AA and CA children in predicting future behavior problems. This is important because one of the most fundamental purposes of child mental health assessment is to identify children who might be more at risk for more severe behavior problems in the future. Furthermore, social workers and other mental health professionals need assessment tools that are accurate (i.e., reliable and valid) in identifying and monitoring behavior problems.

The SFA analyses are important to our overall assessment of the results of this study because failure to detect significant differences in validity coefficients alone is not evidence that a measure is valid for ethnic-minority groups (Malgady, 1996). The SFA results can be viewed from two perspectives. Statistically, the results of the measurement and structural equivalence of the SAC across the AA and CA samples in this study was mixed. There were

some fit indices indicating adequate and close fit, and others that revealed that the model did not meet the criteria for good fit. Moreover, the model comparison statistics of both the initial and follow-up SFA analyses showed that the factor structure of the SAC was non-equivalent across groups. However, these results also revealed that the SAC had substantive and practical significance across groups. The standardized weights of the items suggested that the overall scale did not favor one group over another. These results indicate that for practical purposes, some SAC items appear to work better for AA children and others appear to work better for CA children. It might be the case that all items work well for both groups but in different ways.

The results of the SFA highlight the difficulty that exists in developing a measure that can pass one of the most rigorous tests of measurement equivalence. Previous research has shown that obtaining adequate fit indices when conducting SFA of latent constructs across different ethnic and cultural groups is rare and that it is often the case that researchers make minor adjustments to items (Peterson, Smith, Akande, Ayestaran, Bochner, Callan, & Cho, (1995). Making adjustments to the SAC items might not be prudent because of the small differences that were found in this study. However, the implications of the small differences in item weights are important to this discussion.

Inspection of the differences in standardized item weights revealed two important findings. First, among the five items that had differences greater than .1, AA children had higher (i.e., stronger) parameters estimates than CA children. The only item that favored CA children was the INT item 17-“keeps to himself/herself a lot.” This item apparently posed some conceptual problems for AA caregivers. It might be that the disproportionate level of stressors involved in AA communities (McLoyd, 1998) contributes to AA caregivers feeling

relieved to have a child that keeps to himself/herself. AA parents also might be less concerned about their children spending time alone because they may see this as assurance that their children are staying out of trouble. Another possibility is that the difference in this INT item could be due to the gender differences between the two groups. There were more proportionately more girls in the CA group (i.e., 8% more) than there were in the AA group and it is well-documented that adolescent girls exhibit significantly higher rates of internalizing problems than adolescent boys (Keenan & Shaw, 1997). However, because there was such a small difference between the two groups in terms of gender and only one INT item had a difference greater than .10 that favored the CA group, it is unlikely that gender issues are what caused the SFA to be non-equivalent across the two ethnic groups. Nonetheless, this finding poses some very interesting questions and it would be important to control for gender when analyzing the cross-ethnic equivalence of measures in future studies.

A second important finding from inspection of differences on item weights is that several items loaded stronger for AA caregivers. Specifically, item 38-“swears or curses”, item 31-“is unsure of self or easily embarrassed” and item 32-“clowns around or shows off”, all had differences in loadings between groups that were greater than .1, in favor of AA children. It might be that average levels of being uncomfortable with attention, easily embarrassed, clowning around, and swearing are perceived by AA caregivers to indicate greater than average behavior problems. There is some research that suggests AA parents tend to be strict disciplinarians (Lambert et al., 2001; 2002; MacAdoo, 2002; Peters, 2002) and might place greater emphasis on limiting clowning around and swearing types of behaviors than CA parents. The differences in the above EXT items appear consistent with recent research conducted by Lambert et al., (2002). Lambert et al., found that the most

recorded behavior problem by AA parents in a large sample of clinically referred children was “uncooperative behavior.” Moreover, Lambert et al., (2001) suggested that in order to better understand and treat psychopathology in AA children, it is important to become familiar with AA values and family practices, particularly in terms of parenting. AA parents place greater emphasis on respect of elders in the home and in the community. Therefore, swearing and clowning around would be strongly discouraged by AA parents. Without further information, these explanations are purely speculative and future research should include additional qualitative data to compliment the type of data provided here.

There is another possible explanation for the dissimilar psychometric functioning of the small set of items found in this study. It might be that these items performed well in identifying actual differences between AA and CA children in how INT and EXT behavior problems are manifested. For example, it is possible that item 17 – “keeps to him/herself a lot” – is more characteristic of INT problems for CA children than it is for AA children. In a similar fashion, it might be that items 31, 32, and 38 are more characteristic of EXT problems for AA children than they are for CA children. This would further suggest that the SAC is reliable and valid for both AA and CA children and in some cases it is reliable and valid with both groups for different reasons.

Summary and Implications of Unexpected Findings

There were additional findings in this study that warrant further discussion. There were several significant differences in the validity coefficients that were contrary to what was expected, given the direction in differences of base-rates of the criteria. For example, postdictively, AA children total SAC and EXT SAC scores predicted services from mental health professionals better than did CA scores. Yet, CA children had higher base-rates on this

criterion than AA children. It might be that AA children have less access to mental health professionals so that only those with the most severe behavioral and mental health problems are likely to receive services.

There were several differences in the predictive validity of caregivers' total and EXT SAC scores that were not in the direction expected. For example, although AA and CA children had similar base-rates for any offense, CA children had significantly higher total and EXT validity for this criterion. It is possible that those AA children that either re-offended or committed their first offenses at follow-up, were more likely to have had lower SAC scores than their CA counterparts at baseline. We do know from these results that the SAC worked equally well for both AA and CA groups in predicting which children would return to a detention-type facility. This appears to support the theory that those AA children who committed offenses at follow-up, were more likely to be first offenders and were therefore, less likely to experience detention.

The findings related to detention might have the most troubling implications of this study. Consistently, AA children were more likely to experience detention and other more restrictive services. This has been documented in previous studies (e.g., Nugent & Glisson, 1999). In this study, AA children were detained at more than twice the rate than CA children at baseline. This disparity might be explained in part by the fact that AA children appeared to have committed the most offenses in all three categories recorded at baseline. However, we have no further information on which children were referred to juvenile court or after court referral, which children were subsequently charged with an offense. It is possible that AA and CA children with similar behavior profiles were not referred to juvenile court at similar

rates. It is also possible that once referred to juvenile court, AA and CA children with similar mental health profiles were not subsequently charged with an offense at similar rates.

We know from these results that CA children consistently had significantly higher SAC scores than their AA counterparts. For example, at baseline CA children had significantly higher total and INT SAC scores than did AA children. We also know that at follow-up CA children had significantly higher total, INT and EXT SAC scores than AA children. These findings suggest that while CA children consistently had significantly more mental and behavioral health problems than AA children, they experienced less restrictive pathways to service than did AA children. However, without further analyses, little can be said about the underlying mechanisms that might explain this disparity.

Limitations of the Study

The results presented above must be interpreted within the context and after careful consideration of the limitations of this study. First, not all children and families that were invited to participate agreed to do so (i.e., 65%). No data, demographic or otherwise, were collected from the estimated 35% non-responders. The implication is that although data were collected in 20 different counties across one state, this sample might not be representative of all children referred to juvenile court in this state. However, we do have statistical data that suggests the demographic profiles of these children, including their reason for referral and previous custody, are similar to children previously referred to juvenile court across the state in which this study was conducted (Tennessee Council of Juvenile and Family Court Judges, 2002).

A second limitation of this study is also related to the sample. The sample was not truly random. The children in this study were all referred to juvenile court and were selected

only after meeting certain criteria (i.e., delinquency, status offense, or dependency and neglect issues). This population might be considered a “high-risk” population, because they are likely to have a greater probability of developing emotional and behavioral problems. However, the SAC was specifically developed to aid service providers in child welfare and juvenile justice systems (Glisson et al, 2002; Hemmelgarn et al, 2003). Although these results might not generalize to all children, the sample used in this study was appropriate to test the measurement equivalence of the SAC and the results might be applicable to children in other child welfare and juvenile justice systems. Nonetheless, these results must be replicated using similar samples to give greater weight to their meaning and applicability.

A third possible limitation also concerns the generalizability of these results to other groups. A major criticism of research on ethnic-minority children is that while all ethnic-minority children are under-represented in these studies, the majority of research published in the literature tends to focus on AA children (McLoyd, 1998). The present study continued this pattern, as AA was the only ethnic-minority group represented in this research. This raises some questions regarding the generalizability of the results discussed above to other ethnic-minority children. These results need to be replicated with other ethnic-minority groups, such as Hispanic children. This would provide further evidence of the measurement equivalence and practical application of the SAC across different ethnic and cultural groups in the United States.

A final limitation relates to the measurement procedures used in this study. The data used in this study came exclusively from a single source (i.e., primary caregivers of the children). Therefore, the results discussed above might have been affected by a particular form of mono-method bias (Cook & Campbell, 1979). The typical case of mono-method bias

involves problems associated with interpreting results that were generated from a single source. In this study the results might have been affected by “single source (i.e., primary caregivers) bias.” This is a related principal that suggests using a single source to provide all the information in a study that included multiple measures raises concern that each measure might have been vulnerable to the same source of error, resulting in a form of systematic error. In this regard, “consistency among the measures would not really tell us whether that source of systematic error was being avoided,” (Rubin & Babbie, 1997, p. 173). However, the primary purpose of this study was to compare outcomes on measures and relationships among outcomes on these measures across groups. It is likely that possible errors from single source bias were randomly distributed between the two groups. As a result, the substantive findings of this study are interpretable. Nonetheless, future studies should include multiple sources of data to strengthen the validity of the results found in this study.

Conclusion and Directions for Future Research

Despite its limitations, this study provides important knowledge that can help to address a critical gap in research on cross-ethnic measurement equivalence. Past research has shown identified concerns for the cross-ethnic measurement equivalence of tools regularly used to assess child mental health (Lambert et al., 2002; Okazaki & Sue, 1995; Weisz & McCarty, 1999). Tools used to assess and monitor treatment protocols for child mental health issues such as ADHD, must first rule out the presence of systematic bias in favor of or against any one particular ethnic group (Jarvinen & Sprague, 1995) prior to gaining wide acceptance in the field. Unfortunately, few behavior rating scales used in practice and research (e.g., the CBCL) in social work, psychology and related mental health fields have

been systematically analyzed with methods similar to those incorporated in this study (Tyson, in press).

The comprehensive approach to investigating measurement equivalence that was taken in this study found that there were substantially more similarities than there were differences between AA and CA children in terms of the psychometric properties of the SAC. This research gives practitioners and researchers greater confidence that scores obtained from the SAC provide reliable and valid estimates of the level of emotional and behavioral problems of children, regardless of ethnic or cultural background. Although the present study has moved the literature on cross-ethnic equivalence forward, there is much work to be done in this area.

Now that we have a substantial body of evidence supporting the validity of the SAC as a measure of child mental health with children in child welfare and juvenile justice systems, future studies should explore the appropriateness of the SAC in other settings, such as the educational and broader mental health communities. Additionally, more cross-ethnic research is also needed that includes different informants. These findings highlight the importance of including assessments from multiple informants, providing data that describes children from various perspectives. Specifically, we know that teachers can provide valid and useful information with the SAC. Future research should include a comprehensive analysis of the measurement equivalence of the SAC using teachers as informants. Research of this kind would give greater confidence in the cross-situational validity of the SAC across different ethnic groups in the United States. The SAC also should be examined as a child-report assessment tool. One of the limitations of this study is that behavioral and emotional health problems were estimated from the perspective of the parents. It would be important to

investigate how well the SAC is able to provide valid and useful data using children from different ethnic groups as informants. Finally, future research is needed to identify and disentangle the underlying mechanisms that might explain the disparities between AA and CA children in terms of pathways to services that were found in this study.

To summarize, these results must be viewed as a comprehensive set of findings and, together, provide preliminary evidence of the measurement equivalence of the SAC for AA children, as well as add to its validity as a tool to assess CA children. With increased concern about the mental health of children in child welfare and juvenile justice systems, practitioners and service providers require this type of evidence-based approach to selecting appropriate assessment tools for different ethnic groups. This is particularly important because these tools might be used for important placement and treatment decisions.

In terms of the few items that were found to favor one group over another, item analysis would be an important next step in future attempts to explain these differences. Specifically, IRT methods might be particularly effective in identifying items with differential functioning across groups and the extent of these differences.

While several possible weaknesses of this study have been discussed, certain characteristics of the sample contributed to the strength of these data in an important way. The two groups of children were very similar in terms of their demographic profiles. The most notable difference was that nearly all of the AA children lived in urban areas (i.e., 90%) and the majority of CA children (i.e., 67%) lived in rural counties. Although statistically significant, the other differences between these youth were practically negligible, with the possible exception of the number of parents living together. Both AA and CA families had low incomes, with AA families having incomes that were slightly lower. Both groups had

large proportions of males, with CA youth having slightly more males. And both groups of children had similar caregivers, although it appears AA children were twice as likely to have a grandmother as their primary caregiver. The remaining variables in the demographic profiles of these youth were nearly identical. Specifically, these children were the same age, came from families that were nearly identical in size, were referred to juvenile court for the same reasons, and had previous custody histories that were almost identical.

One implication is that the similarities in demographic background found between the two groups might explain the strikingly similar reliability and validity evidence provided by their scores on the SAC. These data appear to support the contention that ethnicity and race might not be as relevant as SES and other important demographic characteristics. What is particularly noticeable about this sample is that although AA families lived in urban areas and CA families primarily came from rural areas, they seemed to have similar family structures and similar problems. One conclusion might be that regardless of race, youth entering the child welfare and juvenile justice systems are similar in ways that are important to service system administrators and practitioners. However, this raises further speculation that the differences found in pathways to services between these two groups of children do not appear to have justification and must be addressed.

In conclusion, this research has practical significance beyond this study. Policy makers and service system directors in child mental health services must identify effective, evidence-based assessment tools in this era of outcome-based services. Large service systems are becoming increasingly reliant on management information systems to ensure compliance with program requirements and monitor client outcomes. The importance of reliable and valid standardized instruments that are applicable across diverse ethnic groups cannot be

overstated. Therefore, policy makers need to become conversant with psychometric research and determine whether measures are valid for members of different ethnic groups that we serve. Taking this approach will not only lead to better assessment tools being used in large service systems responsible for large numbers of AA and other ethnic minority children, this will ultimately lead to better treatment and placement outcomes for these children.

REFERENCES

REFERENCES

- Achenbach, T. M. (1966). The classification of children's psychiatric symptoms: A factor analytic study. *Psychological Monographs: Applied and General*, 80(7), 1-37.
- Achenbach, T. M. (1980). DSM-III in light of empirical research on the classification of child psychopathology. *Journal of the American Academy of Child and Adolescent Psychiatry*, 19, 395-412.
- Achenbach, T. M. (1988). Integrating assessment and taxonomy. In M. Rutter, A. H. Tuma, & I. S. Lann (Eds.), *Assessment and diagnosis in child psychopathology* (pp. 300-343). New York: Guilford Press.
- Achenbach, T. M. (1991a). *Manual for the Child Behavior Checklist/4-18 and 1991 profile*. Burlington, VT: University of Vermont Department of Psychiatry.
- Achenbach, T. M. (1991b). *Manual for the Teacher's Report Form and 1991 profile*. Burlington, VT: University of Vermont Department of Psychiatry.
- Achenbach, T. M., & Edlebrock, (1978). The classification of child psychology: A review and analysis of empirical efforts. *Psychological Bulletin*, 85, 1275-1301.
- Achenbach, T. M., Howell, C. T. Quay, H. C., & Conners, C. K. (1992). National survey of competencies and problems among 4- to 16-year-olds: Parents reports for normative and clinical samples. *Monographs of Social Research in Child Development*, 56, 225.
- Achenbach, T. M., and McConaughy, S. H. (1997). *Empirically-based assessment of child and adolescent psychopathology: Practical applications*. (2nd ed.). Newbury, CA: Sage.
- Allen, R. L., & Oshagan, H. (1995). The UCLA loneliness scale: Invariance of social structural characteristics. *Personality and Individual Differences*, 19(2), 185-195.

American Psychiatric Association. (1980). *Diagnostic and statistical manual of mental disorders* (3rd ed.). Washington, DC: Author.

American Psychiatric Association. (1987). *Diagnostic and statistical manual of mental disorders-revised* (3rd ed., rev.). Washington, DC: Author.

American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th ed.). Washington, DC: Author.

American Psychiatric Association. (2001). *Diagnostic and statistical manual of mental disorders-text revision* (4th ed.-TR). Washington, DC: Author.

American Psychological Association, American Educational Resource Association, and National Council on Measurement in Education. (1985). *Standards for educational and psychological testing*. Washington, DC: American Psychological Association.

American Psychological Association, American Educational Resource Association, and National Council on Measurement in Education. (1992). *Standards for educational and psychological testing*. Washington, DC: American Psychological Association.

American Psychological Association, American Educational Resource Association, and National Council on Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington, DC: American Educational Resource Association.

Anderson, J. C., & Gerbing, D. W. (1988). Structural equation modeling in practice: A review and recommended two-step approach. *Psychological Bulletin*, 103, 411-423.

Angold, A., Costello, E J., Erkanli, A. (1999). Comorbidity, *Journal of Child Psychology and Psychiatry and Allied Disciplines*, 40(1), 57, 31.

Arbuckle, J. L. (1999). *SPSS/AMOS IV*. Chigago, IL: SmallWaters Corporation.

Arbuckle, J. L., & Wothke, W. (1999). *AMOS 4.0 User's Guide*. Chicago, IL: SmallWaters Corporation.

Arnold, B. R., & Matus, Y. E. (2000). Test translation and cultural equivalence methodologies for use with diverse populations. In I. Cuellar & F. A. Paniagua (Eds.). *Handbook of multicultural mental health: Assessment and treatment of diverse populations* (pp. 121-136). New York: Academic Press.

Barbarin, O. A., & Soler, R. E. (1993). Behavioral, emotional, and academic adjustment in a national probability sample of African American children: Affects of age, gender, and family structure. *Special Issues: Emotional Development of African American Children. Journal of Black Psychology, 19*, 423-446.

Barber, C. C., Neese, D. T., Coyne, L., Fultz, J., & Fonagy, P. (2002). The target symptom rating: A brief clinical measure of acute psychiatric symptoms of children and adolescents. *Journal of Clinical Child and Adolescent Psychology, 31*(20), 181-192.

Barber, B. K., Stolz, H. E., Olsen, J. A., & Maughan, S. L. (manuscript submitted to SRCD Monographs). Parental support, psychological control, and behavioral control: Longitudinal and cross-national validations.

Bennett, I. (1960). *Delinquent and neurotic children*. New York: Basic Books.

Bentler, P. M., & Bonnett, D. G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin, 88*, 588-606.

Berk, R. A. (1982). *Handbook of methods for detecting test bias*. Baltimore, MD: John Hopkins University Press.

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

Bollen, K. A., & Long, J. S. (1993). *Testing structural equation models*. Newbury Park: Sage.

Brady, E. U., & Kendall, P. C. (1992). Comorbidity of anxiety and depression in children and adolescents. *Psychological Bulletin*, 111, 244-255.

Breen, M. J., & Fielder, C. R. (Eds.) (1996). *Behavioral approach to assessment of youth with emotional/behavioral disorders: A handbook for school-based practitioners*. Austin, TX: Pro-ed, Inc.

Brown, J. S. & Achenbach, T. M. (1996). *Bibliography of published studies using the Child Behavior Checklist and related materials: 1996 edition*. Burlington, VT: University of Vermont Department of Psychiatry.

Browne, M. W. & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 136-162). Newbury Park, CA: Sage.

Bruce-Webb, M. & Jones-Harden, B. (2003). Beyond child protection: Promoting mental health for children and families in the child welfare system. *Journal of Emotional and Behavioral Disorders*, 11(1), 49-58.

Burchard, E. G., Ziv, E., Coyle, N., Gomez, S. L., Tang, H., Karter, A. J., Moutain, J. L., Perez-Stable, E. J., Sheppard, D., & Risch, N. (2003). The importance of race and ethnic background in biomedical research and clinical practice. *The New England Journal of Medicine*, 348(12), 1170-1175.

Burns, B. J., Landsverk, K., Kelleher, K., Faw, L., Hazen, A. & Keeler, G. (2001). Mental health, education, child welfare, and juvenile justice service use. In R. Loeber & D. P.

Farrington (Eds.). *Child delinquents: Development, intervention, and service needs*. (pp. 273-304). Thousand Oaks: Sage.

Byrne, B. M. (1998). *Structural equation modeling with LISREL, PRELIS, and SIMPLIS: Basic concepts, applications, and programming*. Mahwah, NJ: LEA.

Byrne, B. M. (2001). *Structural equation modeling with AMOS: Basic concepts, applications, and programming*. Mahwah, NJ: LEA.

Byrne, B. M., & Campbell, T. L. (1999). Cross-cultural comparisons and the presumption of equivalent measurement and theoretical structure: A look beneath the surface. *Journal of Cross-Cultural Psychology*, 30(5), 555-574.

Cattell, R. B. (1949). A note on factor invariance and the identification of factors. *British Journal of Psychology*, 11, 134-138.

Center for the Study of Social Policy (1993). *Kids count data book*. Washington, DC: Author.

Cicchetti, D., & Toth, S. L. (1991). A developmental perspective on internalizing and externalizing disorders. In D. Cicchetti & S. L. Toth (Eds.), *Internalizing and externalizing expressions of dysfunction: Rochester symposium and development psychopathology* (pp. 1-20). Hillsdale, NJ: Erlbaum.

Cohen, J. (1992). A power primer. *Psychological Bulletin*, 112(1), 155-159.

Cohen, J., & Cohen, P. (2003). *Applied multiple regression analysis for the behavioral sciences* (3rd. ed.). Mahwah, NJ: Erlbaum.

Cook, T., & Campbell, D. (1979). *Quasi-experimentation*. Boston: Houghton-Mifflin.

Cooke, D. J., Kosson, D. S., & Michie, C. (2001). Psychopathy and ethnicity: Structural, item, and test generalizability of the Psychopathy Checklist-Revised (PCL-R) in Caucasian and African American participants. *Psychological Assessment*, 13(4), 531-542.

Cooper, R. S., Kaufman, J. S., & Ward, R. (2003). Race and genomics. *The New England Journal of Medicine*, 348(12), 1166-1170.

Cone, J. D. (1998). Psychometric considerations: Concepts, contents, and methods. In A. S. Bellack & M. Hersen (Eds.). *Behavioral assessment: A practical handbook* (4th ed., pp. 22-46). Boston, MA: Allyn and Bacon.

Conners, C. K. (1973). Rating scales for use in drug studies with children. *Psychopharmacology Bulletin*, 9, 24-29.

Conners, C. K. (1999). Conners Rating Scales-Revised. In M. E. Maruish (Eds.), *The use of psychological testing for treatment planning and outcome assessment* (2nd ed., pp. 467-496). Mahwah, NY: Earlbaum.

Corcoran, K., & Fischer, J. (2000). *Measures for clinical practice: Vol. 1. Couples, families, and children*. (3rd ed.). New York, NY: Free Press.

Crijnen, A. A. M., Achenbach, T. M., & Verhulst, F. C. (1997). Comparisons of problem reports by parents of children in 12 cultures: Total problems, externalizing and internalizing. *Journal of the American Academy of Child and Adolescent Psychiatry*, 36, 1269-1277.

Dana, R. H. (2000). Culture and methodology in personality assessment. In I. Cuellar & F. A. Paniagua (Eds.). *Handbook of multicultural mental health: Assessment and treatment of diverse populations* (pp. 97-120). New York: Academic Press.

Deater-Deckard, K., & Dodge, K. A. (1997). Externalizing behavior problems and discipline revisited: Nonlinear effects and variation by culture, context, and gender.

Psychological Inquiry, 8(3), 161-175.

Depaul, G. J. (1991). Parent and teacher ratings of ADHD symptoms: Psychometric properties in a community-based sample. *Journal of Clinical Child Psychology*, 20, 245-253.

Depaul, G. J. (1992). How to assess attention deficit hyperactivity disorder within school settings. *School Psychology Quarterly*, 7, 60-74.

Doueck, H. J., English, D. J., Defanfil, D., & Moote, G. T. (1993). Decision-making in child protective services: A comparison of selected risk-assessment systems. *Child Welfare*, 122, 441-451.

Drasgow, F., & Hulin, C. L. (1987). Cross-cultural measurement. *Interamerican Journal of Psychology*, 21(1), 1-24.

Duchnowski, A. J., & Friedman, R. M. (1990). Children's mental health: Challenges for the nineties. *Journal of Mental Health Administration*, 17, 3-12.

Dupaul, G. J., Power, T. J., Anastopolous, A. D., Reid, R., McGoe, K. E., & Ikeda, M. J. (1997). Teacher ratings of attention deficit hyperactivity disorder symptoms: Factor structure and normative data. *Psychological Assessment*, 9, 436-444.

Eaton, W. W. (1980). *The sociology of mental illness*. New York: Praeger.

Eckert, T. L., & Depaul, G. J. (1996). Youth-completed and narrow-band child behavior questionnaires. In M. J. Breen & C. R. Fiedler (Eds.). *Behavioral Approach to Assessment of Youth With Emotional/Behavioral Disorders* (pp. 289-358). Austin: Pro-ed.

Elliott, S. N., Busse, R. T., & Gresham, F. M. (1993). Behavior rating scales: Issues of use and development. *School Psychology Review*, 22(2), 313-321.

Epstein, J. N. March, J. S., Conners, C. K., & Jackson, D. L. (1998). Racial differences on the Conners Teacher Rating Scale. *Journal of Abnormal Child Psychology*, 26(2), 109-118.

Evans, I. (1993). Constructional perspectives in clinical assessment. *Psychological Assessment*, 5, 264-272.

Farmer, A. D., & Bierman, K. L. (2002). Predictors and consequences of aggressive-withdrawn problem profiles in early grade school. *Journal of Clinical Child and Adolescent Psychology*, 31(3), 299-311.

Floyd, F. J. & Widaman, K. F. (1995). Factor analysis in the development and refinement of clinical assessment instruments. *Psychological Assessment*, 7(3), 286-299.

Furstenberg, F. F., Cook, T. D., Eccles, J., Elder, G. H., & Sameroff, A. (1999). *Managing to make it: Urban families and adolescent success*. Chicago: University of Chicago Press.

Glisson, C. (1996). Judicial and service decisions affecting children in state custody: The limited role of mental health. *Social Service Review*, 70, 257-281.

Glisson, C., Hemmelgarn, A., & Post J. A. (2002). The Shortform Assessment for Children (SAC): An assessment and outcome measure for child welfare and juvenile justice. *Research on Social Work Practice*, 12(1), 82-106.

Goldman, H. H., & Skodal, A. E. (1992). Revising Axis V for DSM-IV: A review of measures of social functioning. *American Journal of Psychiatry*, 149, 1148-1156.

Goldstein, H. (1996). Group differences and bias in assessment. In H. Goldstein & T. Lewis (Eds.). *Assessment: problems, Developments, and Statistical Issues* (pp. 85-93). New York: Wiley & Sons.

Gottfredson, L. S. (1994). The science and politics of race norming. *American Psychologist*, 49, 955-963.

Guerra, N. G., & Jagers, R. (1998). The importance of culture in the assessment of children and youth. In C. V. McLoyd & L. Steinberg (Eds.). *Studying minority adolescents: Conceptual, methodological and theoretical issues*. (pp. 183-210). Mahwah, NJ: Erlbaum.

Guertin, W. H. (1952). A factor analytic study of schizophrenic symptoms. *Journal of Consulting Psychology*, 16, 308-312.

Harrington, G. M. (1984). An experimental model of bias in mental testing. In C. R. Reynolds & R. T. Brown (Eds.). *Perspectives on mental testing* (pp. 101-138). New York: Plenum.

Harrington, G. M. (1988). Two forms of minority-group test bias as psychometric artifacts with an animal model (*Rattus norvegicus*). *Journal of Comparative psychology*, 102, 400-407.

Hartman, C. A., Hox, J., Auerbach, J., Erol, N., Fonseca, A. C., Mellenbergh, G., Novik, T. S., Oosterlaan, J., Roussos, A. C., Shalev, R. S., Zilber, 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 and Allied Disciplines*. 40(7), 1095-1116.

Harvey, R. J., & Hammer, A. L. (1999). Item response theory. *Counseling Psychologist*, 27(3), 353-383.

Hemmelgarn, A. L., Glisson, C., & Sharp, S. R. (2003). The shortform assessment for children: A validity comparison with the child behavior checklist and teacher report form. *Journal of Research on Social Work Practice*, 13(4), 510-530.

Heubeck, B.G. (2000). Cross-cultural generalizability of CBCL syndromes across three continents: From USA and Holland to Australia. *Journal of Abnormal Child Psychology*, 28(5), 439-450.

Hewitt, L. E. & Jenkins, R. L. (1946). *Fundamental patterns of maladjustment: The dynamics of their origin*. Springfield: State of Illinois.

Hoagwood, K., Horwitz, S., Stiffman, A., Weisz, J., Bean, D., Rae, D., Compton, W., Cottler, L., Bickman, L., & Leaf, P. (2000). Concordance between parent reports of children's mental health services and service records: The service assessment for children and adolescents (SACA). *Journal of Child and Family Studies*, 9(3), 315-331.

Hobbs, N. (1975). *The futures of children: Categories, labels, and their consequences*. San Francisco: Josey-Bass.

Hoge, R. D. (2002). Standardized instruments for assessing risk and need in youthful offenders. *Criminal Justice and Behavior*, 29(4), 380-396.

Horney, K. (1945). *Our inner conflicts*. New York: Norton.

Howell, J. C. (2001). Juvenile justice programs and strategies. In R. Loeber & D. P. Farrington (Eds.). *Child delinquents: Development, intervention, and service needs*. (pp. 305 – 322). Thousand Oaks: Sage.

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

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

Hui, C. H., & Triandis, H. C. (1989). Effects of culture and response format on extreme response style. *Journal of Cross-Cultural Psychology*, 20, 296-309.

Jarvinen, D. W. , & Sprague, R. L. (1995). Using ACTers to screen minority children for ADHD: An examination of item bias. *Journal of Psychoeducational Assessment*, 13, (Special Issue on ADHD), 172-184.

Jensen, P. S., Watanabe, H. K., Richters, J. E., Roper, M., Hibbs, E. D., Salzberg, A. D., & Liu, S. (1996). Scales, diagnoses, and child psychopathology: II. Comparing the CBCL and the DISC against external validators. *Journal of Abnormal Child Psychology* , 24(2), 151-168.

Justice for Juveniles Initiative. (1999). *How communities respond to mental health and substance abuse needs of youth in the juvenile justice system-executive summary*. National Mental Health Association.

Kampaas, R. W., Petoskey, M. D.; & Rowe, E. W. (2000). Current trends in psychological testing of children. *Professional Psychology: Research and Practice*, 31(2), 155-164.

Kassius, M.C., Ferdinand, R.F., van de Berg, H. and Verhalst, F.C. (1997). Association between difficult diagnostic approaches for child and adolescent psychopathology. *Journal of Child Psychology and Psychiatry*, 34(6), 625-632.

Kaufman, J. S., Cooper, R. S. (2001). Considerations for racial/ethnic classification in etiologic research. *American Journal of Epidemiology*, 154, 291-298.

Keenan, K., & Shaw, D. (1997). Developmental and social influences on young girls' early problem behavior. *Psychological Bulletin*, 121(1), 95-113.

Keith, T. Z., Fugate, M. H., DeGraff, M., Diamond, C. M., Shadrach, E. A., & Stevens, M. L. (1995). Using multi-sample confirmatory factor analysis to test for construct bias: An example using the K-ABC. *Journal of Psychoeducational Assessment, 13*, 347-364.

Knight, G. P., & Hill, N. E. (1998). Measurement equivalence in research involving minority adolescents. In C. V. McLoyd & L. Steinberg (Eds.). *Studying minority adolescents: Conceptual, methodological and theoretical issues*. (pp. 183-210). Mahwah, NJ: Erlbaum.

Kovacs, M. (1992). *Children's Depression Inventory*. Los Angeles: Multi-Health Systems.

Kovacs, M., & Devlin, B. (1998). Internalizing disorders in childhood. *Journal of Child Psychology and Psychiatry, 39*(1), 47-63.

Lambert, M. C., Markle, F., & Bellas, V. F. (2001). Psychopathology and assessment in African American children and families: A historical, ecological, and strengths-based perspective. A. Neal-Barnett, J. Contreras, K. A. Kerns (Eds.), *Forging links: African American children clinical developmental perspectives*. (pp. 246-267). Westport, CT: Prager.

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

Latkovich, S. A. (1996). An investigation of the concurrent, predictive, and construct validity of the Child Behavior Checklist for African-American children. (Doctoral Dissertation, University of Akron, 1996). *Dissertation Abstracts International, 57*, 11B.

Loeber, R., Stouthamer-Loeber, M., & White, H.R. (1999). Developmental aspects of delinquency and internalizing problems and their association with persistent juvenile substance abuse between ages 7 & 18. *Journal of Clinical Child Psychology, 28*(3), 322-332.

Loehlin, J. C. (1998). *Latent variable models: An introduction to factor path and structural analysis*. New Jersey: LEA.

MacAdoo, H. P. (2002). The village talks: Racial socialization of our children. In H. P. MacAdoo (Ed.), *Black children: Social, educational, and parental environments*. (2nd ed., pp. 47-55). Thousand Oaks, CA: Sage.

Malgady, R.G. (1996). The question of cultural bias in assessment and diagnosis of ethnic minority clients: Let's reject the null hypothesis. *Professional Psychology Research and Practice*, 27(1), 73-77.

Marsella, A. J., & Kameoka, V. A. (1989). Ethnocultural issues in the assessment of psychopathology. In S. Wetzler (Ed.), *Measuring mental illness in psychometric assessment for clinicians* (pp. 231-256). Washington, DC: American Psychiatric Press.

Martin, P., Peters, C. L., & Glisson, C. (1998). Factors affecting case management recommendations for children entering state custody. *Social Service Review*, 72, 521-544.

Mash, E. J. & Terdal, L. G. (1997). Assessment of child and family disturbance: A behavioral-systems approach. In E. J. Mash & L. G. Terdal (Eds.), *Assessment of Childhood Disorders* (3rd ed., pp. 3-68). New York: Guilford Press.

Mash, E. J., & Terdal, L. G. (Eds.). (1997). *Assessment of Childhood Disorders* (3rd ed.). New York: Guilford Press.

Mattison, R.E., & Spitznagel, E.L. (1999). Long-term stability of child behavior checklist profile types in a child psychiatric clinic population. *Journal of the American Academy of Child & Adolescent Psychiatry*, 38(6), 700-707.

Mayfield, J. W., & Reynolds, C. R. (1998). Are ethnic differences in diagnosis of childhood psychopathology an artifact of psychometric methods? An experimental evaluation

of Harrington's hypothesis using parent-report symptomology. *Journal of School Psychology, 36*(3), 313-334.

McConaughy, S. H., Stanger, C., & Achenbach, T. M. (1992). Three-year course of behavioral/emotional problems in a national sample of 4- to 6-year-olds: I. Agreement among informants. *Journal of the Academy of Child and Adolescent Psychiatry, 31*, 932-940.

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

McLoyd, V. C. (1998). Changing demographics in the American population: Implications for research on minority children and adolescents. In V. C. McLoyd & L. Steinberg (Eds.), *Studying minority adolescents: Conceptual, methodological, and theoretical issues* (pp. 89-109). Mahwah, NJ: Erlbaum.

McMahon, R.J. (1994). Diagnosis, assessment, and treatment of externalizing problems in children: The role of longitudinal data. (Special section: Childhood psychopathology). *Journal of Consulting and Clinical Psychology, 62*(5), 901-917.

Merrell, K. W. (1999a). *Behavioral, social, and emotional assessment of children*. New Jersey: LEA.

Merrell, K. W. (1999b). Informant report: Rating scale measures. In E. S. Shapiro & T. R. Kratochwill (Eds.), *Conducting School-Based Assessments of Child and Adolescent Behavior* (pp. 203-234). New York: Guilford Press.

Meyers, K. & Winters, N. (2002). Ten-year review of rating scales. I: Overview of scale functioning, psychometric properties, and selection. *Journal of the American Academy Child and Adolescent Psychiatry, 41*(2), 114-122.

Messick, S. (1989). Validity. In R. L. Linn (Ed.). *Educational Measurement*. (3rd ed.). (pp. 13-103), New York: McMillan.

Moras, K., & Barlow, D. H. (1992). Dimensional approaches to diagnosis and the problem of anxiety and depression. In W. Feigenbaum, A. Ehlers, J. Margraf, & I. Florin (Eds.). *Behavioral assessment: A practical handbook*. (3rd ed.). (pp. 280-312).

Neter, J. & Wasserman, W. (1987). *Applied linear statistical models*. Homewood, IL: Irwin.

Nugent, W. R. (1993). Differential validity in social work measurement. *Social Service Review*, 631-650.

Nugent, W. R., & Glisson, C. (1999). Reactivity and responsiveness in children's service systems. *Journal of Social Service Research*, 25(3), 41-60.

O'Donnell, J. P., Stein, M. A., Machabanski, H., & Cress, J. N. (1982). Dimensions of behavior problems in Anglo-American and Mexican-American preschool children: A comparative study. *Journal of Consulting and Clinical Psychology*, 50(5), 643-651.

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

Ollendick, T.H., & King, N.J. (1994). Diagnosis, assessment, and treatment of internalizing problems in children: The role of longitudinal data. *Journal of Consulting and Clinical Psychology*, 62(5), 918-927.

Oommen, T. K. (1989). Ethnicity, immigration, and cultural pluralism: India and the United States. In M. L. Kohn (Ed.), *Cross-national research in sociology*. Sage Publications.

Pelham, W.E., Milich, R., Murphy, D.A., & Murphy, H.A. (1989). Normative data on the IOWA Conners Teaching Rating Scale. *Journal of Clinical Child Psychology*, 18, 259-262.

Peters, M. F. (2002). Racial socialization of young black children. In H. P. MacAdoo (Ed.), *Black children: Social, educational, and parental environments*. (2nd ed., pp. 57-72). Thousand Oaks, CA: Sage.

Peterson, M. F., Smith, P. B., Ayestaran, S., Bochner, S., Callan, V., Cho, N. G., Jesuino, J. C., D'Amorim, M., Francois, P. H., Hofmann, P. L., Leung, K., Lim, T. K., Mortazavi, S., Munene, J., Radford, M., Ropo, A., Savage, G., Setiadi, B., Sinha, T. N., Sorenson, R., & Viedge, C. (1995). Role conflict, ambiguity, and overload. A 21 nation study. *Academy of Management Journal*, 38(2), 429-452.

Phillips, L., & Rabenovich, M.S., (1958). Social role and patterns of symptomatic behaviors. *Journal of Abnormal and Social Psychology*, 57, 181-186.

Quay, H. C. (1979). Classification. In H. C. Quay and J. S. Werry (Eds.), *Psychopathological disorders of childhood* (2nd ed., pp 1-42). New York: Wiley.

Ramsey, M. C., Reynolds, C. R., & Kamphaus, R. W. (2002). *Essentials of Behavioral Assessment*. New York: Wiley & Sons.

Raykov, T., & Marcoulides, G. A. (2000). *A first course in structural equation modeling*. Mahwah, NJ: LEA.

Reise, S. P., Widaman, K. F., & Pugh, R. H. (1993). Confirmatory factor analysis and item response theory: Two approaches to exploring measurement invariance. *Psychological Bulliten*, 114, 552-566.

Reid, R., Dupaul, G. J., Power, T. J., Anastopoulos, A. D., Rogers-Adkinson, D., Noll, M. & Ricco, C. (1998). Assessing culturally different students for attention deficit hyperactivity disorder using behavior rating scales. *Journal of Abnormal Child Psychology*, 26(3), 187-198.

Reid, R., Casat, C. D., Norton, H. J., Anastopolous, A. D., & Temple, E. P. (2001). Using behavior rating scales for ADHD across ethnic groups: The IOWA Conners. *Journal of Emotional and Behavioral Disorders*, 9(4), 210-218.

Reynolds, C. R., & Kamphaus, R. (1992). *Behavior Assessment System for Children*. Circle Pines, MN: American Guidance Service.

Reynolds, C. R., & Richmond, B. O. (1985). *Revised Children's Manifest Anxiety Scale*. Los Angeles: Western Psychological Services.

Rigdon, E. E., & Ferguson, C. E. (1991). The performance of polychoric correlation coefficient and selected fitting functions in confirmatory factor analysis with ordinal data. *Journal of Marketing Research*, 28, 491-497.

Ross, A. O. (1980). *Psychological disorders of children*. New York: McGraw-Hill.

Rowe, D. C. (1997). Group differences in developmental processes: The exception or the rule? *Psychological Inquiry*, 8(3), 218-222.

Rubin, A., & Babbie, E. (1997). *Research methods for social work*. (3rd ed.). New York, NY: Brooks/Cole.

Samantrai, K. (1992). To prevent unnecessary separation of children from families: Public law 96-272. *Social Work*, 34, 295-302.

Segal, D. L. & Falk, S. B. (1998). Structured interviews and rating scales. In M. J. Breen & C. R. Fiedler (Eds.). *Behavioral Approach to Assessment of Youth With Emotional/Behavioral Disorders* (4th ed., pp. 158-178). Austin: Pro-ed.

Shaffer, D., Fisher, P., Lucas, C.P., Dulcan, M.K., & Schwab-Stone, M.E. (2000). NIMH Diagnostic Interview Schedule for Children version IV (NIMH DISC-IV): Description, differences from previous versions, and reliability of some common diagnoses. *Journal of the American Academy of Child and Adolescent Psychiatry*, 39(1), 28-38.

Shapiro, E. S., & Kratochwill, T. R. (2000). Introduction: Conducting a multidimensional behavioral assessment. In E. S. Shapiro & T. R. Kratochwill, (Eds.). *Conducting school-based assessments of child and adolescent behavior*, (pp. 1-20). New York: Guilford Press.

Skiba, R. J., Kesting, K. & Bush, L. D. (2002). Culturally competent assessment: more than non-biased tests. *Journal of Child and Family Studies*, 11(1), 61-78.

Silverman, W. K. & Serafini, L. T. (1998). Assessment of child behavior problems: Internalizing disorders. In M. J. Breen & C. R. Fiedler (Eds.). *Behavioral Approach to Assessment of Youth With Emotional/Behavioral Disorders* (4th ed., pp. 342-360). Austin: Pro-ed.

Snyder, H. N. (2000). *Juvenile Arrests 1999*. Washington, DC: Office of Juvenile Justice and Delinquency.

Stanger, C., McConaughy, S. H., & Achenbach, T. M. (1992). Three-year course of behavioral/emotional problems in a national sample of 4- to 6-year-olds: II. Predictors of syndromes. *Journal of American Academy of Child and Adolescent Psychiatry*, 31, 941-950.

Stiffman, A. R., Horwitz, S. M., Hoagwood, K., Compton, W., Cottler, L., Bean, D. L., Narrow, W. E., & Weisz, J. R. (2000). The service assessment for children and adolescents (SACA): Adult and child reports. *Journal of the Academy of Child and Adolescent Psychiatry*, 39(8), 1032-1044.

Tatum, B. D. (1997). *Why are all the black kids sitting together in the cafeteria?* New York, NY: Basic Books.

Tennessee Council of Juvenile Court Judges. (2002). *2001 Tennessee annual juvenile court statistical report*. Nashville, TN: Author.

Teplin, L. A., Abram, K. M., McClelland, G. M., Dulcan, M. K., & Mericle, A. A. (2002). Psychiatric disorders in youth in juvenile detention. *Archives in General Psychiatry*, 59, 1133-1143.

Trendafilov, N. (2003). Dynamical systems approach to factor analysis parameter estimation. *British Journal of Mathematical and Statistical Psychology*, 56, 27-46.

Tucker, L. R. (1951). *A method for syntheses of factor analyses studies* (Report No. 984). Washington, DC: Department of the Army, Personnel Research Section.

Tyson, E. H. (in press). Ethnic differences in using behavior rating scales to assess the mental health of children: A conceptual and psychometric critique. *Journal of Child Psychiatry and Human Development*, 34(4).

Tyson, E. H., Dulmus, C. N., Wodarski, J. S. (2002). Assessing violent behavior. In L. A. Rapp-Paglicci, A. R. Roberts, & J. S. Wodarski (Eds.), *Handbook of Violence*. (pp. 148-168). New York, NY: John Wiley & Sons.

Ulmann, R. K., Sleator, E. K., & Sprague, R. L. (1991). *Manual for the ADD-H Comprehensive Teacher's Rating Scale (ACT-eRS)*. Champaign, IL: MeriTech, Inc.

- U.S. Bureau of the Census. (2000). *1999 Census Profile (No. 117)*. Washington, DC: U.S. Department of Commerce.
- U.S. Department of Health and Human Services. (1999). *The AFCARS Report*. Washington, DC: Author.
- U.S. Department of Health and Human Services, Administration on Children, Youth and Families. (2001). *Child maltreatment*. Washington, DC: Author.
- U.S. Department of Justice. (1999). *COJJDP Fact Sheet No 96*. Washington, DC: Author.
- van de Vijver, F. J. R. & Leung, K. (1997). *Methods and Data Analysis for Cross-Cultural Research*. Thousand Oaks: Sage.
- van de Vijver, F. J. R. & Poortinga, Y. H. (1997). Towards and integrated analysis of bias in cross-cultural assessment. *European Journal of Psychological Assessment*, 13, 21-29.
- Verhulst, F.C., & Achenbach, T.M. (1995). Empirically based assessment and taxonomy of psychopathology: Cross cultural applications. A review. *European Child and Adolescent Psychiatry*, 4, 61-76.
- Vogt, W. P. (1999). *Dictionary of Statistics and Methodology: A Non-technical Guide for Social Sciences*. (2nd ed.). Thousands Oaks: Sage.
- Weisz, J. R., & McCarty, C. A. (1999). Can we trust parent reports in research on cultural and ethnic differences in child psychopathology? Using the bicultural family design to test parental culture effects. *Journal of Abnormal Psychology*, 108(4), 598-605.
- Weisz, J. R., & McCarty, C. A., Eastman, K. E., Suwanlert, S., & Chaityasit, W. (1997). Developmental psychopathology and culture: Ten lessons from Thailand. In S. S.

Luthar, J. A. Barack, D. Cicchetti, & J. Weisz (eds.). *Developmental psychopathology: Perspectives on risk and disorder* (pp.568-592). Cambridge: Cambridge University Press.

Widiger, T. A., Frances, A. J., Pincus, H. A., Davis, W. W., First, M. B. (1991). Toward and empirical classification for the DSM-IV. *Journal of Abnormal Psychology*, 100, 280-288.

Wittenborn, J. R. & Holzberg, J. D. (1951). The generality of psychiatric syndromes. *Journal of Consulting Psychology*, 15, 372-380.

Yau-Fai Ho, D. (1994). Introduction to cross-cultural psychology. In L. L. Alder & U. P. Gielen (Eds.), *Cross-Cultural Topics in Psychology*. (pp. 3 -13). Praeger: Westport, CT.

Zide, M. R., & Gray, S. W. (2001). *Psychopathology: A competency-based assessment model for social workers*. Belmont, CA: Brooks/Cole.

Zwick, R., Donogue, J. R., & Grima, A. (1993). Assessment of differential item functioning for performance tasks. *Journal of Educational Measurement*, 30, 233-251.

APPENDICES

APPENDIX A

TABLES

Table 1.

Demographic Profile, Reason For Referral, and Prior Custody of Participants.

Variable	AA (N=562)			CA (N=692)		
	%	<i>M</i>	<i>SD</i>	%	<i>M</i>	<i>SD</i>
Age (range 4-18)		14.4	2.4		14.5	2.4
Family Members (range 0-14)		5.5	2.4		4.3	1.6
Mo. Income (range 0-\$9500) [†]		\$1414	1015		1773	1372
Gender ^{††}						
Male	70.1			62.0		
Female	29.9			38.0		
Type of County ^{††}						
Urban	93.4			33.2		
Rural	6.6			66.8		
Primary Caregiver/Informant ^{††}						
Bio-Mother	68.0			75.6		
Bio-Father	6.0			8.7		

[†]Significantly different at $p < .01$. ^{††}Chi-square significantly different at $p < .01$

Table 1. (cont.)

Variable	AA (N=562)			CA (N=692)		
	%	<i>M</i>	<i>SD</i>	%	<i>M</i>	<i>SD</i>
Grandmother	12.5			6.9		
Other Relative	8.2			2.5		
Parents Live Together ⁺⁺	6.4			21.1		
Reason For Court Referral						
Delinquent	9.4			9.8		
Status Offense	27.2			33.3		
Dependency/Neglect	60.7			54.2		
Previously in State Custody	14.6			13.3		

Table 2.

Means, Standard Errors and Standard Deviations of SAC Scores by Ethnicity.

	Total SAC			Internalizing			Externalizing		
	<i>M</i>	<i>SE</i>	<i>SD</i>	<i>M</i>	<i>SE</i>	<i>SD</i>	<i>M</i>	<i>SE</i>	<i>SD</i>
AA (n=562)	33.32 ^a	.80	19.00	11.98 ^b	.38	8.97	21.34	.52	12.44
CA (n=692)	36.34 ^a	.69	18.05	14.19 ^b	.36	9.46	22.15	.46	12.05

^{a,b} group means significantly different at $p < .001$.

Table 3.

Internal Consistency and Test-Retest Reliabilities of the SAC for Both Groups.

	Internal Consistency		Test-Re-test	
	AA	CA	AA	CA
Total SAC	$\alpha = .95$	$\alpha = .94$	$r = .66^{***}$	$r = .63^{***}$
Internalizing	$\alpha = .90$	$\alpha = .90$	$r = .60^{***}$	$r = .63^{***}$
Externalizing	$\alpha = .94$	$\alpha = .94$	$r = .68^{***}$	$r = .62^{***}$

*** $p < .001$.

Table 4.

Intercorrelations of the SAC and Convergent-Divergent Validity With the Columbia Impairment Scale (CIS) for AA (n = 562) and CA (n=692)*.

	Total SAC	INT	EXT	CIS
Total SAC	-	.79***	.88***	.65***
INT	.84***	-	.40***	.43***
EXT	.92***	.57***	-	.63***
CIS	.63***	.48***	.62***	-

*CA in top diagonal and AA in bottom diagonal. INT = Internalizing; EXT = Externalizing. *** $p < .001$.

Table 5.

Frequency and Univariate Statistics of Postdictive-Validity Criterion Variables.

Criterion Variable	AA		CA		<i>t</i> / χ^2	<i>p</i>
	<i>M</i> (<i>SD</i>)	% (n)	<i>M</i> (<i>SD</i>)	% (n)		
Status Offenses	5.3(19.0)	(558)	2.8(13.7)	(679)	2.65	<.05
Non-Violent Offenses	3.1(5.4)	(556)	1.9(5.8)	(677)	3.74	<.001
Violent Offenses	1.1(1.4)	(557)	.7(2.6)	(678)	3.06	<.01
Violent Behavior		37.7(212)		35.5(242)	2.43	ns
SED		10.5(59)		7.5(52)	3.42	ns
ADHD		19.4(109)		30.5(211)	20.10	<.001
School MH Services		29.0(163)		30.2(209)	.21	ns
Any MH Service*		76.9(432)		79.3(549)	1.11	ns
Comm. Mental Hlth. Center		32.0(180)		42.5(294)	14.24	<.001
Mental Health Professional		20.6(116)		29.2(202)	11.98	<.01
Any Custody**		86.5(486)		55.5(384)	30.51	<.001
Psychiatric Hospital		10.7(60)		17.3(120)	11.21	<.01
Detention/Prison/Jail		83.8(471)		41.8(289)	229.63	<.001
Taking Any Medication		13.9(78)		29.2(202)	41.93	<.001

Table 6.

Postdictive Criterion Validity Coefficients of Total SAC for AA and CA Children and Tests for Significant Differences.

Criterion Variable	Total SAC		z-statistic	p
	AA	CA		
Status Offenses	.15	.14	-	ns
Non-Violent Offenses	.14	.14	-	ns
Violent Offenses	.14	.15	-	ns
Violent Behavior	.25	.25	-	ns
SED	.24	.23	-	ns
ADHD	.21	.22	-	ns
School MH Services	.19	.20	-	ns
Any MH Service	.22	.26	-	ns
Comm. Mental Hlth. Cntr.	.27	.27	-	ns
Mental Health Professional	.25	.18	-3.02	<.001
Any Custody	.22	.26	-	ns
Psychiatric Hospital	.22	.21	-	ns
Detention/Prison/Jail	.20	.16	-	ns
Taking Any Medication	.26	.28	-	ns

Table 7.

Postdictive Criterion Validity Coefficients of Internalizing Subscale for AA and CA Children and Tests for Significant Differences.

Criterion Variable	Total SAC		z-statistic	<i>p</i>
	AA	CA		
SED	.25	.21	-	ns
School MH Services	.18	.16	-	ns
Any MH Service*	.16	.19	-	ns
Comm. Mental Hlth. Cntr.	.20	.23	-	ns
Mental Health Professional	.18	.12	-	ns
Any Custody**	.15	.17	-	ns
Psychiatric Hospital	.20	.17	-	ns
Taking Any Medication	.25	.26	-	ns

Table 8.

Postdictive Criterion Validity Coefficients of Externalizing Subscale for AA and CA

Children and Tests for Significant Differences.

Criterion Variable	Total SAC		z-statistic	p
	AA	CA		
Status Offenses	.17	.14	-	ns
Non-Violent Offenses	.18	.15	-	ns
Violent Offenses	.16	.16	-	ns
Violent Behavior	.27	.29	-	ns
SED	.18	.18	-	ns
ADHD	.20	.24	-	ns
School MH Services	.16	.17	-	ns
Any MH Service*	.22	.24	-	ns
Comm. Mental Hlth. Cntr.	.27	.23	-	ns
Mental Health Professional	.26	.18	-3.42	<.001
Any Custody**	.22	.26	-	ns
Psychiatric Hospital	.19	.18	-	ns
Detention/Prison/Jail	.21	.20	-	ns
Taking Any Medication	.22	.22	-	ns

Table 9.

Frequency and Univariate Statistics of Predictive-Validity Criterion Variables.

Criterion Variable	AA	CA	χ^2	<i>p</i>
	% (n)	% (n)		
Any Offense	47.9 (205)	47.4 (252)	-	ns
Violent Behavior	14.3 (59)	12.7 (66)	-	ns
SED	9.3 (40)	5.8 (31)	4.41	<.05
ADHD	17.5 (75)	29.4 (158)	18.58	<.001
Any MH Service*	34.0 (146)	41.0 (221)	5.05	<.05
Comm. Mental Hlth. Cntr.	15.1 (65)	28.9 (156)	25.97	<.001
Any Custody**	38.2 (165)	36.1 (199)	-	ns
Detention/Prison/Jail	31.2 (134)	18.7 (101)	20.10	<.001
Adjudicated Unruly	14.3 (60)	21.1 (107)	7.32	<.01
Alcohol Use	10.3 (44)	18.6 (100)	13.16	<.001
Drug Use	19.3 (83)	24.2 (130)	-	ns

Table 10.

Predictive Criterion Validity Coefficients of Total SAC for AA and CA Children and Tests for Significant Differences .

Criterion Variable	Total SAC		z-statistic	<i>p</i>
	AA	CA		
Any Offense	.09	.20	-4.74	< .001
Violent Behavior	.17	.23	-2.30	< .05
SED	.19	.16	-	ns
ADHD	.20	.21	-	ns
Any MH Service*	.17	.21	-	ns
Comm. Mental Hlth. Cntr.	.16	.24	-3.06	< .01
Any Custody**	.21	.25	-	ns
Detention/Prison/Jail	.17	.18	-	ns
Adjudicated Unruly	.24	.22	-	ns
Alcohol Use	.16	.13	-	ns
Drug Use	.11	.11	-	ns

Table 11.

Predictive Criterion Validity Coefficients of Internalizing Subscale for AA and CA Children and Tests for Significant Differences.

Criterion Variable	Internalizing		z-statistic	p
	AA	CA		
SED	.19	.14	-	ns
Any MH Service*	.12	.10	-	ns
Comm. Mental Hlth. Cntr.	.14	.20	-2.44	< .05
Any Custody**	.12	.15	-	ns
Alcohol Use	.13	.09	-	ns
Drug Use	.19	.14	-	ns

Table 12.

Predictive Criterion Validity Coefficients of Externalizing Subscale for AA and CA Children and Tests for Significant Differences.

Criterion Variable	Externalizing		z-statistic	<i>p</i>
	AA	CA		
Any Offense	.13	.23	-3.97	< .001
Violent Behavior	.18	.23	-	ns
SED	.15	.13	-	ns
ADHD	.16	.23	-2.70	< .01
Any MH Service*	.18	.24	-2.26	< .05
Comm. Mental Hlth. Cntr.	.14	.20	-2.44	< .05
Any Custody**	.23	.26	-	ns
Detention/Prison/Jail	.19	.20	-	ns
Adjudicated Unruly	.24	.25	-	ns
Alcohol Use	.15	.13	-	ns
Drug Use	.10	.11	-	ns

Table 13.

Initial Simultaneous Factor Analysis (SFA) of the SAC Across AA (N = 562) and CA (N = 692) Children.

Model	<i>df</i>	X^2	<i>p</i>	Fit Indices			
				RMR	GFI	CFI	RMSEA
All Parameters Free	2138	6991.4	< .001	.039	.782	.818	.043
Constrained Phi	2139	7006.5	< .001	.043	.782	.817	.043
Constrained Factor Pattern	2186	7112.9	< .001	.042	.779	.815	.042
Constrained Phi and Pattern	2187	7127.0	< .001	.044	.779	.814	.042

Table 14.

Initial Chi-square Difference Tests for Parameter Free and Constrained Models.

Models Compared	<i>df</i>	X^2	Δdf	ΔX^2	<i>p</i>	X^2/df
Comparison A						
All Parameters Free	2138	6991.4				
Constrained Phi	1699	7006.5	1	15.1	<.001	
Comparison B						
All Parameters Free	2138	6991.4				
Constrained Factor Pattern	2186	7112.9	48	121.5	<.001	2.51
Comparison C						
All Parameters Free	2138	6991.4				
Constrained Pattern and Phi	2187	7127.0	49	135.6	<.001	2.71

Table 15.

Ethnic Group Differences in Internalizing (INT) Standardized Regression Weights Within the All Free Estimates Model of the Simultaneous Factor Analysis*.

Factor	Item	Ethnic Group		Difference
		CA	AA	
INT	3	.508	.545	-.037
INT	4	.556	.508	.048
INT	11	.543	.616	.073
INT	12	.388	.590	-.202
INT	13	.631	.710	-.079
INT	14	.709	.723	-.014
INT	17	.460	.301	.159
INT	19	.541	.535	.006
INT	20	.551	.512	.039
INT	21	.529	.533	.004
INT	22	.392	.445	-.053
INT	23	.360	.386	-.026
INT	24	.435	.352	.083
INT	25	.415	.335	.080

*Items in bold had differences greater than .10 between the two groups.

Table 15. (cont.).

Factor	Item	Ethnic Group		Difference
		CA	AA	
I INT	26	.312	.320	-.008
NT	28	.516	.553	-.037
INT	30	.503	.546	-.043
INT	31	.426	.530	-.104
INT	33	.393	.339	.054
INT	34	.527	.473	.054
INT	44	.479	.480	.001
INT	45	.756	.751	.005
INT	47	.536	.530	.006
INT	48	.589	.656	-.067

Table 16.

Ethnic Group Differences in Externalizing (EXT) Standardized Regression Weights Within the All Free Estimates Model of the Simultaneous Factor Analysis*.

Factor	Item	Ethnic Group		Difference
		CA	AA	
EXT	1	.676	.667	.009
EXT	2	.495	.560	-.065
EXT	5	.731	.711	.020
EXT	6	.594	.508	.086
EXT	7	.696	.654	.042
EXT	8	.661	.650	.011
EXT	9	.541	.620	-.079
EXT	10	.591	.562	.029
EXT	15	.615	.679	-.064
EXT	16	.451	.499	-.048
EXT	18	.641	.680	-.039
EXT	27	.640	.636	.004
EXT	29	.688	.627	.061

*Items are in bold had differences greater than .1 between the two groups.

Table 16. (cont.).

Factor	Item	Ethnic Group		Difference
		CA	AA	
EXT	32	.510	.622	-.112
EXT	35	.518	.501	.017
EXT	36	.626	.648	-.022
EXT	37	.602	.618	-.016
EXT	38	.647	.750	-.103
EXT	39	.453	.541	-.088
EXT	40	.668	.683	-.015
EXT	41	.725	.726	-.001
EXT	42	.765	.767	.020
EXT	43	.787	.772	.015
EXT	46	.670	.630	.040

Table 17.

Follow-Up^a Simultaneous Factor Analysis (SFA) of the SAC Across AA (N = 562) and CA (N = 692) Children.

Model	<i>df</i>	<i>X</i> ²	<i>p</i>	Fit Indices			
				RMR	GFI	CFI	RMSEA
All Parameters Free	1698	5820.5	< .001	.039	.793	.831	.044
Constrained Phi	1699	5833.7	< .001	.042	.793	.830	.044
Constrained Factor Pattern	1741	5893.2	< .001	.041	.791	.830	.044
Constrained Phi and Pattern	1742	5907.2	< .001	.042	.791	.829	.044

^aSFA conducted after deleting five items (items 12, 17, 31, 33, and 38) from the overall model.

Table 18.

Follow-Up Chi-square Difference Tests for Parameter Free and Constrained Models.

Models Compared	df	X^2	Δdf	ΔX^2	p	X^2/df
Comparison A						
All Parameters Free	1698	5820.5				
Constrained Phi	1699	5833.7	1	13.2	<.01	
Comparison B						
All Parameters Free	1698	5820.5				
Constrained Factor Pattern	1741	5893.2	43	72.6	<.01	1.69
Comparison C						
All Parameters Free	1698	5820.5				
Constrained Pattern and Phi	1742	5907.0	44	86.6	<.001	1.97

APPENDIX B

Family Interview Questionnaire Variables (Level of Measurement)-Postdictive Analyses

1. Status Offenses – (interval)

Tobacco Possession; Under-age Consumption; Violation of Aftercare; Unruly; Truancy; Runaway; Violation Court Order; Violation of Curfew

2. Non-Violent Offenses – (interval)

Theft; Arson; Burglary; Vandalism; Illegal Use of Credit Cards; Unauthorized Use of Vehicle; Criminal Attempt; Evading Arrest; DUI; Resisting Arrest; Filing False Report; Criminal Impersonation; Disorderly Conduct; Criminal Trespass; Gambling; Violation of Probation; Escape; Loitering

3. Violent Offenses – (interval)

Especially Aggravated Battery; Aggravated Robbery; Robbery; Rape; Aggravated Sexual Battery; Sexual Abuse of a Child; Sexual Offense; Aggravated Assault; Assault; Reckless Endangerment; Aggravated Arson; Aggravated Burglary; Harassment; Cruelty To Animals; Weapons Charges

4. Violent Behavior – (dichotomous)

What level of violence has your child committed? No Violence vs. Threat of Violence, Felony Violence, or Misdemeanor Violence

5. Seriously Emotionally Disturbed (SED) – (dichotomous)

Does this apply to your child? Yes or No

6. Attention Deficit-Hyperactivity Disorder (ADHD) – (dichotomous)

Does this apply to your child? Yes or No

7. Taking Any Medication – (dichotomous)

Caregiver report of medication their child was taking and reason. For an internalizing problem (e.g., anxiety or depression), an externalizing problem (e.g., anger management, hyperactivity), or mental health/psychotic problem (e.g., bipolar, hallucinations)

APPENDIX C

Family Interview Questionnaire Variables (Level of Measurement)-Predictive Analysis

1. Any Offense – (dichotomous)

Status offense (see Appendix A), *Non-Violent Offenses* (see Appendix A) or *Violent Offenses* (see Appendix A), or *Drug Offense*

2. Violent Behavior – (dichotomous)

What level of violence has your child committed? No Violence vs. Threat of Violence, Felony Violence, or Misdemeanor Violence

3. Seriously Emotionally Disturbed (SED) – (dichotomous)

Does this apply to your child? Yes or No

4. Attention Deficit Hyperactivity Disorder (ADHD) – (dichotomous)

Does this apply to your child? Yes or No

5. Adjudicated Unruly – (dichotomous)

Was your child adjudicated unruly since our last interview? Yes or No

6. Alcohol Use – (dichotomous)

Does this apply to your child (since our last interview)? Yes or No

7. Drug Use – (dichotomous)

Does this apply to your child (since our last interview)? Yes or No

APPENDIX D

Shortform Assessment for Children (SAC)

Use the behaviors listed below to describe this child as you know him or her.

0= Never

1= Sometimes

2= Often

Internalizing Items

3. Says he/she feels lonely
4. Cries or appears tearful a lot
11. Is afraid he/she might do something bad
12. Is uncomfortable with attention*
13. Says he/she is not loved by anyone
14. Thinks he/she is worthless/second-rate
17. Keeps to him/herself a lot
19. Is overly anxious or afraid
20. Blames him/herself a lot
21. Feels overly tired
22. Has pains without medical reason
23. Has headaches without medical reason
24. Feels sick without medical reason
25. Has stomachaches without medical*
26. Throws up without medical reason
28. Refuses to speak or answer
30. Is secretive and doesn't share thoughts
31. Is unsure of self or easily embarrassed
33. Doesn't stick up for him/herself
34. Stares into space or at nothing
44. Moves slowly or lacks energy
45. Is sad, unhappy, or feels down
47. Is withdrawn/keeps apart from people
48. Is easily worried

Externalizing Items

1. Argues too much
2. Brags or boasts
5. Is cruel or mean
6. Demands attention*
7. Doesn't follow rules
8. Destroys other's things*
9. Disobeys at school
10. Does not feel guilty*
15. Fights a lot
16. Hangs out with troublemakers
18. Lies or cheats
27. Attacks or hits others
29. Screams too much
32. Clowns around or shows off
35. Steals from others
36. Is stubborn or irritable
37. Has sudden mood swings
38. Swears or curses
39. Talks more than should*
40. Teases or provokes others*
41. Loses temper*
42. Threatens or frightens others
43. Has no respect for others
46. Is overly loud

*Items were abbreviated

APPENDIX E

Columbia Impairment Scale (CIS)

0=No Problem 1-3=Some Problem 4=A Very Big Problem 8=Not Applicable

How much of a problem does he/she have with....?

1. Getting into trouble
2. Getting along with his/her mother
3. Getting along with his/her father
4. Feeling unhappy or sad
5. With his/her behavior at school
6. With having fun
7. Getting along with adults other than you or his/her father/mother
8. With feeling nervous or worried
9. Getting along with his/her brother/sister
10. Getting along with other kids his/her age
11. Getting involved with activities like sports or hobbies
12. With his/her schoolwork (doing his/her job)
13. With his/her behavior at home

APPENDIX F

Services Assessment for Children and Adolescents (SACA). (Dichotomous, Y/N)

School Mental Health Services were: Being in a special school for students with behavioral/emotional problems

Being in a special class in a regular school for behavioral/emotional/drug problems

Special help (tutoring or training, etc.) for behavioral/emotional/drug problems

Counseling or therapy in school for behavioral/emotional/drug problems

Any Mental Health Service was defined as:

Psychiatric Hospital*

Psychiatric or Medical unit in a general hospital for emotional or behavioral problems

Drug or Alcohol treatment unit

Residential Treatment center

Group Home

Foster Home

Detention Center/Prison or Jail*

Emergency Shelter for behavioral or emotional problems

Community Mental Health Center or other outpatient mental health clinic*

Mental Health Professional like a psychologist, social worker, family counselor*

Partial Hospitalization or day treatment

Drug or alcohol clinic

In-Home Therapist or counselor or preservation worker

Emergency room for behavioral or emotional problems

Pediatrician for behavioral or emotional problems

Probation for behavioral or emotional problems

Priest, Minister or Rabbi for behavioral or emotional problems

Healer, Shaman, Spiritualist

Acupuncturist/chiropractor

Any self-help group like AA or NA

Respite Care Provider

Any custody:

State custody or any out of home placement

*These services were also individual criterion variables because of high base-rates

VITA

Edgar H. Tyson graduated from Miami-Dade Community College, Miami, FL in 1992 with an Associates Degree in Liberal Arts. He was awarded a Minority Access to Research Careers (MARC) Fellowship Grant to attend an honors program in psychology at Barry University, Miami Shores, FL. Edgar graduated Magna Cum Laude from Barry University in 1994 with a Bachelor of Science in Psychology degree. Edgar returned to Barry University in 1995 to pursue a Master in Social Work degree, and completed his degree while working full time as a therapist in a span of three years. He earned his MSW degree from Barry University in 1998, maintaining a perfect 4.0 G.P.A. Edgar was a member of the NASW governing body in Florida as the Student Representative, attending state meetings and setting state policy. He was Social Work Student of the Year for Broward County, FL in 1997 and Social Work Student of the Year for the State of Florida in 1998. Edgar was a therapist and counselor, working for several non-profit organizations in Miami, including, Tacolcy Center, Miami Bridge Youth Shelter, New Horizons Mental Health Community Center, Family Counseling Services, and Cool School Inc. He then returned to school to pursue his doctorate degree at The University of Tennessee, Knoxville, TN in 2000. While at UT, Edgar wrote a book chapter that was published in 2002 in the Handbook of Violence. He also was the sole-author of his first peer-reviewed article while a doctoral student at UT. Edgar gained additional practice experience as a behavioral therapist at Peninsular Behavioral Health Hospital, where he worked for two years. Edgar accepted a position at Florida State University as an Assistant Professor, where he is to begin his teaching career in August 2003. Edgar has a lovely wife, Lorraine Tyson and six children, Tahir, Aminah, Wayne, Shatriya, Steven, and Christine.

3444 1896 43
03/31/04

