

January 01, 2010

Race, gender and age effects on the assessment of bipolar disorder in youth

Mary Ann McDonnell
Northeastern University

Recommended Citation

McDonnell, Mary Ann, "Race, gender and age effects on the assessment of bipolar disorder in youth" (2010). *Nursing Dissertations*. Paper 1. <http://hdl.handle.net/2047/d20000351>

This work is available open access, hosted by Northeastern University.

Race, Gender and Age Effects on the Assessment of Bipolar Disorder in Youth

A Dissertation Presented

By

Mary Ann McDonnell

To

The School of Nursing in Bouve College

in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

In Nursing

Northeastern University

Boston, Massachusetts

August 2010

APPROVALS

**Northeastern University
Bouve College of Health Sciences
School of Nursing**

**Ph.D. Program in Nursing
Approval of Dissertation Proposal**

To: Betsy Howard, PhD
Director, PhD Program in Nursing

From: Carol Glod, PhD
Major Advisor Date

Mary Ann McDonnell
Student's Name SS#

PhD in Nursing 9-07
Area of Concentration Date Admitted

The Doctoral Committee of the above named doctoral student has approved a dissertation proposal entitled:

A copy of this proposal is attached.

Approvals: Signatures below indicate approval of the attached dissertation proposal and recommendation to candidacy for the Ph.D. Degree.

Carol Glod, PhD
Major Advisor Date

Eric Youngstrom, PhD
Committee Member Date

Alisa K. Lincoln, PhD
Committee Member Date

Betsy Howard, PhD
Director, Graduate Programs Date

Carol Kenner, PhD
Dean, School of Nursing Date

ACKNOWLEDGEMENTS

I cannot express enough appreciation and gratitude to my friend, advisor and committee chair, Professor Carol Glod, for her unwavering support, encouragement and academic guidance. Professor Glod has played a major role in my academic pursuits, mentoring and guiding me over the past 13 years. Her influence and example led me to pursue a PhD in Nursing and her persuasion and guidance kept me on track throughout the years.

I will always be indebted to Dr. Eric Youngstrom for his encouragement and unbelievable generosity. He has been a brilliant mentor and friend who patiently guided me through the beginning stages of the research proposal and the data analyses to the interpretation of the results of this study. Without his inspiration, assistance and patience, the evolution and completion of this study would not have been possible. I am also truly thankful to all of the members of Dr. Youngstrom's research team at UNC who were very welcoming and helpful throughout this process. I would like to especially thank Andrew Freeman, who was very generous with his time and shared his statistical knowledge and problem solving skills when I was in need of guidance and support.

Dr. Alisa Lincoln brought a unique perspective to this research project and I thank her for her guidance and support and for taking the time to be on my dissertation committee. I am very thankful to Professor Betsy Howard who played a major role in my academic endeavors and helped me to navigate my way through the trials and tribulations of being the first graduate of the PhD in Nursing program at Northeastern University.

I would like to thank my many friends, professors and mentors from Northeastern University who helped me along the way, especially Yevgenia LaVrova, Ruth Balal, and Janet Nescarella for their tremendous support and guidance. You have all contributed to making my life at Northeastern University rich and unforgettable.

Last, but not least, I thank my family, especially my parents, brothers and my children for unconditionally supporting my endeavors, no matter how grandiose they may be at times, and for always being there when I needed them. Your encouragement and support have always been a driving force in my life.

CONTENTS

APPROVALS.....1

ACKNOWLEDGMENTS.....2-3

CONTENTS.....4

LIST OF TABLES.....5

LIST OF FIGURES.....6

ABSTRACT.....7

RACE, GENDER AND AGE EFFECTS ON THE ASSESSMENT OF BIPOLAR
DISORDER IN YOUTH

Introduction: Chapter 1.....8-15

Review of the Literature and Theoretical Framework: Chapter 2.....16-45

Methods: Chapter 3.....46-64

Results: Chapter 465-89

Discussion: Chapter 5.....90-104

CONCLUSION.....105-112

REFERENCES.....113-142

APPENDIXES.....143-169

LIST OF TABLES

Table 1: Definitions for the purpose of this study.....	143-144
Table 2: DSM-IV-TR Diagnostic Criteria for Manic/Hypomanic Episode.....	144
Table 3: Demographics.....	145
Table 4: Diagnosis Stratified by Group.....	145
Table 5: Eigenvalues.....	146-147
Table 6: One Group CFA.....	148
Table 7: P-YMRS: RACE DIF.....	149
Table 8: P-GBI RACE DIF.....	150
Table 9 P-CMRS: RACE DIF.....	151-152
Table 10: P-MDQ RACE DIF.....	153
Table 11: P-YMRS AGE DIF.....	154
Table 12: P-GBI AGE DIF.....	155
Table 13: P-CMRS AGE DIF.....	156-157
Table 14: P-MDQ AGE DIF.....	158-159
Table 15: P-YMRS GENDER DIF.....	160
Table 16: P-GBI GENDER DIF.....	161
Table 17: P-CMRS GENDER DIF.....	162-163
Table 18: P-MDQ GENDER DIF.....	164-165

LIST OF FIGURES

1. GATEWAY PROVIDER MODEL.....	166
2. SAMPLE: HIGHEST LEVEL OF EDUCATION.....	167
3. SAMPLE: ESTIMATED ANNUAL INCOME.....	168
4. SAMPLE: HOLLINGSHEAD OCCUPATIONAL SCALE SCORE.....	169

ABSTRACT

Background: Little is known about racial, age or gender disparities in the assessment and diagnosis of pediatric bipolar disorder in children and adolescents. **Objectives:** To assess the psychometric properties of five commonly-used screening instruments in detecting bipolar disorder in youth.

Design: A secondary analysis will be used to determine if a test item bias (or differential item functioning) exists within each of these instruments. Data from Dr. Eric Youngstrom's existing database of 758 African American and White families, will be used. **Results:** Measurement variance was found in the instruments between different racial and age groups and between genders. As seen in the results section, DIF was detected on specific items, indicating that the probability of a parent endorsing that item differs between equal groups due to group membership, rather than the construct (mania), which it is intended to measure. However, most differences were small and unlikely to be clinically-meaningful or influence diagnostic or treatment decisions in research or clinical practice.

Conclusions: Further research using advanced methodology is needed to better understand how these instruments function between different racial and age groups and between genders and to better understand how the differences will affect research and clinical practice. **Keywords:** pediatric bipolar disorder, assessment, diagnosis, race, age, gender, screening instruments, rating scales and disparities.

RACE, GENDER AND AGE EFFECTS ON THE ASSESSMENT OF BIPOLAR DISORDER IN YOUTH

Chapter 1

Introduction

Pediatric Bipolar Disorder (PBD) is a serious public health concern in the United States today. Studies suggest that PBD is a severe chronic illness that necessitates early, appropriate diagnosis and treatment to prevent a worsening of symptoms and an increase in psychopathology (Geller, Tillman, Craney, & Bolhofner, 2004). The diagnosis of PBD has increased dramatically in recent years (Blader & Carlson, 2007), with the incidence tripling over the course of a decade (Youngstrom, Meyers, Youngstrom, Calebrese, & Findling, 2006). Prevalence rates of PBD vary from 1.3% up to 5.7% (Lewinsohn, Klein, & Seeley, 2000). Recent findings from a meta-analysis of epidemiological studies of pediatric bipolar disorder suggest that PBD occurs in 2 to 3% of youths in community samples (Van Meter, Moreira, & Youngstrom, 2009).

Children and adolescents with PBD have a significantly increased risk of suicidal behavior and completed suicide (Pfeffer, 2001), are severely impaired and account for a large percentage of those utilizing services in the school, juvenile justice and community systems. Since 2003, PBD is the most common diagnosis for hospitalized children with psychiatric conditions by 2003 (Blader & Carlson, 2007). Under-detection, misdiagnosis and inappropriate treatment of PBD

continue to be serious problems in the U.S. population, with few studies addressing the effects of race, age or gender in assessment and diagnosis.

Children and adolescents with PBD have been overlooked for many years due to the atypical expression of symptoms in youth compared to adults, developmental differences and the high rates of comorbid psychiatric disorders. Among the many challenges that researchers and clinicians have faced for many years is the lack of appropriate, age-specific diagnostic criteria and screening instruments which can be used in children and adolescents.

Parent reports are the most efficacious in diagnosing pediatric bipolar disorder compared to self-report and teacher rating scales (Youngstrom et al., 2004). The most commonly-used screening instruments for PBD include: 1) Young Mania Rating Scale-parent version (P-YMRS). The P-YMRS is an 11-item questionnaire adapted from the YMRS (Young, Biggs, Ziegler, & Meyer, 1978). Parents rate their child's manic symptoms on five explicitly defined grades of severity, with item scores ranging from 0 to 4 (and 3 items ranging from 0 to 8); 2) General Behavior Inventory-parent version (P-GBI). The GBI was originally developed for adults with bipolar disorder and later modified as a parent-report inventory (P-GBI) for youth as young as five years of age (Findling et al., 2001; Youngstrom, Findling, Danielson, & Calabrese, 2001). Parents complete the P-GBI to rate the depressive, hypomanic, manic, and biphasic mood symptoms of their children aged 5-17 (Findling, et al., 2001; Youngstrom, et al., 2001); 3) Child Mania Rating Scale-parent version (P-CMRS) is a 21-item scale. Each item is developmentally specific for children and adolescents and corresponds with the

DSM-IV-TR diagnostic criteria symptoms (American Psychiatric Association, 2000); 4) Mood Disorders Questionnaire - parent version (P-MDQ) is a youth scale (Wagner, Emslie, Findling, Gracious, & Reed, 2004) which is a slightly modified version of the MDQ. Parents complete the 15 item questionnaire to report potential manic symptoms in their child; 5) Child Behavior Checklist (CBCL) is a rating scale that is filled out by parents in reference to their child or adolescent. The 2001 version of the Child Behavior Checklist for ages 6-18 (CBCL/6-18) (Achenbach & Rescorla, 2001) is one of the most widely used instruments in children and adolescents in research and clinical settings (Sattler, 2002).

The DSM-IV criteria and the most commonly-used screening instruments used to assess bipolar disorder were designed for adults, not children. Only in recent years, have researchers begun to study and agree on the diagnostic presentation of symptoms in youth and to adapt the screening instruments for the purpose of assessing youth for PBD.

Pediatric bipolar disorder is also referred to as early-onset bipolar disorder and juvenile-onset bipolar disorder in the literature. These terms are used interchangeably to describe bipolar disorder I and II, as defined in the DSM-IV-TR (APA, 2000) and bipolar disorder NOS or Bipolar Spectrum disorder, which are used as a working diagnosis of children and adolescents who are impaired with symptoms of bipolar disorder but who do not meet the full diagnostic criteria as defined in the DSM-IV-TR (APA, 2000).

The *Diagnostic and Statistical Manual, Fourth Edition (DSM-IV-TR)* (APA, 2000) defines Bipolar I disorder characterized by one or more manic episodes, (a

distinct period with an abnormally and persistently elevated, expansive, or irritable mood, for at least one week); or mixed episodes, (in which criteria for a manic episode and major depressive episode are met nearly every day for at least one week); and Bipolar II Disorder as having at least one or more major depressive episodes, with at least one hypomanic episode (a distinct period of persistently elevated, expansive or irritable mood lasting at least 4 days, different from the usual non-depressed mood). The presence of manic or mixed episode precludes the diagnosis of Bipolar II Disorder (APA, 2000). Cyclothymic disorder is a chronic, fluctuating mood disturbance, involving numerous periods of hypomanic symptoms, alternating or mixed with depressive symptoms. To meet diagnostic criteria for cyclothymia, youth must not experience any symptom free intervals lasting longer than 2 months during a one-year period. Bipolar Disorder (NOS) not otherwise specified or the term Bipolar Spectrum Disorder is used when some, but not all, symptoms are present or when duration or severity criteria fail to meet criteria for Bipolar I or II (APA, 2000). Although individuals may not meet full diagnostic criteria for Bipolar I or II disorder, they suffer from high levels of impairment that require intervention. Interventions which begin early in the illness may delay the first onset of mania or even prevent the development and progression to full bipolar disorder (Chang, Howe, Gallelli, & Miklowitz, 2006). Whether or not children and adolescents with Bipolar Disorder NOS will go on to develop Bipolar Disorder I or II is unclear.

It remains unclear if the bipolar symptoms and scores on screening instruments perform similarly among different racial and age groups and between genders. It is paramount for clinicians to use unbiased screening tools for the

diagnosis of PBD, given the increasing number of minorities living in the U.S. and research findings indicating that PBD left untreated, can follow a progressive course, with recurrent episodes that become increasingly severe and resistant to treatment (Blader & Carlson, 2007; Geller & Luby, 1997).

A review of the literature revealed no published data on how the most commonly-used screening instruments perform in different racial groups, between younger children 5-10 years of age and older children 11-18 years of age, or between genders. Thus, there is virtually no data to support their use in different ethnic or racial groups. Accurate diagnosis and treatment necessitates awareness of critical differences between minority groups and others, in beliefs and sensitivities and report of mental health symptoms. Ignoring racial differences reflects a certain type of bias in the assessment of individuals (Snowden, 2003). Understanding the relationship between race, gender and age and the function of commonly-used screening instruments for PBD is essential to ensure proper and timely diagnosis and treatment for youth.

It is essential that we understand and address disparities in the assessment and diagnosis of PBD. Accurate diagnosis and timely and appropriate intervention are greatly influenced by valid measurement techniques. In order to address the well-documented problem of disparities in mental health care, we need to address the issue of the limitations in the screening and assessment of children and adolescents. While rates of bipolar diagnosis between boys and girls are equivalent, symptoms typically manifest themselves differently between genders. Boys express more externalizing symptoms and violent behavior, while

girls express more internalizing symptoms than boys, with more somatic complaints (Kashani, Beck, Hooper, et al., 1987; Michaud, Narring, Dubuis-Arber, & Paccaud, 1993). A study of AA urban teenagers found that girls were more likely to report emotional distress, a primary feeling of disadvantage, and expressions of volatile anger, interpersonal sensitivity and loneliness (Offord, Boyle, & Szatmari, 1987). Both AA and white adolescent females perceive more emotional distress than do AA and white adolescent males (Casper, Belanoff, & Offer, 1996). To improve assessment and diagnostic techniques in both boys and girls, it is essential to understand if any gender bias exists in the screening instruments. Using developmentally-appropriate measures, which are sensitive to gender may lead to more timely and accurate diagnosis for children and adolescents.

Assessing mental health issues in children and adolescents can be complicated. The presentation of symptoms and the course of BD can be quite different in children compared with adults. This is particularly true in young children age (Youngstrom, Youngstrom, & Starr, 2005). It is important to consider the age of the individual at the time of the assessment in relation to his or her developmental stage and clinical presentation. It is also important to consider the age in which they experienced an onset of symptoms. The quality of presenting symptoms and the range of comorbidity may vary greatly according to age and duration of illness (Post & Kowatch, 2006). Older children and adolescents report more depressive symptoms than younger children do. Therefore, careful consideration should be paid to the age of the individual, in relation to the total

scores of screening instruments to reduce bias in assessment (Duax, Youngstrom, Calabrese, & Findling, 2007). Discrepancies appear to be greater when data from children with prepubertal-onset bipolar disorder are combined with those with adolescent-onset bipolar disorder. Since research on screening instruments for PBD have relied on diagnostic criteria and instrumentation developed for adults, there is a need for developmental appropriateness in the item content of screening instruments among different age groups (Geller et al., 2002). The overall short-term goal of this study is to assess the psychometric properties of five commonly-used screening instruments in detecting bipolar disorder in youth. A secondary analyses will be conducted to determine if a test item bias exists within each of these instruments. Data from Dr. Eric Youngstrom's existing database (of 758 African American and White families, will be used for the analyses.

The primary specific aim of this study is to answer the following question: 1) Does a test item bias (differential item functioning) exist between race (African American and White) when using the P-YMRS, P-GBI, P-CMRS, P-MDQ or CBCL in urban youth? The secondary aim of this study is to answer the following two questions: 1) Does a test item bias (differential item functioning) exist between age (children 5-10 years of age and adolescents 11-18 years of age) when using the P-YMRS, P-GBI, P-CMRS, P-MDQ or CBCL in urban youth?; and 2) Does a test-item bias (differential item functioning) exist between gender (boys versus girls) when using the P-YMRS, P-GBI, P-CMRS, P-MDQ or CBCL in urban youth?

The overall long-term goal of this program of research is to improve identification of racial, gender and age disparities in the assessment and diagnosis

and treatment of PBD, and to prevent untoward outcomes in youth who have been undiagnosed and untreated. The findings from this study will lead to a better understanding of the bias of assessment tools. The proposed study is important for the examination of health care disparities to guide early diagnosis and treatment in youth. In addition, the reliability and validity of screening instruments will be assessed. Well-validated instruments can increase the rate of early symptom identification, facilitate accurate and timely diagnosis and ultimately address the unmet treatment needs for specific subgroups of youth with bipolar disorder.

Chapter 2

Review of the Literature and Theoretical Framework

Introduction

This chapter is a state of the science review of the literature on the assessment and diagnosis of pediatric bipolar disorder (PBD). It includes a theoretical framework which can be used to guide the study of pediatric bipolar disorder, and a review of the literature pertaining to the epidemiology, diagnosis, assessment, and the gene environment interaction of PBD, as well as the influence of race, gender and age in the assessment and diagnosis of mental health issues in children and adolescents.

Theoretical Framework

Linking a theoretical framework is the critical first step to understanding and evaluating screening tools which are commonly used to assess pediatric bipolar disorder in children and adolescents (Andersen, 1995). The Network-Episode Model (Pesconsolido, 1991, 1992), a framework which is focused on the dynamic process underlying the use of mental health services and the mode of entry into the service sector, highlights social influence as a major determinant in when, how and if an individual will receive services. It infers that individuals will either seek care as a “rational” choice, by making a conscious decision to enter into treatment, or they will lose their ability to choose and be forced into treatment by members of their family, community or the police. This framework emphasizes the influence of social, psychological, economic, cultural, medical and system factors on an individual’s direct illness career and on how he or she will respond to

health problems over a period of time and that the route in which an individual enters treatment ultimately affects his or her perception and report of mental illness (Pescosolido, Gardner, & Lubell, 1998).

The study of mental illness in youth presents unique challenges because children and adolescents do not typically seek mental health care on their own. They rely on knowledgeable family members, teachers, juvenile justice authorities and other adults in the community, referred to as “gateway providers” in the NEM, to recognize their symptoms as pathological and to refer them for treatment. However, findings from research studies consistently report low rates of identification and recognition of symptoms in youth, (Stiffman, Pescosolido, & Cabassa, 2004) with 46-86% of mental health symptoms and diagnoses being overlooked by general practitioners, pediatricians and primary health care providers (Stiffman, Pescosolido, & Cabassa, 2004). It is essential that “gateway providers” be educated about predisposing factors such as culture, economic and environmental factors that influence the presentation of symptoms in youth to improve timely recognition and treatment of psychopathology in youth.

Based on the NEM, gateway providers and social influence are significant factors in determining whether youth will receive mental health services. The children’s version of the NEM suggests that the gateway provider’s knowledge, attitude and preference may affect their response to a child’s mental health symptoms and shape the pathway in which the child will enter and travel through the mental health system. The Gateway Provider Model (GPM) shown in Figure 1 (Stiffman, Pescosolido, & Cabassa, 2004), is an elaborated testable subset of the

Network-Episode Model, focused on the central influences that can affect a child or adolescent's access to treatment. In this model, the "gateway provider" is defined as the individual who first identifies a problem and refers the child or adolescent to treatment (Stiffman, Pescosolido, & Cabassa, 2004).

The GPM focuses on 3 central influences: 1) the key role of the gateway provider (GP) in the initiation or direction of treatment for youth; 2) the importance of the GP having access to information so that he or she can offer youth consistent assistance, advice and referrals, tailored to fit their needs; and 3) recognizing that the GP's attitude, knowledge and impression of support for treatment and the system burden can inhibit or facilitate the implementation of new approaches in treatment systems.

Enhancing the education, training and support of GP's who play a significant role in the identification of symptoms and in referring youth to treatment should be a major priority. Well-validated screening instruments may lead to better recognition of symptoms and increased rates of referrals by GP's. It is essential that the instruments used by GP's be developmentally-appropriate and sensitive to race, age and gender to facilitate more timely and accurate diagnosis and treatment for children and adolescents with PBD, which may ultimately increasing youth's access to and utilization of mental health services.

Before we can truly understand how individuals, the health care system and external environments interact, we must clearly identify the needs of individuals within specific ethnic and socioeconomic groups in urban areas. The GPM goes beyond the assessment of objective need and operates independently of the child

or adolescent's perception of need, which is key in assessing children and adolescents (Stiffman, Pescosolido, & Cabassa, 2004). It has been used in previous studies to examine the role of providers, bridging the gap between the child and adolescent's need for services and the actual receipt of services.

The strengths and weaknesses of the family and community may contribute significantly to perceptions about mental illness and may delay or speed up the process of getting an evaluation and an accurate diagnosis. To develop interventions aimed at addressing accurate diagnosis and timely treatment for youth with bipolar disorder, it is essential to improve assessment tools and to increase the gateway provider's knowledge and understanding of their value. Identifying psychopathology is an essential first step in increasing the GP's ability to make quality care decisions and appropriate referrals for youth (Stiffman, Pescosolido, & Cabassa, 2004). Knowledge gained from this study may lead to future studies aimed at educating gateway providers in assessment and accurate diagnosis and treatment of African American youth with PBD.

Review of the Literature

Since 2003, pediatric bipolar disorder (PBD) has been the most common diagnosis for hospitalized children with psychiatric conditions (Blader & Carlson, 2007). Under-detection, misdiagnosis and inappropriate treatment of PBD continue to be a serious problem in the U.S. population, with few studies addressing the effects of race, age or gender in assessment and diagnosis.

An extensive review of the literature revealed little to no systematic study on racial, age or gender disparities of PBD. Prevalence rates of PBD are estimated

to be between 2-3% (Van Meter, et al., 2009). While 2-3% of the population having PBD may not seem like a large enough number to deem it a serious public health concern, the severity and chronicity of the illness and its deleterious effect on individuals, families, and the community at large warrant serious consideration and attention. Individuals with bipolar disorder who are between the ages of 15 to 44 have high rates of death and disability (Woods, 2000). It is estimated that by the age of 25, females with PBD who have gone untreated can expect to lose 14 years of effective functioning in work and school and with family, and 12 years of normal health, in addition to a life expectancy shortened by nine years, compared to controls (DHEW, 1979). PBD may carry an even greater risk of mortality from suicide (Geller et al., 2002). The lifetime rates of suicide attempts in individuals with bipolar disorder is 29.2% and an estimated 8.6-18.9% actually die by suicide (Chen & Dilsaver, 1996). Individuals with bipolar disorder (BD) are at the highest risk of suicide attempts, with nearly one third having attempted in early years (Carter, 2003). They are also at an increased risk for substance use, abuse, and dependence (Tohen, Zarate, & Turvey, 1995), alcohol abuse, obsessive compulsive disorder, school failure, and legal problems (Lin et al., 2006; Wilens et al., 1999).

Epidemiology

Prevalence rates of PBD vary considerably according to different methodology used, populations studied, and differences in the inclusion/exclusion criteria used. Some studies included only bipolar I or II disorder, while others included both bipolar I and II disorder, as well as bipolar disorder NOS and bipolar

spectrum disorders. A study of youth with bipolar disorder revealed lifetime prevalence rates of 1.3%, with equal rates between genders (Kessler, 1994). The National Co-morbidity Study conducted in 1990 included 8,000 subjects, 15-24 years of age, reported a lifetime prevalence of bipolar I disorder as 1.7%, with a median age of onset of 21 years of age (Kessler, 1994).

Bipolar Disorder (BD) in the adult population has a typical onset occurring in late adolescence or young adulthood, affecting roughly 4% of the population according to the most recent epidemiological data (Kessler, Berglund, Demler, Jin, & Walters, 2005). Previous adult studies of bipolar disorder have reported rates as high as 6% when subthreshold or bipolar spectrum cases were included in the study (Judd & Akiskal, 2003). Lifetime prevalence rates of bipolar I range from 0.3 to 1.6%, with approximately 0.5% having bipolar II (Lewinsohn, et al., 2000). The prevalence rates of bipolar I and II combined were reported as 2.6% in the National Comorbidity Survey Replication (Kessler, Chiu, Demler, Merikangas, & Walters, 2005).

Retrospective studies of adults have further supported the existence of PBD. The National Institute of Mental Health STEP-BD study reported that of the first consecutive 1000 adult bipolar patients who entered the study, 27.7% had a prepubertal onset of symptoms (under the age of 13 years of age), and 37.6% experience adolescent onset, defined as between 13 to 18 years of age (Perlis et al., 2004). An epidemiological study of the lifetime prevalence of bipolar disorders I and II found onset rates of 1% during adolescence and 2% during young adulthood, and of 5% for bipolar spectrum disorder (Lewinsohn, et al., 2000).

Previously, a lifetime prevalence of 1% for bipolar II disorder and 1% for cyclothymia in youth 14 to 18 years of age was reported in an epidemiological study in a community sample of older adolescents (Lewinsohn, Klein, & Seeley, 1995). Bipolar spectrum disorder and bipolar disorder not otherwise specified (NOS), subthreshold forms of the disorder occur more frequently, and are reported as affecting up to 5.7% of children and adolescents in some research studies (Lewinsohn, et al., 2000).

More recent findings from a meta-analysis suggest that PBD occurs in 2 to 3% of youth in community samples (Van Meter, et al., 2009). However, clinical diagnosis of PBD has been reported as high as 5.9% in patients from an urban youth mental health clinic in Ohio (Youngstrom et al., 2005) and 3.5 to 6% in young adult samples (Grant et al., 2005; Kessler, Berglund, Demler, Jin, & Walters, 2005). Furthermore, the US Center for Disease Control data for 2002-2003, reported rates as high as 40% in acute psychiatric hospitalizations (Blader & Carlson, 2007). This is not surprising, given the severity of impairment found in children and adolescents who have been diagnosed with bipolar disorder at a young age. The large discrepancy between rates of bipolar disorder in large-scale community surveys and prospective longitudinal studies (Angst, 2004; Bauer & Pfenning, 2005; Grant et al., 2005; Pini et al., 2005; Waraich, Goldner, Somers, & Hsu, 2004), may be accounted for by differences in the samples, methodology, and screening instruments.

Heredity/Environment

The single best predictive risk factor for bipolar disorder is high familial loading. Family and adoptive studies have further supported heredity as a significant predictor of the development of bipolar disorder (Smoller & Finn, 2003). Children who have a biological parent with bipolar disorder average a five-fold increase in the likelihood of having bipolar disorder (Youngstrom & Duax, 2005), and a 2.5 fold increase in the risk if they have a second degree relative, such as a grandparent with bipolar disorder (Hodgins, Faucher, Zarac, & Ellenbogen, 2002). Genetic studies of adult-onset bipolar disorder show clear familiarity (Althoff, Rettew, Faraone, Boomsma, & Hudziak, 2006), including a 10-fold increased risk of bipolar disorder in first degree relatives of bipolar probands compared to control families (Smoller & Finn, 2003). Rates in offspring of bipolar parents range from 4 to 15% versus from 0 to 2% in offspring of healthy parents (Birmaher et al., 2009). In a study of 388 offspring of 233 parents with bipolar disorder and 251 offspring of 143 demographically matched control parents, greater than 75% of offspring of parents with bipolar disorder reportedly had their first mood episode prior to the age of 12 and most met the criteria for bipolar disorder NOS or major depressive disorder during their first episode (Birmaher, et al., 2009). Offspring of parents with BP are at high risk for the development of bipolar spectrum disorders, mood, anxiety and other Axis I disorders (Birmaher, et al., 2009). While many factors have been linked to the development and onset of bipolar disorder, other than family history, there is no other risk factor that has been documented sufficiently to

justify its integration into clinical decision making regarding the diagnosis of pediatric bipolar disorder (Tsuchiya, Byrne, & Mortensen, 2003).

Findings from research studies suggest that if two individuals with the same genetic predisposition for bipolar disorder are exposed to an urban versus nonurban environment, each will fare differently. Those in urban settings have a higher likelihood of developing psychotic illness versus nonurban individuals who may develop mania (Kaymaz et al., 2006). Despite documentation of increased stressors in urban youth and the interaction between psychosocial stressors and the onset of bipolar disorder, there is no data to suggest that there is an increase in the diagnosis of bipolar disorder in urban youth.

Specific Risk Factors for Urban Youth

Environmental Influence. Children and adolescents living in urban environments are exposed to more psychosocial stressors, which has been linked to an increased risk of developing bipolar disorder (Post, Leverich, Xing, & Weiss, 2001). In addition to environmental factors, there is a high rate of misdiagnosis in minority groups (Good, 1997). While the impact of environment on the development of PBD is unclear, evidence suggests that environment may increase the risk of psychopathology in a child who has a genetic predisposition (Helm, Newport, Bonsall, Miller, & Nemeroff, 2001). Stress-induced alteration in prepubertal adolescents may increase the risk of vulnerability to the onset of psychopathology during adolescence (Romeo & McEwen, 2006).

Compared to nonurban areas, urban areas have higher rates of poverty, school drop out, crowded living conditions and lower income levels. Youth who

live in these areas are exposed to more crime, violence, high level stress and substance abuse (Weist et al., 2000) and are more likely to be Hispanic or African American, come from families who speak English as a second language and to live in a single parent household. Psychosocial stressors may interact with genetic predisposition to induce the full expression of bipolar disorder (Post, et al., 2001). Understanding disparities that exist in the assessment and diagnosis of urban individuals is essential to improve accurate and timely diagnosis of bipolar disorder in this population.

Importance of Accurate Assessment of Pediatric Bipolar Disorder

Kindling. Early intervention may delay the development of bipolar disorder: the onset of the first manic episode and may even prevent the development of full bipolar disorder (Chang, et al., 2006). The theory of kindling in mood suggests that the combination of psychosocial stress and genetic vulnerability may lead to a full mood episode and that once a full mood episode has occurred, future episodes are triggered more easily with a mild stimulus. Over time, episodes begin to develop spontaneously with no stimulus at all. Interventions during the early course of kindling may prevent progression of the illness (Chang, Steiner, & Ketter, 2000). To intervene early in the course of illness, an accurate and timely diagnosis must be made. Early intervention may ultimately help to prevent the high psychosocial and medical morbidity and mortality associated with bipolar disorder (Goodwin & Jamison, 2007).

Diagnosis of Pediatric Bipolar Disorder

DSM-IV-TR definition. The *Diagnostic and Statistical Manual, Fourth Edition (DSM-IV-TR)* (APA, 2000) does not distinguish age specific criteria for bipolar disorder, which presents a challenge to identifying and diagnosing children and adolescents who have an atypical symptom presentation. The DSM-IV TR (APA, 2000) defines Bipolar I disorder characterized by one or more manic episodes, (a distinct period with an abnormally and persistently elevated, expansive, or irritable mood, for at least one week); or mixed episodes, (in which criteria for a manic episode and major depressive episode are met nearly every day for at least one week); and Bipolar II Disorder as having at least one or more major depressive episodes, with at least one hypomanic episode (a distinct period of persistently elevated, expansive or irritable mood lasting at least 4 days, different from the usual non-depressed mood). The presence of manic or mixed episode precludes the diagnosis of Bipolar II Disorder.

Cyclothymic disorder is a chronic, fluctuating mood disturbance, involving numerous periods of hypomanic symptoms, alternating or mixed with depressive symptoms. To meet diagnostic criteria for cyclothymia, youth must not experience any symptom free intervals lasting longer than 2 months during a one-year period.

Developmental Differences in the Diagnosis of Pediatric Bipolar Disorder

For many years, it was believed that bipolar disorder only existed in adults. However, there is now consensus among experts that unequivocal cases of bipolar disorder occur in prepubertal children, as well as adolescents (Leibenluft, 2008), and evidence suggests that characterization of early phases in the

development of PBD is feasible (Bauer, Juckel, Correll, Leopold, & Pfennig, 2008). Children and adolescents with bipolar disorder have an atypical presentation of symptoms compared to that of adults. Currently, there is no consensus statement on the “true” diagnostic criteria for pediatric bipolar disorder (PBD) (Youngstrom, Birmaher, & Findling, 2008) which has contributed to under-diagnosis and misdiagnosis in children for many years.

PBD is one of many terms used to describe children and adolescents with bipolar disorder. Other terms which are used interchangeably include: early-onset bipolar disorder, very-early onset bipolar disorder, juvenile bipolar disorder, pediatric mania, and prepubertal and pubertal bipolar disorder (Carlson et al., 2009). It is unclear if these terms, while used interchangeably, actually refer to the exact same condition. The lack of continuity in both defining and naming bipolar disorder in youth has contributed to diagnostic confusion (Carlson, et al., 2009).

There are many factors which currently impede accurate and timely diagnosis including (1) the lack of child-specific diagnostic criteria as defined by the DSM-IV-TR, (2) the variable clinical presentation within and across episodes, (3) its symptomatic overlap with more common disorders of childhood, (4) the constraints placed on symptom expression by the developmental stage of a child, (5) issues of comorbidity and (6) the lack of child-specific screening instruments which have established validity in different racial and age groups or between genders.

The primary barrier to advancing research in PBD is the lack of consensus on the definition and diagnosis in children and adolescents (Carlson, et al., 2009).

As researchers and clinicians struggle to fit children and adolescents into the DSM-IV diagnostic criteria for adult bipolar disorder, the diagnosis has been missed in many youth, prompting the development of age-specific criteria (Geller & Luby, 1997). To establish core features of PBD, researchers use a developmental, age-specific viewpoint for pediatric patients who do not have the adult-type onset of BD. Despite differences of opinions between researchers, the core features of bipolar disorder in prepubertal and early-adolescent bipolar disorder phenotype (included in table 2) as defined by Geller et al (2002) are currently considered the gold standard in diagnosing youth.

In an effort to move beyond the diagnostic tangles of PBD impeding the advancement of research, the National Institute of Mental Health (NIMH) funded initiatives bringing experts from around the world together. As a result, many reviews were conducted which support the validity of PBD in youth (Faedda et al., 1995; Leff, Fischer, & Bertelsen, 1976; Weller, Weller, & Fristad, 1995). This expert group conceded that the diagnosis of PBD is possible using psychiatric assessment instruments and that while some phenotypes of bipolar I and II fit the DSM-IV criteria, other phenotypes of the disorder exist which do not fit the criteria (Nottelmann, 2001).

For the non-DSM-IV phenotype of PBD, a working diagnosis of bipolar disorder not otherwise specified (NOS) was agreed upon. Subthreshold forms of bipolar disorder are subsumed under the category of “bipolar spectrum disorders.” It is important to note that subthreshold disorders, while not meeting all of the specific diagnostic criteria as defined in the DSM-IV-TR, are still associated with

significant impairment and early recognition may provide a window of opportunity for early intervention and the prevention of full blown bipolar disorders (Shankman, Klein, Lewinsohn, Seeley, & Small, 2008). Researchers differ in their opinions about whether the term “bipolar spectrum disorders” accurately captures the seriously impaired children who have severe aggression, problems with hyperactivity, impulsivity, inattention and mood instability (Carlson, et al., 2009). Findings from research studies report that bipolar spectrum disorders or bipolar disorder NOS (BP-NOS), are common, clinically significant and under-detected in treatment settings. This category is recommended to describe youth who do not have the classic adult presentation of bipolar disorder (Nottelmann, 2001). This diagnostic term is typically used when some, but not all, symptoms of bipolar disorder are present or when duration or severity criteria fail to meet criteria for Bipolar I or II.

The most frequent course of PBD is characterized by continuous psychopathology with few well periods, including long duration of episodes with rapid cycling and mixed mania (Geller et al., 2000a; Nottelmann et al., 2001). Rapid, brief, recurrent episodes lasting hours to days are a more typical pattern of illness in children and adolescents (Geller, Tillman, et al., 2004; Geller, et al., 2000a). Other researchers report that youth with chronic mania present with severe impairment with mood disturbances that are characterized by irritability and aggressiveness (Biederman et al., 2000; Carlson, Bromer, Driessens, Mojtabei, & Schwartz, 2002; Geller, Craney et al., 2002; Geller, Zimmerman, et al., 2002; Pavuluri, Birmaher, & Naylor, 2005; Wozniak et al., 1995), and that an earlier age

of onset is associated with a higher risk of recurrence of a mood episode and a worse outcome (Yatham, Kauer-Sant'Anna, Bond, Lam, & Torres, 2009).

The DSM-IV-TR definitions of bipolar disorder are a common topic of discussion in the pediatric literature. Mixed episodes are defined as a period of at least 7 days in which symptoms of a manic and depressive episode are met. Rapid cycling is defined as at least 4 episodes per year, with the prerequisite criteria for at least 7 days being met for a manic episode. This is different than the rapid cycling discussed in the pediatric literature, in which, cycling refers to mood changes within an episode (Geller, Tilman, Craney, & Bolhofner, 2004).

Subthreshold hypomania was reported to be predictive of bipolar II and an increased risk for suicidality in the Oregon study (Angst, Gamma, & Lewinsohn, 2002). Despite the fact that subthreshold cases of PBD do not meet the full criteria for bipolar disorder I (BP-I) or bipolar disorder II (BP-II), most of these children and adolescents have moderate to severe clinical severity and role impairment (Merikangas et al., 2007) that require intervention. The earlier treatment is initiated, the more responsive children seem to be to medications and other interventions. Whether or not children and adolescents with Bipolar Disorder NOS will go on to develop Bipolar Disorder I or II is unclear.

Presentation of Symptoms.

Studies of populations at high-risk for BD development have indicated that children with strong family histories of BD, who are themselves experiencing symptoms of attention-deficit/hyperactive disorder (ADHD) and/or depression or have early mood dysregulation, may be experiencing prodromal states of BD. Many

children and adolescents with bipolar disorder present with the following symptoms: 1) chronic irritability; 2) psychosis; 3) mixed episodes with rapid cycling, with little inter-episodic recovery; and 4) high rates of comorbidity with ADHD, oppositional defiant disorder and conduct disorder. Episodes tend to be long in duration, often with rapid cycling and mixed mania, and include symptoms of impulse control or conduct problems (Birmaher et al., 2006; Geller, et al., 2000a; Geller, 2005; Wozniak, et al., 1995). Mixed episodes are described as rapid shifting moods, ranging from irritability or elation to depression; usually with manic symptoms coexisting with depressive symptoms for at least 4 hours per day.

Definitions used in the PBD literature, but not in the DSM-IV-TR include: 1) ultrarapid cycling, described as brief, frequent manic episodes lasting hours to days, not meeting the 4 day prerequisite for hypomania, and occurring from 5 to 364 times per year (Geller, et al., 2000a) and 2) Ultradian cycling, defined as repeated, brief cycles that occur daily and last from minutes to hours (Geller, et al., 2000a).

Children and adults with an atypical presentation of symptoms, characterized by periods of elated, expansive or irritable mood are increasingly being diagnosed with bipolar disorder (Carlson, 2005). Patterns of bipolar disorder in youth have been described as very rapid, brief, recurrent episodes lasting hours to a few days (Geller, Craney, et al., 2002; Geller, Tillman, et al., 2004; Geller, Zimmerman, et al., 2002).

There is increasing data on the greater severity of PBD from both prospective studies of children and retrospective studies of adult subjects about their childhood. Prospective follow up studies of children and adolescents with PBD have revealed that an earlier age of onset and low maternal warmth predicts more weeks spent ill

(Geller, Tillman, Bolhofner, & Zimmerman, 2008). In a longitudinal study investigating the continuity of child and adult BP I and characteristics of later episodes, 115 children, with a mean age of 11.1 were followed for 8 years. In this study, Geller and colleagues (2008) found a similar presentation of symptoms during first, second and third episodes, characterized by psychosis, rapid cycling, and long duration of illness in PBD. They also found that the duration of episodes lasted for long periods, with a mean of 55.2 weeks for the second episode and 40 weeks for the third episode. The mean number of cycles occurring per day in individuals with PBD is 3.7, with most of the day being spent in a pathological mood state. Daily cycles continue into adulthood, with a 44% frequency of manic episodes, which is 13-44 times higher than the population prevalence, strongly supporting the continuity between child and adult bipolar I disorder (Geller, Tillman, Bolhofner, & Zimmerman, 2008).

Psychotic symptoms appear to be prevalent in PBD. A study of two groups of youth, ages 6-9 years old and 10-16 years old, with bipolar disorder I reported the lifetime prevalence of delusions and pathological hallucinations as 76.3% (Tillman et al., 2008). The most common type of delusions reported were grandiose and the most common type of hallucinations were visual. Pathological hallucinations occurred in 37.4%, and benign hallucination occurred in 43.6% of youth with bipolar disorder I. The most common type of benign hallucinations consisted of one hearing his or her name being called. There were no differences found between the two age groups in this study. This data supports the prevalence of psychosis in outpatient prepubertal and early adolescent bipolar disorder I which is similar to that reported for adults (Tillman, et al., 2008).

Significance of Race, Age and Gender

Accurate diagnosis and timely and appropriate intervention are greatly influenced by valid measurement techniques. The evidence of bias in diagnosis of African American individuals has been documented for two decades. Bias may influence a clinician's clinical judgment and evaluation of an individual's personal or clinical characteristics (Van Ryn, 2002; Van Ryn & Burke, 2004). Studies have documented rater bias in coding more psychotic symptoms and missing more affective symptoms in minority groups (Sherazi, McKeon, McDonough, Daly, & Kennedy, 2006). High rates of comorbidity, a lack of knowledge among clinicians and biased child-specific screening instruments for youth of different racial groups, age groups and genders, are some of the factors which may contribute to under-diagnosis and misdiagnosis in youth.

Disparities in mental health assessment and diagnosis may be attributable to language barriers, a cultural influence which affects the expression of symptoms in youth, their beliefs about mental illness, and utilization of mental health services (Rios-Ellis, 2005). Stigma is a major deterrent to seeking mental health treatment. Negative perceptions of the mentally ill appear to be greater among African Americans than among whites (Pescosolido, Monahan, Link, Stueve, & Kikuzawa, 1999). Individuals who are concerned with potential stigmatization or family disapproval report less willingness to enter into treatment (Barney, Griffiths, Jorm, & Christensen, 2006). Beliefs endorsed by racial minority groups such as assigning personal responsibility to the development of illness may influence an individual's report and expression of mental health symptoms. Children are more

likely to convey psychological distress in a manner consistent with their family's customs (Lopez & Guarnaccia, 2000). Consistent with the Gateway Provider Model, research suggests that family, peers and community are major influences in the utilization of mental health services. The major determinants in whether an individual will enter into and continue with mental health treatment include perceived need and referrals from mental health and primary care providers, family members, friends and other members of the community (Ayalon & Alvidrez, 2007). As the U.S. African American (AA) population continues to grow, careful examination of the validity of standardized mental health screening tools for use with this population is essential to decrease disparities in identifying, diagnosing and treating AA youth.

Race

There are multiple ways in which racism can affect health (Williams & Mohammed, 2009) and ethnicity is significantly associated with misdiagnosis of bipolar patients (Brodie, 2004). Although the rate of BD is similar among African Americans and other Americans, African Americans (17.3%) are less likely to receive a diagnosis of a mood disorder than Whites (50.7%):(Barnes, 2008). A study of 2991 child and adolescent African-American, Hispanic/Latino and white patients, treated in an urban psychiatric emergency service over 13 months examined the role of race and disposition decision making. The findings demonstrated that African American youth are more likely to receive a diagnosis of psychosis and/or a behavioral disorder, and to be hospitalized and less likely to receive a diagnosis of bipolar disorder, major depression disorder and/or a

substance abuse disorder than white youth are (Muroff, Edelson, Joe, & Ford, 2008). Findings from a study of youth, ages 5-17 years of age, reported that African American (AA) youth were less likely to receive the diagnosis of PBD than Euro-American youth of the same age, despite the fact that there were no differences found in the manic symptom presentation and overall depression scores between the two groups (Constant et al., 2007). African American (AA) adolescents, 12 to 18 years of age, assessed during their first psychiatric hospitalization were diagnosed more frequently as having psychotic features and higher ratings for auditory hallucinations (Patel, Delbello, & Strakowski, 2006).

Significant race differences remain even when the DSM-IV diagnostic criteria and semi-structured interviews are used in the diagnosis of bipolar disorder (BD) (Neighbors, Trierweiler, Ford, & Muroff, 2003). One study of youth with PBD reported that African American adolescents were nearly twice as likely to receive treatment with an antipsychotic agent as white adolescents were (DelBello, Soutullo, & Strakowski, 2000). AA individuals have higher rates of diagnosis of schizophrenia and lower rates of affective disorders than expected (Baker & Bell, 1999). Given the under-diagnosis of African American youth with PBD and the increased rates of treatment with atypical antipsychotic agents, which have significant, and perhaps unknown long term side effects, it is essential that these disparities be addressed.

The evidence of bias in diagnosis of AA individuals has been documented for 2 decades. Bias may influence a clinician's clinical judgment and evaluation of an individual's personal or clinical characteristics (Van Ryn, 2002; Van Ryn &

Burke, 2004). It can result from many factors including unfounded assumptions about a particular group, an intolerance of minority individuals with mental illness, and an expectation that behavior should conform to specific social norms.

Misconceptions, inaccuracies and stereotypes associated with African Americans can contribute to misdiagnosis (Williams & Williams-Morris, 2000). Determining the role of bias in mental health assessment is essential (Snowden, 2003).

Understanding the relationship between race and the function of commonly-used screening instruments for PBD may reduce the excess morbidity, mortality, and economic burden due to health disparities and incremental validity of mania instruments in the AA population.

Research concerning the diagnosis of PBD in African American (AA) population has been meager, despite the fact that racial disparities in the diagnosis exist in adolescence through adulthood (Strakowski et al., 1996; Strakowski, McElroy, Keck, & West, 1996). There is a large body of research and theoretical work indicating that culture shapes the expression, perception and reporting of mental illness symptoms (Lopez & Guarnaccia, 2000), yet there have been no studies addressing how African American (AA) parents interpret and describe their child's mood, as reflected in their responses on questionnaires or to the questions that clinicians ask when using screening tools for PBD. Given that the standards for educational and psychological testing caution against using measures that have not been validated within a particular group (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 1999), it is highly problematic to

use untested instruments in assessing PBD in the AA population.

Currently, there is no research that has examined bias in screening instruments used for PBD among different racial groups, or between different age groups or genders. Failure to recognize symptoms of BD in children and adolescents of different racial and age groups and between genders may reduce the precision of estimates and lead to a bias in the assessment, diagnosis and treatment of this serious, chronic illness in youth.

In 2001, the surgeon general's report, "Mental Health: Race, Culture, and Ethnicity, and Mental Health" documented disparities in access to care and treatment of mental health issues in minority groups access who go untreated or improperly treated (General, 2001; SurgeonGeneral, 2001). Some researchers have proposed the idea that while AA and white adults do not vary in symptom presentation of psychiatric disorders, the bias of providers makes AA adults more vulnerable to diagnosis of schizophrenia or other psychotic disorder diagnoses (Neighbors, Jackson, Campbell, et al., 1989).

In a study of PBD comparing 17 African American, mean age of 15.9 years and 61 white adolescents, mean age 15.3 years, AA adolescents were how much more likely to present with or be identified with psychotic features than white adolescents were (Patel, et al., 2006). Findings from a study of youth, ages 5-17 years of age, reported that AA youth were much less likely to receive the diagnosis of PBD than Euro-American youth of the same age, despite the fact that there were no differences found in the manic symptom presentation and overall depression scores between the two groups (Constant, et al., 2007). Significant

race differences remain even when the DSM-IV diagnostic criteria and semi-structured interviews are used in the diagnosis of Bipolar Disorder (BD) (Neighbors, et al., 2003). Furthermore, AA adolescents with PBD are nearly twice as likely to receive treatment with an antipsychotic agent as white adolescents are (DelBello, et al., 2000). The reasons for these differences remain unclear.

Differences in the utilization of mental health services has also been reported, with younger and older age groups of African Americans being less likely to utilize services, and AA women being more likely to utilize services than AA men (Neighbors et al., 2007). The root of the diagnostic and treatment bias, as well as inequality in service utilization in the AA population remains unclear. These findings highlight and support the urgent need to study bias and disparities in the AA population. It is essential to uncover this discrepancy immediately and move towards the resolution of these racial effects in the assessment and diagnosis of PBD.

Discrepancy in Diagnosis. Demographics, social structure, health beliefs, personal, family and community resources, insurance, income level, access to health care sites, transportation and the individual's perceived and actual needs all affect diagnosis and treatment of illness (Andersen, 1995; Andersen & Aday, 1978).

With increasing rates of diagnosis and the continuing problem of under-detection and misdiagnosis in youth, the need for empirically-derived and clinically useful assessment strategies and instruments is more crucial than ever. Effective, reliable screening instruments will help to facilitate screening for PBD, and to

increase early recognition, diagnosis and treatment. In turn, this may ultimately lead to the prevention of the serious risks, such as hospitalization and suicide, which are associated with a delay in, or lack of diagnosis and treatment of pediatric bipolar disorder. Understanding the relationship between race and the function of commonly-used screening instruments for PBD may reduce the excess morbidity, mortality, and economic burden due to health disparities and incremental validity of mania instruments in the African American population. In order to address the well-documented problem of disparities in mental health care, we need to assess the presence of bias in the most commonly used screening instruments in PBD research studies.

Age and Gender. Earlier age of onset is associated with more exposure to risk factors such as increased rates of familial mood disorders, comorbidity and physical and sexual abuse (Youngstrom, Meyers, Youngstrom, Calebrese, et al., 2006). Retrospective studies of adults with childhood onset bipolar disorder often report that the initial episode was depressive (Lish, Dine-Keenan, Whybrow, Price, & Hirschfeld, 1994). However, those who have experienced subtle hypomanic symptoms in the past often fail to report the symptoms and clinicians frequently fail to inquire about such symptoms when an individual presents with depressive symptoms (Angst & Cassano, 2005). Misdiagnosis has important prognostic and treatment implications. Youth who present with depression early are typically diagnosed with unipolar depression and treated with antidepressants, which can contribute to a worsening of symptoms.

To investigate the rates and predictors of bipolar disorder I and II in children

6-12 years old, a prospectively, blindly rated study of 79 children with MDD and 31 normal control children matched for age, gender, and socioeconomic status was conducted. Twenty five (31.7%) of subjects who were followed over 5 years spontaneously switched from prepubertal depression to prepubertal mania (Geller, Fox, & Clarke, 1994). A study of 60 adolescents with major depressive disorder, who were treated with a tricyclic antidepressant during hospitalization and followed over 3-4 years reported a 20% switch rate from depression to bipolar disorder (Strober & Carlson, 1982). The base rate of pharmacological hypomania during treatment with an antidepressant was 4%, and predicted a bipolar outcome with 100% confidence in adults of those who switched from depression to mania (Strober & Carlson, 1982). The rates of switching from depression to mania in many research studies may be conservative because of the probable under-diagnosis of childhood mania previously discussed.

Premorbid anxiety and dysphoria are common in those whose first episode is a depressive disorder (McClellan, Kowatch, Findling, et.al., 2007). Many children with an appropriate diagnosis of unipolar depression in childhood will go on to develop bipolar disorder in adolescence or young adulthood (Birmaher, Arbelaez, & Brent, 2002). Up to 20% of youth with major depression will go on to develop mania in adulthood (Geller, Fox, & Clark, 1994; Geller, Zimmerman, Williams, Bolhofner, & Craney, 2001; Kovacs, 1996; Rao, Ryan, Birmaher, et al., 1995). Developmental factors, and the progression and course of the illness, appear to differ from the “classic” presentation of adults.

Youth have less symptom-free periods, briefer episodes and higher rates of

irritability (Findling, et al., 2001; Geller, Zimmerman, Williams, Delbello, Bolhofner, et al., 2002) and comorbidity, (Findling, et al., 2001; Geller, et al., 2000a; Wozniak, Biederman, Monuteaux, Richards, & Faraone, 2002) which complicates clinical assessment and diagnosis. Symptoms of mania and hypomania can differ between adults and youth. Adolescents with mania often present with psychotic symptoms, labile moods, and or mixed manic and depressive features (Pavuluri, et al., 2005). Manic symptoms tend to manifest as externalizing behavior problems in youth (Youngstrom, Meyers, Youngstrom, Calabrese, & Findling, 2006) and lack of insight into one's behavior is common in mania (Dell'Osso et al., 2002), which can complicate the diagnostic picture.

It is important to consider the age of onset and the age of the child at the time of assessment in relation to the clinical presentation. Grandiosity and excessive involvement in pleasurable activities can vary as a function of age and development (Bowring & Kovacs, 1992; Geller & Luby, 1997) and comorbidity appears to be higher in PBD (Carlson, 1995; Wozniak, et al., 1995). The range of comorbidity and the quality of presenting symptoms can vary greatly according to age and duration of illness (Post & Kowatch, 2006). Since research on screening instruments for PBD rely on diagnostic criteria and instrumentation developed for adults, there is a need for developmentally-appropriate item content of screening instruments (Geller, Zimmerman, et al., 2002). A case in point, older children report greater depression than younger children, therefore total scores on screening instruments may lead to a bias in assessment (Duax, et al., 2007). A study of 529 youth and caregiver pairs who sought services at a community mental health or

academic medical center, revealed that youth with bipolar disorder reported lower quality of life than youth without bipolar disorder. Adolescents reported more mood symptoms and lower quality of life scores than children with bipolar disorder (Freeman et al., 2009).

While rates of bipolar disorder are equal among males and females later in life, early onset cases, especially prior to the age of 13, are predominantly male (McClellan, et al., 2007). Symptoms manifest themselves differently between boys and girls. Prior to puberty, boys and girls have equal rates of depression. However, during adolescence, the rate of depression among girls doubles (Coyle et al., 2003). Boys present with more manic symptoms and girls present with more depressive symptoms and lower quality of life scores than boys with bipolar disorder do (Freeman, et al., 2009).

Children with an early-onset of affective symptoms are much more likely (71%) to have bipolar disorder than those with late-onset (51%) affective symptoms (Schurhoff et al., 2000). Those with early-onset bipolar disorder are characterized as having the most severe form of the illness (Schurhoff, et al., 2000), which emphasizes the need for early diagnosis and treatment. Rates of youth with serious mood disorders who are undiagnosed and/or inadequately treated are reportedly greater than 70% (Coyle, et al., 2003). Using developmentally-appropriate measures, which are sensitive to age and gender may lead to more timely and accurate diagnosis for children and adolescents.

The proposed study will help us to determine if a test item bias exists in screening instruments between children 5 to 10 years of age and 11 to 18 years of

age, and will help us to develop better assessment and diagnostic strategies for children and adolescents of both genders in the future. The effects of race, age and gender in the assessment and diagnosis of PBD need to be examined so as to impact favorably the reduction of health disparities experienced by AA youth, and to understand and better address issues related to age and gender.

Conclusion

Significant progress in diagnosing children with PBD has been made in the past decade. However, many limitations in screening and assessment of youth currently exist. Left unrecognized and untreated, PBD has a devastating impact on the well being of children (Youngstrom, Findling, Kogos Youngstrom, & Calabrese, 2005), can lead to an increased risk for morbidity and mortality, and places a large economic burden on the U.S. (Murray & Lopez, 1997). Developing accurate diagnostic instruments and incorporating parental measurements are key components to improving the diagnosis of PBD. Youngstrom and colleagues (Youngstrom et al., 2004) report that the General Behavior Inventory-Parent Version (P-GBI) (Youngstrom, Findling, Danielson, & Calabrese, 2001), and the Child Behavior Checklist (CBCL) (Achenbach & Rescorla, 2001) helped to identify cases with PBD, and aided in decreasing the rate of false-positive diagnoses in their study sample. A study of 157 children with BP I, II and not otherwise specified (NOS) showed that BP children have more severe psychopathology than healthy controls and children with other psychopathology. However, the CBCL and the CBCL-PBD were not useful as a proxy for the diagnosis of BP using DSM-IV diagnostic criteria (Diler et al., 2009). Further data from an ongoing

investigation (NIMH R01 MH066647, PI: Youngstrom) have replicated the superiority of the P-GBI to the CBCL in a low-income community mental health sample and have demonstrated that parent report using the Mood Disorders Questionnaire-Parent (P-MDQ) (Hirschfeld et al., 2000) performs comparably to the P-GBI and significantly better than adolescent report on either the P-MDQ or P-GBI (Youngstrom et al., 2005). A study assessing the sensitivity of full and brief forms of the CMRS-P in parents of 150 subjects diagnosed with BD, attention deficit hyperactivity disorder or healthy controls reported that the Brief CMRS-P is highly sensitive and specific for differentiating mania from ADHD and has high accuracy in the detection of variations between presentations of pediatric mania (Henry, Pavuluri, Youngstrom, & Birmaher, 2008).

With increasing rates of diagnosis and the continuing problem of under-detection and misdiagnosis in youth, the need for empirically-derived and clinically useful assessment strategies and instruments is more crucial than ever.

Research concerning the diagnosis of PBD in African American (AA) population has been meager, despite the fact that racial disparities in the diagnosis exist. Assessment tools need to be validated to ensure that they are linguistically and culturally relevant.

Given symptom differences between older and younger children and that research on screening instruments for bipolar disorder has relied on diagnostic criteria and instrumentation developed for adults, it remains unclear if the item content of these instruments is developmentally appropriate for children and adolescents. Furthermore, while the rates of diagnosis between boys and girls are

equivalent, symptoms may manifest themselves differently between boys and girls. These and other factors, yet to be examined, may lead to bias in the assessment of youth with PBD.

Accurate diagnosis and timely and appropriate intervention are greatly influenced by valid measurement techniques. Well-validated screening instruments may increase accurate and timely diagnosis and treatment of PBD and may also lead to better utilization of mental health services in the future.

Chapter 3

DESIGN AND METHODOLOGY

Research Questions

The overall short-term goal of this research study is to assess the psychometric properties of five screening instruments commonly-used to detect bipolar disorder in youth (see Table 1). A secondary analysis will be used to determine if a test item bias (or differential item functioning) exists within each of these instruments. Data from Dr. Eric Youngstrom's existing database (of 758 African American and White families) will be used.

The primary specific aim of this study is to obtain an answer to the following question: 1) Does a test item bias (or differential item functioning) exist between race (African American and White) when using the Young Mania Rating Scale – Parent Version (P-YMRS), General Behavior Inventory – Parent Version (P-GBI), Child Mania Rating Scale-Parent Version (P-CMRS), Mood Disorders Questionnaire-Parent Version (P-MDQ) or the Child Behavior Checklist (CBCL) in urban youth? The secondary aim of this study is to obtain an answer to the following two questions: 1) Does a test item bias (or differential item functioning) exist between age (children 5-10 years of age and adolescents 11-18 years of age) when using the P-YMRS, P-GBI, P-CMRS, P-MDQ or CBCL in urban youth?; and 2) Does a test-item bias (or differential item functioning) exist between gender (boys versus girls) when using the P-YMRS, P-GBI, P-CMRS, P-MDQ or CBCL in urban youth?

Design

This is a secondary analysis of archival data that was collected by Dr. Eric Youngstrom and his colleagues from an urban community mental health center and an academic outpatient mood disorders clinic. The specific aims of the study for which Dr. Youngstrom collected this data were: 1) To develop and prospectively cross-validate a juvenile BPSD screening protocol optimized for use in community settings with representative, racially diverse populations; 2) To clarify the characteristic features of “subsyndromal” bipolar spectrum disorders, currently referred to as “Bipolar NOS” (Nottelmann, et al., 2001), in both research and community settings; and 3) To investigate developmental changes in symptom occurrence and presentation across the age-span from ages 5 to 17 years (including an examination of the validity of this diagnosis in terms of functional impairment and longitudinal course).

Validity

Validity refers to the approximate truth of an inference about a relationship (Cook & Campbell, 1979). Establishing the validity of a test can be difficult because psychological variables such as mood are usually abstract (Anastasi & Urbina, 1997). The standards for Educational and Psychological Testing list three main methods of establishing validity as content related, criterion related and construct related (American Educational Research Association, et al., 1999).

Threats to validity are factors that may lead to an incorrect inference. It is essential that test validation go through continual refinement as conceptualization and research on constructs change over time (Smith & McCarthy, 1995).

Ecological validity

Ecological validity is the ability for a measure to relate to real world functioning (Kazdin, 1999; Messick, 1995). Many past environmental epidemiology studies have not included measures that focus specifically on everyday function. For example, some research studies have used self report and teacher measures of bipolar disorder in children. However, research in the field of PBD shows that parent rating scales are better in terms of assessing PBD (Youngstrom, Meyers, Youngstrom, Calebrese, et al., 2006) which is why the parent versions of the YMRS, GBI, CMRS and the MDQ were chosen for the purpose of this study.

Criterion Validity

Criterion validity, also referred to as empirical or predictive validity is determined by comparing test scores with a performance on an outside measure (Anastasi & Urbina, 1997). Criterion-Related Validity, the degree to which scores on an instrument are correlated with some external criterion will be assessed in this study. In this study we will determine whether each of the scales, the P-YMRS, P-GBI, P-CMRS, P-MDQ and the CBCL, are useful predictors of PBD diagnosis in youth of different racial and age groups and between genders.

An important consideration is the degree to which a specific test can be applied to a unique environment (Hogan, Hogan, & Roberts, 1996). In other words, can these instruments, which have adequate criterion validity in diagnosing PBD in white adolescents, also be used to provide accurate assessments and

predictions for the diagnosis of PBD in different racial and age groups and between genders?

Predictive validity

Concurrent predictive validity, the diagnostic efficiency of a measure in demonstrating validity in terms of assigning children into categories such as clinical diagnosis, is most commonly reported in terms of sensitivity and specificity. Sensitivity, the percentage of children that truly have PBD and that are classified correctly, and specificity, the rate of children who do not have PBD and are not classified correctly (Kraemer, 1992). Both have been established for the 5 instruments which will be used in this study in previous studies of primarily white adolescents. This study will examine concurrent predictive validity in different racial and age groups and between genders.

Sample

The archived data previously collected by Dr. Youngstrom and his research team from an urban community mental health center and an academic outpatient mood disorders clinic consists of 758 African American and White families. Twenty percent of the subjects in this sample have a diagnosis of bipolar disorder. The following definition of terms will be used for the purpose of this study. African American is defined as an individual of African heritage. The African Americans in this sample do not necessarily share the same cultural, religious or ethical beliefs. The urban youth in this sample are derived from two sources. Some urban youth are from families (N=621) with low SES, as defined by Medicaid status, who lived

and received services in an urban area. The remaining (N=209) families included in this sample sought outpatient evaluation at an urban academic medical center. These families came from a much larger catchment area, including an urban Midwestern city, but also extending into the suburbs and surrounding rural regions. These families had a wide range of education, race/ethnicity, and socio-economic status, with the median being middle class and European American with at least some post-secondary education.

Study participants were recruited and screened for more than a dozen different pharmacotherapy studies from an urban setting. Target diagnosis for protocols included bipolar disorder (bipolar I and cyclothymia or bipolar not otherwise specified NOS), unipolar depression, ADHD, conduct disorder, and aggressive behavior regardless of diagnosis. Recruitment was based on presenting symptoms and willingness to participate in treatment protocols. Advertisements and referrals described treatment studies, and those families interested in various treatment studies completed the diagnostic assessment as a screening or baseline evaluation. The sample was enriched by referrals of children whose parents had a diagnosed bipolar disorder and were participating in treatment or research at an affiliated adult mood disorder clinic. In addition, youths (including normal controls, without a psychiatric diagnosis) were recruited by flyers and word of mouth. Their parents were asked to complete these descriptive psychometric instruments at the Child and Adolescent Psychiatric Clinical Research Center. Assessments took place at an outpatient clinic in an urban Midwestern city. To be included in this study, individuals had to meet the

following **Inclusion criteria:** (i) youths between the ages of 5 years, 0 months and 17 years, 11 months of age, (ii) of either gender, (iii) of any ethnicity, (iv) presenting for an outpatient evaluation for which the youth provided written assent and the guardian provided written consent for participation, and (v) both the youth and the primary caregiver presented for the assessment. Individuals were excluded from the study based on the following **Exclusion criteria:** (i) inability of both the youth and the parent to communicate orally at a conversational level in English to complete the interview, (ii) having a pervasive developmental disorder, as determined by psychiatry history, psychiatric interview, or having an Autism Screening Questionnaire score of 15 or higher (Berument, Rutter, Lord, Pickles, & Bailey, 1999), and (iii) suspected moderate, severe or profound mental retardation, documented via educational history, standardized cognitive ability test scores of less than 70, or a Peabody Picture Vocabulary Test-Third Edition (Dunn & Dunn, 1997) score of less than 70 as a screener.

All eligible participants completed the same assessment procedures, including the index tests and reference standard diagnostic interview, regardless of presenting symptoms or treatment study eligibility. The design was “prospective” in the sense that data collection and analysis were planned before the index test and reference standard were performed (Bossuyt, Bruns, et al., 1992) as opposed to *post hoc* examination of a variety of measures collected for a different purpose.

Power Analysis

Power Analyses. With $N = 769$, a power = .80 will be used to detect effects of Cohen's (1988) $f^2 = .010$ or larger for slope or intercept bias parameters in a regression model with alpha level = .05, and $f^2 = .015$ for alpha = .01 (Faul, Erdfelder, Buchner, & Lang, 2009) which would allow using a more conservative threshold to reduce the risk of Type I error due to the number of analyses needed to investigate symptom/item level bias. Thus, the study is powered to detect even small effects of race, gender, or age on item performance: According to Cohen's (Cohen, 1988) criteria, $f^2 = .02$ would be a "small" effect, whereas .15 would be a "medium" and .35 would be a "large" effect for the social sciences. Power will be higher for the analyses based on multi-group confirmatory factor analysis and Item Response Theory (IRT) because these are multivariate techniques that will concentrate on latent variables instead of using observed variables as approximations, thus minimizing the effects of imperfect reliability and validity of the observed variables.

The application of Item Response Theory (IRT) methodology enables the identification of items exhibiting item bias or differential item functioning (DIF), in which responses to an item are not only affected by the level of theta, but also by extraneous characteristics, such as sex, race, (Teresi, 2001) or gender. The use of IRT-based methods in the proposed study will be used to evaluate the P-YMRS, P-GBI, P-CMRS, P-MDQ and the CBCL in the assessment of PBD.

It is also worth noting that the samples used in the research project exceed all recommended minimal thresholds for both Confirmatory Factor Analyses (CFA)

and IRT. Specifically, CFA authorities recommend having a minimum of 100 cases ($n = 573$ for the main subsample, and $n > 100$ for all proposed subsamples) (Kline, 1998), and smaller sample sizes are even less of an issue when the main sample is larger; all analyses will maintain at least a 20:1 ratio of subjects to indicators in the primary sample, and all factors will have between four and 28 indicators with moderate to high loadings. All of these are parameters that increase the power of the factor analysis and improve the likelihood of recovering the true structure (Guadagnoli & Velicer, 1988; MacCallum, Widaman, Zhang, & Hong, 1999). Similarly, the sample sizes exceed the “rule of thumb” thresholds established for IRT analyses, including having at least 500 observations in the primary subsample when using graded response models (du Toit, 2003).

Study Measures

Young Mania Rating Scale (P-YMRS). (Gracious, Youngstrom, Findling, & Calabrese, 2002). The P-YMRS is an 11-item questionnaire adapted from the YMRS (Young, et al., 1978). Parents rate their child’s manic symptoms on five explicitly defined grades of severity, with item scores ranging from 0 to 4 (and 3 items ranging from 0 to 8). The P-YMRS yields a total score that can range from 0 to 56, with higher scores representing greater psychopathology. Ratings are based on the reported presence of symptoms over the past week. Preliminary work suggests that the P-YMRS does well in discriminating bipolar spectrum disorders from unipolar depression, disruptive behavior disorders, and a mix of other psychiatric disorders when modified for use as a parent-report (Gracious, et al., 2002) and when used as a clinical rating scale (Danielson, Youngstrom,

Findling, Gracious, & Calabrese, 2001). Internal consistency is adequate (e.g. $\alpha = 0.80$ in the age 5-10 sample, and 0.69 in the older sample).

General Behavior Inventory (Parent Version) (P-GBI), (Depue, Krauss, Spont, & Arbisi, 1989). The GBI was originally developed for adults with bipolar disorder and later modified as a parent-report inventory (P-GBI) for youth as young as five years of age (Findling, et al., 2001; Youngstrom, et al., 2001). Parents complete the P-GBI to rate the depressive, hypomanic, manic, and biphasic mood symptoms of their children aged 5-17. The two scales of depressive and hypomanic/biphasic symptoms have strong construct validity (Youngstrom, et al., 2001) and exceptionally high internal consistency (e.g. alphas of 0.97 for depression and 0.94 for hypomanic/biphasic in both age groups) (Youngstrom, et al., 2001). The hypomanic/biphasic score has shown promise as a potential screener for bipolar disorder, based on preliminary analysis of a subsample of these youths (Findling et al., 2002; Youngstrom, et al., 2001).

Child Mania Rating Scale-Parent Version (CMRS-P), (Pavuluri et al., 2006). This is a 21-item rating scale. Each item is developmentally specific for children and adolescents and corresponds with the DSM-IV-TR diagnostic criteria symptoms (American Psychiatric Association, 2000). A score of greater than or equal to 15 (out of 63) indicates a 92% chance of having the diagnosis of PBD (Pavuluri, Henry, Carbray, & Birmaher, 2004). Exploratory and confirmatory factor analysis of the CMRS-P indicated that the scale was unidimensional. The internal consistency and retest reliability were both 0.96. Criterion validity was demonstrated in analysis of receiver operating characteristics curves, which

showed excellent sensitivity and specificity in differentiating children with mania from either healthy controls or children with ADHD (areas under the curve of .91 to .96) (Pavuluri et al., 2006).

Mood Disorders Questionnaire-Parent Version (P-MDQ) (Hirschfeld, Williams, Spitzer, & al., 2000). This youth scale (Wagner, et al., 2004) is a slightly modified version of the MDQ. Parents are asked to complete the 15 item questionnaire to report potential manic symptoms in their child. The P-MDQ is promising as a diagnostic aid because it uses parent report, is brief, and is focused specifically on symptoms of mania. Preliminary data from other studies show that it out-performs self report on the MDQ (Youngstrom, Meyers, et al., 2005).

Child Behavior Checklist (CBCL) (Achenbach, 1991). The 2001 version of the Child Behavior Checklist for ages 6-18 (CBCL/6-18) (Achenbach & Rescorla, 2001) is one of the most widely used instruments in children and adolescents in research and clinical settings (Sattler, 2002). The CBCL is a rating scale that is filled out by parents in reference to their child or adolescent. It includes 118-problem behavior items rated from 0 (not at all typical of the child) to 2 (often typical of the child). One week test/retest stability coefficients are .89 for internalizing problems and .93 for externalizing problems, with alphas for both > .90 (Achenbach, 1991).

The Kiddie Schedule for Affective Disorders and Schizophrenia-Present and Lifetime Version (K-SADS-PL). The K-SADS assessed subjects for major

psychiatric diagnoses in the sample from Dr. Youngstrom's research, which will be used for the secondary analysis in the proposed study. The K-SADS represents the most widely used semi-structured psychiatric diagnostic tool used in investigations of PBD (Nottelmann, et al., 2001) and identifies symptoms of bipolar disorder and schizophrenia in children. The K-SADS is comprised of Nottelmann items that are considered specific to rapid cycling and prepubertal mania, ADHD, and the rest of the DSM-IV diagnoses. The scale includes current and lifetime diagnostic items pertaining to the onset and offset of each symptom and syndrome (Geller, et al., 2000a). One preliminary study of 22 bipolar youth reported a Cronbach's alpha of .94 and an inter-rater reliability of .97 (Axelson et al., 2003). Other studies have found inter-rater reliability equal to 100% using 5 consecutive interviews with raters blind to diagnoses and a 85.7% six month stability measure in bipolar I and bipolar II diagnoses (Geller et al., 2000b; Nottelmann, et al., 2001).

Data Analysis

Data Management and Quality Assurance. A half-time MA-level research assistant and three predoctoral psychology interns conducted the interviews under the supervision of a PhD psychologist (Kogos-Youngstrom). The interns each completed one interview per week, and the MA level completed three interviews per week. Every tenth interview was done conjointly, minimizing rater drift and preserving reliability. Assessments were completed by a MA/RN level research coordinator and doctoral candidate graduate students. Clinical confirmation of diagnoses and LEAD integration of data was performed by Dr. Feeny and Dr. Findling.

Protocols were reviewed for completeness after each interview.

Interviewers, all of whom were psychologists, predoctoral psychology interns, master's level research assistants or clinical psychology graduate students were trained to a high standard of reliability ($\kappa > .85$ across five consecutive interviews), and every tenth interview was done conjointly to minimize rater drift. Interviews were videotaped at both sites, and a random draw of 20% was re-rated to provide an additional reliability check. All data entry and archiving took place at CWRU/UHC. Various procedures were done to assure the quality of the data. All of the questionnaires and interview data were entered into SPSS v. 10 Data Builder/Entry, using the double-entry feature. SPSS Syntax files were then used to create the scales and subscales for the different measures. The syntax files were printed and reviewed for accuracy. Consistent variable and value labels have been developed and were included in the data file. The SPSS data file acts as the source for any data files created for use in other statistical software, thus ensuring that all analyses will rely on appropriately cleaned data. Internal consistency reliability coefficients were calculated on the full sample.

Preliminary Data Screening

All data will be examined using graphical procedures as well as descriptive statistics, including skewness and kurtosis. Univariate and multivariate outliers will be evaluated (using Cook's Distance, Mahalanobis' Distance, studentized deleted residuals, and Mardia's κ) to determine if they will unduly influence the results of statistical analyses, in which case appropriate transformations or other

remedies will be considered (following the general analytic plan in Tabachnick & Fidell (Tabachnick & Fidell, 2007).

Determining Presence of Different Measurement Properties Across Groups

Four distinct but related approaches will be used to evaluate potential changes in measurement across groups (race as primary comparison, gender and age-group as secondary comparisons). These methods provide a cumulative, didactic progression as well as offering a more complete understanding of the measurement properties. The first approach is a classic regression framework. The item score will be the dependent variable and the total score (minus the target item) will be used as a predictor. The total is a proxy for the latent variable (of mania, in this case), and the R^2 measures the extent to which the item shares variance with the construct. A dummy code is also entered for race (AA and White), providing a test of whether the groups show different average scores on the item after controlling for levels of total score on other items. If the coefficient for the group dummy code is significant, that is evidence of “intercept bias” – that the two groups differ in average scores on the item for reasons other than differences in their total score on the measure. An interaction term will also enter the model, testing whether there is a difference in the slope linking total score to predicted item score. If the interaction term is significant, then that indicates that the validity of the item as an indicator of the construct is significantly different across the two groups.

The second approach is to perform exploratory factor analyses separately in the two groups to determine if similar structures emerge. The decision rule

about number of factors will be based on the three most accurate decision rules: The Scree Test, Minimum Average Partial, and Gorfeld's extension of Horn's Parallel Analysis. If it appears that different numbers of factors are evident for each group, then separate models will be developed for each group, and DIF will be examined only for subscales with similar item content.

The third approach is Multi-Group Confirmatory Factor Analysis (MG-CFA). If there are the same number of factors evident in both groups, then the classic Joreskog framework will be used for evaluating measurement invariance (Jöreskog, 1993), specifying equality constraints in all elements of the covariance matrix for the two groups, and then relaxing sets of constraints to evaluate at what level the measurement properties are similar between the two groups. Two changes will be made to adapt the classic, set-wise strategy for use in DIF: (1) adding structured means (analogous to the test of intercept bias in the regression approach) and (2) examining the modification indices individually for all of the factor loadings and means, rather than only treating them as sets, so that the sources of the DIF can be identified. The test of equality of each factor loading is analogous to the test of slope bias in the regression framework. However, the MG-CFA approach is an improvement on the regression framework because it (a) evaluates all the items simultaneously, (b) provides a purer estimate of the latent construct than the approximation offered by the total score in the regression approach, and (c) models the effects of measurement error.

The fourth and final approach will be to use Item Response Theory (IRT), in this case the Graded Response Model (Hambleton & Swaminathan, 1985), to

evaluate whether the items have the same difficulty and information across the two groups. IRT assumes that a single factor underlies the items, so the EFA and MG-CFA are logical precursors, and only single factors will be evaluated in each IRT run. The IRT model will allow calibration of the items, and test whether the items indicate similar degrees of manic severity across the two groups. The four sets of analyses will be repeated for the P-YMRS, P-GBI, P-MDQ, CMRS-P, and the CBCL comparing between races (African American and White), different age groups (ages 5 to 10 versus 11 to 18), and gender in youth. The term differential item functioning (DIF) is used to describe the empirical evidence obtained in the investigation of bias. For example, if the minority group (African American) and majority group (White) differ in their mean performance on an item after controlling for levels of the trait of interest (in this case, mania), then the item shows DIF.

Item response theory (IRT) provides a unified framework for conceptualizing and investigating bias at an item level (Hambleton & Swaminathan, 1985). IRT assumes that there is only one factor underlying the set of items. It is particularly strong in identifying differences in the average amount of the underlying concept (or trait) between the two groups. A noted limitation of IRT and consideration in adding MG-CFA to this analysis, is that IRT does not consider the possibility that other variables may be responsible for the difference in the p-values between groups. Therefore, when there is more than one factor involved, such as in the case of PBD, which is a highly comorbid condition with many overlapping symptoms with other psychiatric disorders, Multi-Group Confirmatory Factor Analysis will produce a pure outcome, rather than an approximation,

essentially eliminating the effects of measurement error. MG-CFA will essentially assess the equality of the regression loading to identify any differences in African American and White youth, setting an equality constraint. SPSS and M-Plus will be used for the analyses. M-Plus allows implementation of both IRT and MG-CFA models (neither of which are included as core methods in SPSS).

Human Subjects

Study procedures have been reviewed and approved by the IRB at the University of North Carolina, Chapel Hill and at Northeastern University. Dr. Youngstrom added the applicant to his protocols through the UNC IRB.

Protection of human subjects will be assured by:

1. Recruitment of large community sample

The archived data to be used for the purpose of this study was previously collected and entered into a data base from a research study which involved a consortium of community mental health centers and an academic mood disorders clinic. All participants freely agreed to participate before they were interviewed during the data collection period.

2. Confidentiality

All data resides on a password-protected computer in the Department of Psychology with only the P.I., Dr. Eric Youngstrom, and key members of the research team having access. No identifying information is included in the data file beyond general demographic information and an arbitrary case identification number. Only de-identified data will be used for the purpose of the proposed secondary analyses.

3. Informed Consent

In order to participate in this study, the guardians of all subjects provided written informed consent. Youths also provided written informed assent prior to participation in this study. Consent forms with identifying information are kept in a locked file cabinet and are only accessible to the PI, Dr. Eric Youngstrom.

4. Risk

The children, adolescents and their parents were interviewed using structured diagnostic interviews and screening instruments. There were no significant risks to the subjects. Subjects were informed of minor risks that included the possibility of them feeling tired or distressed during the interview process. However, they were informed that they could refuse to answer any questions that made them uncomfortable and that they could stop or withdraw from the study if they chose to at any time. There are no current risks to subjects as the proposed study is a secondary analyses of previously collected data, that is stored in a password protected computer.

5. Data Safety and Monitoring Plan

In compliance with NIH policy for the protection of human subjects in clinical studies, Dr. Eric Youngstrom, P.I., implemented the following Data Safety Monitoring Board (DSMB) policy. The DSMB policy serves to protect the health and safety of subjects and provide information relevant to subjects' continuation in clinical studies. If there was an adverse consequence of completing the diagnostic procedures, then research staff immediately notified Dr. E. Youngstrom on site at CWRU/UHC, and Dr. J.K. Youngstrom on site at ACI. If an event occurred at ACI,

Dr. J.K. Youngstrom would inform the PI, Dr. E. Youngstrom, within the same day. They tracked any children who identified as possibly being in harm's way during the course of the assessment (i.e., by endorsing suicidal ideation items on the CDRS-R, P-GBI, A-GBI, or through other disclosure of potential self harm; or due to concerns about abuse): research staff then reviewed these critical indicators on the day of data collection, and informed the P.I. of potential risks. Weekly team meetings were held and a review of adverse events in patients was assessed, along with their disposition. Dr. E. Youngstrom (PI) notified the chair of the DSMB in the event of an emergent adverse event, and he consulted with the chair as necessary about ambiguous cases.

The DSMB is chaired by Daniel J. Rapport, MD (the medical expert monitor from R01 MH-50165, the Supplement to MH-50165, and R21 MH-62650) and includes Eric Youngstrom, PhD, and a consumer/parent. The consumer/parent is Denise O'Brien, who has personally experienced the suicide of an adolescent son with bipolar disorder. The monitoring of the research activities of the center took place in "real time" (within 24 hours during the week and on the next working day following a weekend) and focused on issues of safety. In addition to "real time" monitoring, quarterly meetings were scheduled; ad hoc meetings were called to evaluate unanticipated serious adverse events.

6. Gender and Minority Inclusion

Both male and female African American and White subjects were recruited for this study.

Participation of Children

Because the project goals were to establish the prevalence of bipolar spectrum disorders and to develop appropriate screening procedures in a community child and adolescent mental health care facility, participants recruited included minors between the ages of 5 and 17.

Chapter 4

Results

The archived data was derived from an urban community mental health center and an academic outpatient mood disorders clinic, collected between July 2003 and March 2008. As seen in Table 3, the sample consists of a total of 758 families, 573 African American and 185 White. The urban youth in this sample were derived from two sources, 584 who were evaluated in an urban area community mental health center, and the remaining ($N=174$) who were evaluated in an outpatient urban academic medical center. These families came from a much larger catchment area, including an urban Midwestern city extending into the suburbs and surrounding rural regions.

As seen in the graphs in Figures 2, 3, and 4, the families included in this sample had a wide range of education, socio-economic status and Hollingshead occupational scale scores, with 31% ($n=238$) of the families having at least a high school diploma and 29% ($n=222$) having 1 to 3 years of college, business or trade school. Seventy-two percent ($n=437$) of the families in this sample were considered to be of low SES, with an income less than \$20,000. Eighty percent of the families enrolled at the urban community mental health center were in the low SES group as defined by Medicaid eligibility versus only 20% of the families from the academic center.

Participants at the academic center were seeking treatment from more than a dozen different treatment studies in an urban setting. Target diagnosis for protocols included bipolar disorder (bipolar I and bipolar II, cyclothymia or bipolar

not otherwise specified NOS), unipolar depression, ADHD, conduct disorder, and aggressive behavior regardless of diagnosis. As seen in table 4, approximately 20% of the subjects in this sample have a diagnosis of bipolar disorder I, II, NOS or cyclothymia: African American ($n=74$), White ($n=56$), Younger ($n=62$), Older ($n=90$), Male ($n=83$) and Female ($n=69$).

Recruitment was based on presenting symptoms and desire to participate in a treatment protocol if eligible. Advertisements and referrals described treatment studies, and families interested in various treatment studies completed the diagnostic assessment as a screening or baseline evaluation. The sample was enriched by referrals of children whose parents had a diagnosed bipolar disorder and were participating in treatment or research at an affiliated adult mood disorder clinic. In addition, youths (including normal controls) were recruited by flyers and word of mouth. Their parents were asked to complete these descriptive psychometric instruments at the Child and Adolescent Psychiatric Clinical Research Center. All eligible participants completed the same assessment procedures, including the index tests (screening instruments) and the reference semi-structured diagnostic interview using the Washington University Schedule of Affective Disorders and Schizophrenia for School-Age Children Present-Lifetime Plus (K-SADS) (Geller et al., 2001) regardless of presenting symptoms or treatment study eligibility. The design was “prospective” in the sense that data collection and analysis were planned before the index test and reference standard were performed (Bossuyt, Reitsma, & Bruns, 1992) as opposed to *post hoc* examination of a variety of measures collected for a different purpose.

Data Screening

All data was examined using 1) graphical procedures; 2) descriptive statistics, including skewness and kurtosis; 3) univariate and multivariate outliers (using Cook's Distance, Mahalanobis' Distance, studentized deleted residuals, and Mardia's kappa) to determine if they would unduly influence the results of statistical analyses. The cases with incomplete or missing data including 5 from the age cohort, 2 from the gender cohort and 71 from the racial cohort, were dropped from the total sample of 829. This left 824 cases in the age analysis, 827 in the gender analysis and 758 in the racial analysis. The larger amount of missing data for the racial cohort was due to excluding participants who self-identified as belonging to any ethnic or racial group other than Black or White.

Determining Presence of Different Measurement Properties Across Groups

Four distinct but related approaches were used to evaluate potential changes in measurement across groups (race as the primary comparison: 1) Exploratory Factor Analyses; 2) Classic Regression Framework; and 3) Confirmatory Factor Analysis; and 4) Item Response Theory.

Exploratory Factor Analyses

Exploratory factor analyses (EFA) is an important statistical tool for providing validity evidence concerning the structure of the instruments (Munro, 2005). EFA was used to determine the underlying structure of the items of the P-YMRS, P-GBI, P-CMRS, P-MDQ and the CBCL. Exploratory factor analysis using list wise deletion was conducted separately between the Black and White groups to determine if similar structures would emerge. Principal component analysis was

used for factor extraction; if more than one factor was retained then promax rotation, with Kaiser normalization, was used to facilitate interpretation using SPSS student version 15.0.

Internal Consistency of Items

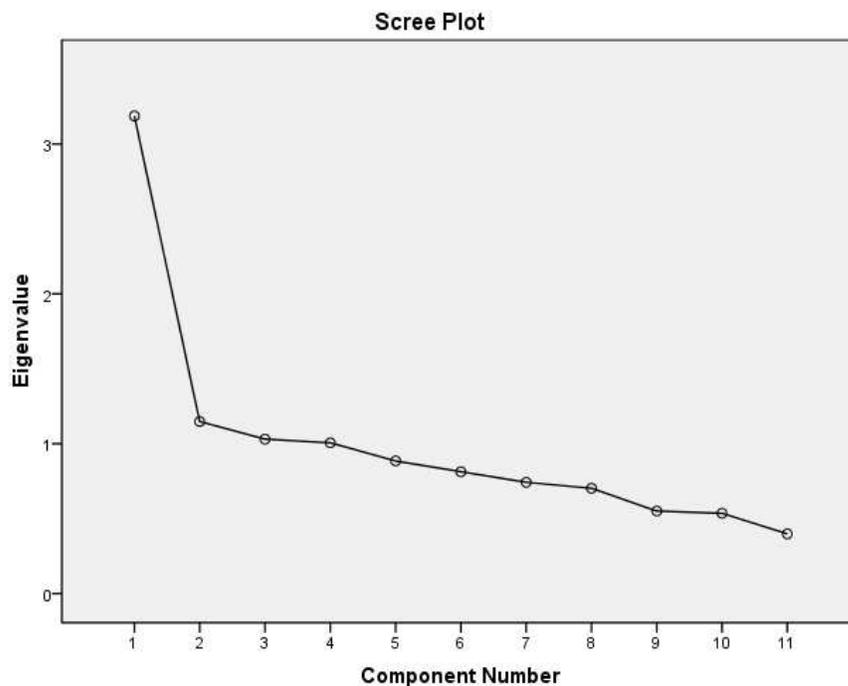
Three methods were used to select the number of factors to retain: the Scree test, Glorfeld's extension of Horn's parallel analysis, and Velicer's Minimum Average Partial Test (Glorfeld, 1995; Horn, 1965; Velicer, 1976). For the P-YMRS, correlations among the 10 items reflected a single underlying dimension with a total score showing: acceptable internal consistency ($\alpha = .75$) replicating previous studies (Gracious 2002; Youngstrom et al., 2002). Similarly, good internal consistency emerged for the 13 items on the P-MDQ ($\alpha = .83$); the ten items on the P-GBI ($\alpha = .88$); the 21 items on the P-CMRS ($\alpha = .90$).

The Scree Test

Bartlett's chi square test and the Scree method have an underlying similarity, based on an analysis (one statistical, the other visual) of the essential equality of the remaining eigenvalues (Horn & Engstrom, 1979) To determine the number of factors, a Scree plot (a 2 dimensional graph with the factors on the x-axis and the eigenvalues on the y-axis) was generated. Eigenvalues represent the variance accounted for by each underlying factor. An 11 item scale such as the YMRS theoretically has 11 possible underlying factors. Each factor has a eigenvalue that indicates the amount of variation in the items accounted for by each factor. EFA uses the Scree plot (a visual representation of eigenvalues) to

determine the number of factors that clearly rise above the Scree (or baseline). If a factor rises above the Scree, then there is a correlation between the item and the factor. The bigger the item, the more strongly it is correlated with the factor. The eigenvalues (Table 5) are arranged in a Scree plot by descending order to identify the Scree, or the point at which the slope of decreasing eigenvalues approaches zero. This is the point at which eliminating additional factors would not eliminate significant variance.

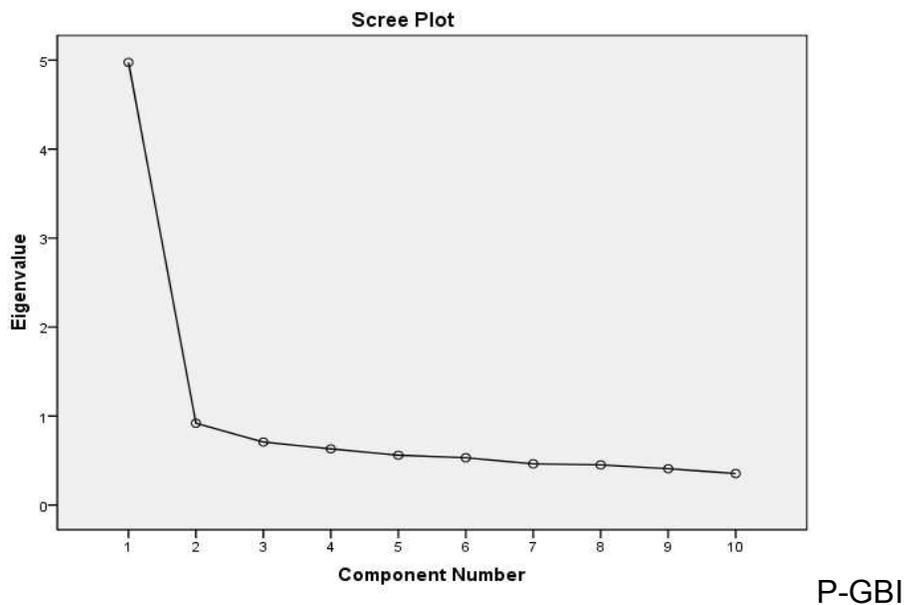
Scree Plot Analysis



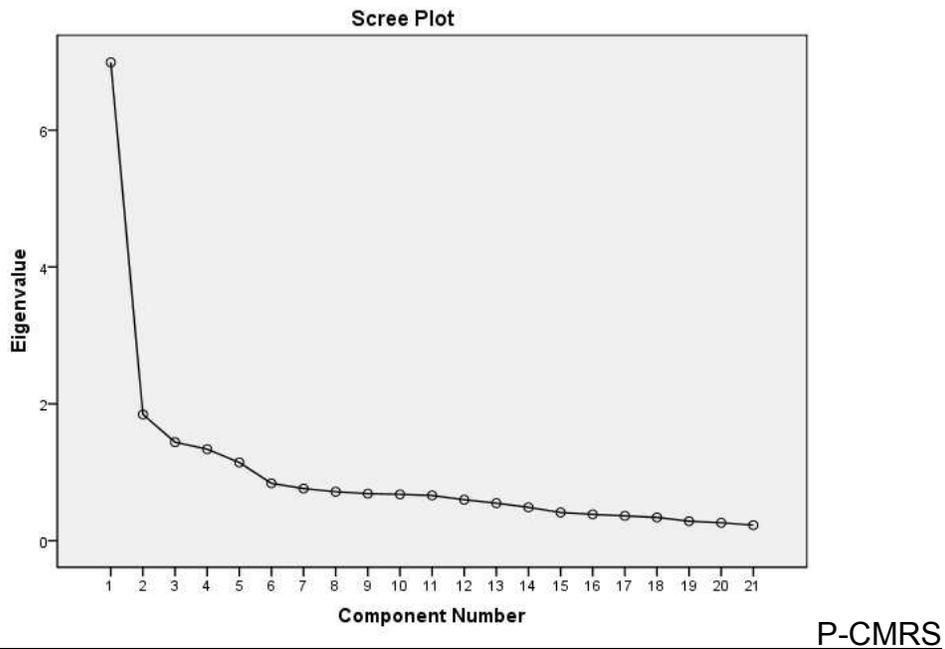
P-YMRS

EFA produced a Scree plot analysis for the 11-item P-YMRS that supported one dominant factor (first eigenvalue =3.2, explaining 29% of the variance). The first factor accounts for most of the variance and the remaining factors have small

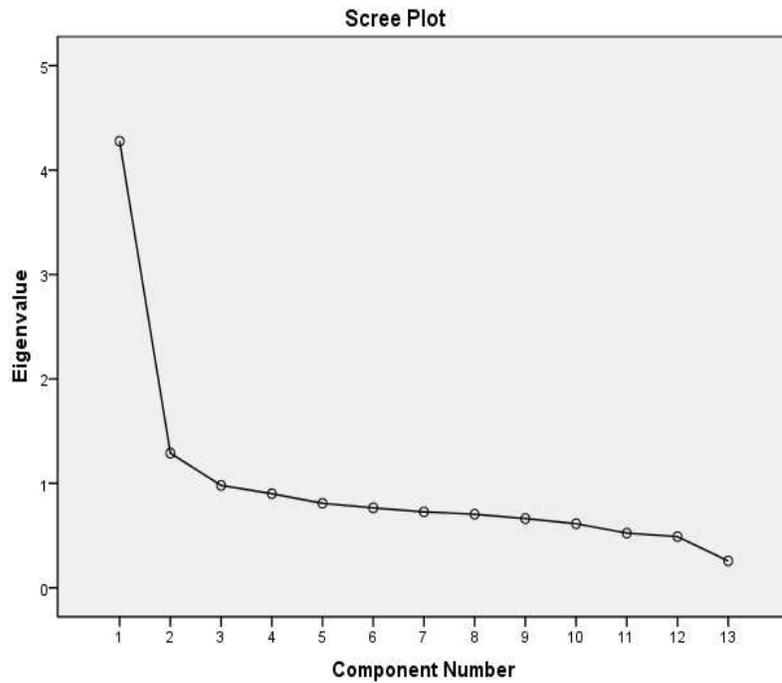
eigenvalues. Item 11 (lack of insight) revealed a very low rate of endorsement (factor loading of 0.081), meaning it does not correlate with the other items in the set so the researcher dropped this item from further analyses of the P-YMRS in this study. This study replicated findings from two previous studies, which raised questions on psychometric grounds, about the inclusion of item 11 in measures of juvenile mania because it appears to be much less related than the other items to the construct of mania (Youngstrom, Danielson, Findling, Gracious, & Calabrese, 2002).



EFA produced a Scree plot analysis for the 10-item P-GBI that supported one dominant factor (first eigenvalue = 9.8, explaining 35% of the variance). The first factor accounts for most of the variance and the remaining factors have small eigenvalues.

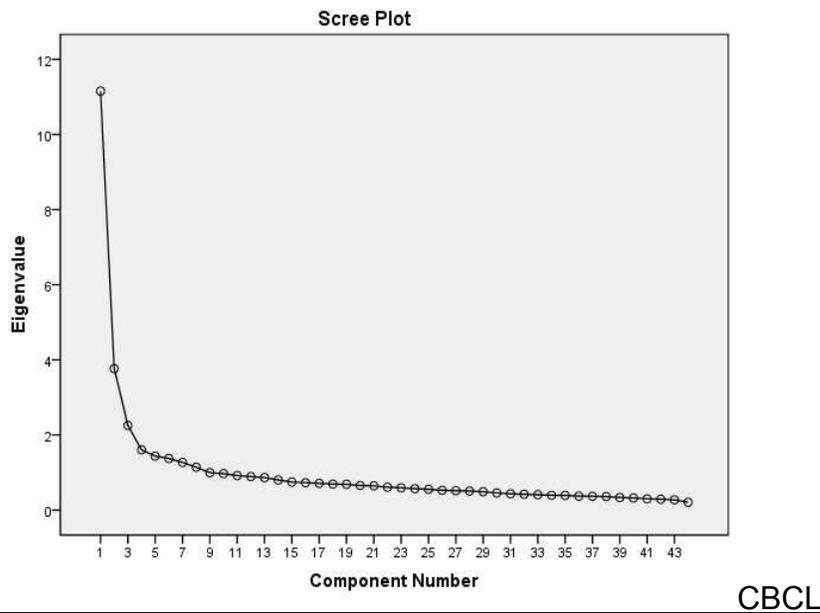


EFA produced a Scree plot analysis for the 21-item P-CMRS that supported one dominant factor (first eigenvalue = 7.0, explaining 33% of the variance). The first factor accounts for most of the variance and the remaining factors have small eigenvalues.



P-MDQ

EFA produced a Scree plot analysis for the 13-item P-MDQ that supported one dominant factor (first eigenvalue =4.2, explaining 33% of the variance). The first factor accounts for most of the variance and the remaining factors have small eigenvalues.



EFA produced a Scree plot analysis for the 43-item CBCL that supported eight dominant factors (first 8 eigenvalues = 11.2, 3.8, 2.3, 1.6, 1.4, 1.4, 1.3, 1.1, 1.0, with the first 8 factors explaining 55% of the variance). The first factor does not account for most of the variance and the remaining factors have significant eigenvalues.

Using a more stringent cut off of eigenvalues of > 2 , the Scree plot analysis supported 3 dominant factors (first 3 eigenvalues = 11.2, 3.8, 2.3, with the first 3 factors explaining 40% of the variance).

EFA methods support three to eight dominant factors for the CBCL, indicating that this instrument is multidimensional, rather than unidimensional. This means that the CBCL is measuring other factors unrelated to mania. Therefore, the CBCL could not be included in further analyses in this study.

Confirmatory Factor Analysis (CFA)

Confirmatory Factor analyses (CFA) is a model which makes the (strong) assumption that the factor loading is due to 1 factor only (in this case mania). CFA allows more precise tests of an instrument's factor structure (Long, 1988). It is a special application of structural equation modeling (SEM). CFA provides a theory-driven method to examine construct validity (Munro, 2005). It assigns items to the factor (mania) according to theoretical expectations. Convergent and divergent validity were examined by assigning items in the instrument to their respective factor (mania) and then indicating according to theoretical expectations which factors are correlated across the instrument and which are not.

Factor Loadings

CFA works optimally when the factor pattern solution exhibits simple structure, this means that each variable is measuring only 1 factor (mania). The quantity of the underlying construct (mania) is presumed to cause an item to take on a certain value. All observed variables are theorized to load on the factor (mania). A factor loading > 0.34 explains 10% of the variance. The observed score (obtained from the items) represents the true score, the quantity associated with the latent variable (mania) minus the measurement error or residual score. Residuals are interpreted as an indication of reliability of observed variables (items).

Item Response Theory (IRT) assumes unidimensionality but because unidimensionality and local independence cannot be strictly met, sets of items with one dominant dimension, in the presence of one or more minor dimensions is

typical. A set of items can be unidimensional but still contain pairs of items that are correlated. Local independence requires that any two items be uncorrelated if the latent variable is fixed (Lord, 1980).

Using CFA to test IRT's unidimensionality, the model is specified in which all items load on 1 factor (mania). The fit of the model is then evaluated using factor model fit statistics and traditional cut offs for these statistics. Assessing item structure and data fit is determined by the success of the selected model in predicting or explaining the data. The classic Joreskog framework was used for evaluating measurement invariance specifying equality constraints in all elements of the covariance matrix for the two groups, and then relaxing sets of constraints to evaluate at what level the measurement properties are similar between the two groups (Jöreskog, 1993).

CFA was used to compare emergent and conventional factor structures. Multigroup CFA was run on the P-YMRS, P-GBI, P-CMRS and the P-MDQ using the MPlus program. Using this approach, multiple problems arose at the item level because the sample size did not always produce two separate positive definite polychoric correlation matrices, making the results uninterpretable and possibly incorrect due to estimation errors. Therefore, the logistical regression approach described and used by Zumbo and colleagues, was used as an alternative approach to assess DIF (Zumbo, 2007) and single-group CFA was run on the P-YMRS, P-GBI, P-CMRS, and P-MDQ to test the model fit.

Test of Model Fit

DIF is the process of looking for extra factors influencing response patterns. Although some fit indices may provide redundant information, each reflects different facets of the model fit (Meade, Johnson, & Braddy, 2008). The researcher examined the model fit and other CFA statistics were examined to clarify whether the items belong on the factor (mania), whether the factors are correlated, and to determine the magnitude of the correlations. IRT analysis makes very strong assumptions about local dependence within the data. The measures in this study showed local dependence, meaning that a secondary non-essential factor (e.g., hyperactivity factor in the mood items) is present. Local dependence could falsely cover true DIF and falsely indicate DIF exists when it does not. This may be contributing to the suboptimal fit indices reported on some of the instruments in this study.

Chi Square

The null hypothesis of perfect fit was tested with chi square. Chi square essentially asks the question, "Does one factor do a good job explaining the items or are there other factors driving the scores of the items?" The relative chi square is the ratio of chi square to the degrees of freedom (χ^2/df) (Munro, 2005). Relative chi-squares less than 3.00 are preferred. However, researcher's definition of a good data-model fit vary from 3.00 to 5.00 (Mueller, 1996). The chi-square statistics are reported as an indicator of the overall model fit. Chi-square as a fit index is sometimes criticized for excessive sensitivity in large samples (>500), which may suggest a poor model fit in the absence of true data issues such as

skewness and kurtosis (Bollen & Curran, 2005). Relative chi-squares ratios of < 3.00 are preferred but some researchers consider ratios between 3.00 and 4.00 and less frequently as high as 5.00 to be a good data-model fit (Geller, et al., 2001; Mueller, 1996).

Other Fit Indices

Findings from research studies suggest using multiple fit indices when evaluating model fit (Fan & Silvo, 2005). Model fit measures assess how well the proposed model captures the covariance between all the items on the test. If some of the items are measuring multiple factors or if they are related to each other more than the other items are, then the fit may be poor (Yu, 2002). Fit Indices are used to quantify features such as the sum of residuals or variance accounted for by the proposed model (Hu & Bentler, 1998) and “to provide information about the degree to which a model is correctly or incorrectly specified for the given data” (Fan, Thompson, & Wang, 1999). Experts recommend using a variety of fit indices so that the strength of a particular index will offset the weakness of another (Gonzalez & Griffen, 2001). Fit indices are often used to supplement the χ^2 test to evaluate the acceptability of latent variable models.

Both Confirmatory factor index (CFI) and the Tucker Lewis Index (TLI) are incremental fit indices, which measure the improvement of fit by comparing a H0 model with a more restricted baseline model. CFI is insensitive to sample size and performs well as a test of invariance (Meade, et al., 2008). It reflects the fit of the hypothesized model relative to an independence model, a model in which all correlations among variables is equal to zero. A CFI of > 0.9 suggests that the

model is useful. A CFI < 0.9 suggests that there may be some inconsistency between the model and the data. TLI can exceed the 0 to 1 range. TLI values tend toward 1 for a correctly specified model (Yu, 2002). Hu and colleagues recommend a cutoff value of TLI close to 0.95 (Hu & Bentler, 1999). Although TLI indicates a greater degree of misspecification, CFI has the advantages of having a 0-1 range and smaller sampling variability. The recommended cutoff value for the CFI is close to 0.95 or higher (Hu & Bentler, 1999). Some researchers suggest that reporting either the CFI or TLI is sufficient because there are high agreements and similarities reported between the two measures. However, the CFI at a cutoff value of 0.95 has better type I and type II error rates than the TLI cutoff value of 0.95 (Yu, 2002).

Root Mean Square Error of Approximation (RMSEA)

RMSEA is another way of looking at the fit of the model. It essentially determines how good the model is in relation to the latent variable (mania), and how good it is for each group. RMSEA performs reasonably well as a test of invariance, is fairly insensitive to sample size, interactions among variables and the reliability of indicators (Meade, et al., 2008), and has known distribution so it permits the calculation of confidence intervals. According to Browne and colleagues, RMSEA values of $> .10$ indicate poor fit; 0.05 to 0.08 indicate fair fit; and < 0.05 indicate close fit (Browne & Cudeck, 1993). The recommended cut off score is close to 0.06 (Hu and Bentler, 1999).

As seen in Table 6, these statistics represent a fair to good fit for the model on the P-YMRS, P-GBI and the P-CMRS. CFI and TLI derived from this model

were less than the recommended cutoff score of 0.95, for the P-YMRS (0.9), P-GBI (0.83), P-CMRS (0.8), indicating some inconsistency between the hypothesized model and the observed data. A CFI of > 0.92 on the P-GBI suggests that the model is useful, as does the TLI > 0.9 for all of the instruments. RMSEA values of $> .10$ for the P-GBI (.19) indicates poor fit. A CFI of 0.96, TLI of 0.97 and a RMSEA of 0.067 indicate good overall fit for the P-MDQ. Assessment of factor loadings and residual correlations reveal that some items correlate with other items on the instruments suggesting that there is local dependence, which can falsely cover true DIF and falsely indicate DIF when it does not exist, contributing to the suboptimal fit indices for some of the instruments in this study.

The chi-square ratio indicated a poor fit for all the instruments in this study. As noted previously, chi-square as a fit index may have excessive sensitivity in large samples (> 500) such as the current sample, which may suggest a poor model fit in the absence of true data issues (Bollen & Curran, 2005).

Classic regression framework

Children and adolescents with the same level of the latent trait (mania) should have similar scores on the items regardless of their race, age or gender. DIF occurs when scores differ between children and adolescents with the same level of mania from different groups. To examine DIF the logistical regression approach described by Zumbo and colleagues was used (Zumbo, 2007).

Best evidence supports the use of principle axis factoring and maximum likelihood approaches. The extraction method used produced factor loadings for each item on every extracted factor. All the instruments, except for the CBCL

showed a simple structure, with most items loading on one factor but having small loadings on other factors. Each item was assessed for DIF using SPSS student version 15.0 software.

The item score (dependent variable) and the total score (minus the target item) was used as a predictor. The total is a proxy for the latent variable (of mania), and the R^2 measured the extent to which the item shares variance with the construct.

A dummy code was entered for race (AA and White): therefore the regression equation is: Score on item = Age dummy code + error, providing a test of whether the groups showed different average scores on the item after controlling for levels of total score on other items. A significant coefficient for the group dummy code is evidence of “intercept bias” (Uniform DIF), which indicates that the two groups differ in average scores on the item for reasons other than differences in their total score on the measure.

RACE: Uniform DIF

A significant coefficient for the group dummy code is evidence of intercept bias (uniform DIF), meaning that the two groups differ in average scores on the item for reasons other than differences in their total score (of mania) on the measure.

DIF was detected only on item 2 of the P-YMRS (table 7) indicating that when all else is held constant, African Americans will score 0.4 points *higher* ($p < .05$) on (**increased motor activity/energy**) than a White at the same level of mania.

DIF was detected on 2 items of the P-GBI (table 8) indicating that when all else is held constant, African Americans will score 0.3 points higher ($p < .05$) on item 4 (**unusually happy/energetic**) and 0.3 points *lower* ($p < .05$) on item 19 (**rapid mood shifts**) than a White at the same level of mania.

DIF was detected on 3 items of the P-CMRS (table 9) indicating that when all else is held constant, African Americans will score 0.4 points *lower* ($p < .05$) on item 2 (**irritable, cranky or mad**); 0.2 points *higher* ($p < .05$) on item 4 (**abilities or powers**) and 0.3 points *lower* ($p < .05$) on item 18 (**rapid mood swings**) than a White at the same level of mania.

DIF was detected on item 6 (**racing thoughts**) of the P-MDQ (table 10) indicating that when all else is held constant, the odds of Whites answering yes to this item was 3.3 times *higher* ($p = 0.009$) than an African American at the same level of mania.

For reasons other than differences in their level of mania, African Americans report higher rates of increased activity, having more energy, feeling unusually happy or energetic, and having special abilities or powers and lower rates of rapid mood shifts, irritable or cranky moods, and racing thoughts than a White at the same level of mania.

RACE: Slope

An interaction term was entered into the model, testing whether there was a difference in the slope linking total score to predicted item score. A significant interaction term indicates that the validity of the item as an indicator of the construct is significantly different across the two groups. As seen in Tables 7, 8,

and 10, no DIF was detected on three out of four scales indicating that the construct is not significantly different between African Americans and Whites. DIF was detected on item 13 (**hypersexuality**) ($p < .05$) of the P-CMRS (table 9), indicating that this item is not equally related to mania between the two groups. The hypersexuality item is more strongly related and does a better job discriminating mania in the Whites than in the African Americans group.

RACE: Nonuniform DIF

The interaction (nonuniform DIF) of the group by the level of mania, is detected if persons from 1 group endorse symptoms more or less often than persons in another group (with an equal level of mania). Groups were matched for mania prior to examining the group effect.

DIF was not detected on any items of the P-YMRS (table 7) or the P-GBI (table 8) between the African American and White groups. As seen in table 9, DIF was detected on 4 items of the P-CMRS between the African American and White groups (with equal levels of mania). The results indicate that the African American group endorsed symptoms *more often* on: 1) item 1 (**super happy**) compared to the White group by 0.02 points ($p = 0.021$), and 2) on item 21 (**visual hallucinations**) by 0.01 points ($p = 0.015$). In contrast, the AA group endorsed symptoms *less often*: 3) on item 4 (**unrealistic abilities or powers**) compared to the White group by 0.01 points ($p = 0.023$); and 4) on item 13 (**hypersexuality**) by 0.02 points ($p = 0.004$).

DIF was detected on item 8 (**more energetic**) of the P-MDQ (table 10) between the African American and White groups (with equal levels of mania). As

seen in table 10, this indicates that the odds of answering yes to this item was *1.3 times higher* ($p=0.033$) in the White group than in the African American group.

Given an equal level of mania, African Americans report feeling super happy and having visual hallucinations more often than Whites do and Whites report having increased energy, unrealistic abilities or powers and hypersexuality more often than African Americans do.

Age Uniform DIF

A dummy code was entered for age (ages 5 to 10 years=0, and 11 to 18 years =1): therefore the regression equation is: Score on item = Age dummy code + error.

DIF was detected on 3 items of the P-YMRS (Table 11), indicating that when all else is held constant, children will score 0.3 points ($p<.05$) higher on item 2 (**increased motor activity/energy**); and 0.2 points ($p<.05$) lower on item 4 (**sleep**) and 0.3 points lower ($p<.05$) on item 5 (**irritability**) than an adolescent with the same level of mania.

DIF was detected on 3 items of the P-GBI (table 12) indicating that when all else is held constant, children will score 0.2 points higher ($p<.05$) on item 4 (**unusually happy/increased motor activity/energetic**), 0.2 points lower ($p<.05$) on item 54 (**unusually happy/energetic/irritable/angry**) and 0.2 points higher ($p<.05$) on item 64 (**racing thoughts**) than an adolescent with the same level of mania.

DIF was detected on 6 items on the P-CMRS (table 13) indicating that when all else is held constant, children will score 0.3 points lower ($p<.05$) on item 2

(irritable/cranky/mad); 0.3 points *higher* ($p < .05$) on item 6 **(too much energy)**; 0.3 points *higher* ($p < .05$) on item 7 **(talks too much/loud or fast)**; 0.5 points *higher* ($p < .05$) on item 11 **(off track)**; and 0.3 points *lower* ($p < .05$) on item 18 **(mood swings)** and 0.3 points *lower* ($p < .05$) on item 19 **(suspicious)** than an adolescent with the same level of mania.

DIF was detected on 4 items of the P-MDQ (table 14) indicating that when all else is held constant, the odds of parents of children answering yes on item 7 **(easily distracted/trouble concentrating)** was 0.4 times *higher* ($p < .05$) than that of adolescents; and the odds of parents of adolescents answering yes was 4.2 times *higher* ($p < .05$) on item 11 **(more interested in sex)**; 2.2 times *higher* ($p < .05$) on item 12 **(unusual/foolish/risky)** and 7.5 times *higher* ($p < .05$) on item 13 **(spending more money)** than that of children.

For reasons other than differences in their level of mania, the parents of children report higher rates of:

- talking too loud or fast
- being off track
- easily distracted/trouble concentrating
- having racing thoughts
- experiencing increased motor activity and more energy
- feeling unusually happy or energetic, except in the context of irritable/angry per item 54 on the P-GBI.

For reasons other than differences in their level of mania, the parents of adolescents report higher rates of:

- decreased need for sleep
- feeling irritable, or unusually happy/energetic with irritability and angry mood
- feeling irritable/cranky/mad
- having mood swings
- feeling suspicious
- being easily distracted
- engaging in foolish risky behaviors
- spending money
- having more interest in sex

Age Slope

DIF was detected on item 3 (**sexual interest**) ($p < .05$); on the P-YMRS (table 11) indicating that this item is not equally related to mania between the two groups. The item is more strongly related and does a better job discriminating mania in the adolescents than in children.

There was no DIF detected in the slope for any items on the P-GBI (table 12), which indicates that the validity of the item as an indicator of the construct (mania) is not significantly different across the two groups.

DIF was detected on item 13 (**hypersexuality**) ($p < .05$); of the P-CMRS (table 13) indicating that this item is not equally related to mania between the two

groups. The item is more strongly related and does a better job discriminating mania in adolescents than in children.

DIF was detected on item 13 (**hypersexuality**) of the P-MDQ (table 14) indicating that this item is not equally related to mania between the two groups. The odds of answering yes to this is *1.4 times higher* ($p<.05$); in adolescents than in children.

Overall, the items related to hypersexuality (or sexual interest) do a better job discriminating mania in adolescents on most measures, with the exception of the P-GBI.

Age Nonuniform DIF

DIF was detected on 2 items of the P-YMRS (table 11) indicating that parents of children report symptoms *more often* than parents of adolescents, adding 0.03 points ($p<.05$) on item 2 (**increased motor activity/energy**) and 0.03 points ($p<.05$) on item 7 (**changes in thought pattern**).

As seen in table 12, DIF was not detected on any items of the P-GBI between the younger and older groups.

DIF was detected on 5 items of the P-CMRS (table 13) indicating that parents of children report symptoms *more often*, adding 0.01 points ($p<.05$) on item 10 (**rushing around/ nonstop**); and 0.01 points ($p<.05$) on item 12 (**productive/creative**); and *less often*, dropping 0.02 points ($p<.05$) on item 11 (**off track**); 0.01 points ($p<.05$) on item 13 (**sexually inappropriate behavior**); and 0.01 points ($p<.05$) on item 20 (**auditory hallucinations**), than parents of adolescents.

DIF was detected on 2 items of the P-MDQ (table 14) between children and adolescents (with equal levels of mania), indicating that the odds of an adolescent answering no was 0.8 times *higher* ($p < .05$) on item 9 (**more active**) and 0.9 times *higher* ($p < .05$) on item 10 (**more social/outgoing**) than that of children.

Given an equal level of mania, the younger group report increased motor activity, changes in thought patterns, rushing around doing things nonstop, and being more productive and creative more often than the older group. Younger subjects are 0.8 times more likely to report being more active, social and outgoing than the older subjects.

Gender Uniform DIF

A dummy code was entered for gender (male=0, and female=1): therefore the regression equation is: Score on item = Age dummy code + error.

DIF was detected on 1 item of the P-YMRS (table 15) indicating that when all else is held constant, males will score 0.4 points *higher* ($p < .05$) on item 9 (**disruptive/aggressive behavior**) than a female at the same level of mania.

As seen in table 16, DIF was not detected on any items of the P-GBI between the male and female groups.

DIF was detected on 1 item of the P-CMRS (table 17) indicating that when all else is held constant, males will score 0.3 points *higher* ($p < .05$) on item 11 (**off track**) than a female at the same level of mania.

DIF was detected on 2 items of the P-MDQ (table 18) indicating that when all else is held constant, the odds of females answering no on item 7 (**easily**

distracted/trouble concentrating) was 0.6 times *higher* ($p < .05$); and the odds of answering yes on item 10 (**more social/outgoing**) was 3.1 times *higher* ($p < .05$) than that of males.

For reasons other than differences in their level of mania, males reported higher rates of disruptive/aggressive behavior, being off track and distracted with more trouble focusing than females, and females were 3.1 times more likely than males to feel more social and outgoing.

Gender Slope

Dif was detected on item 3 (**sexual interest**) ($p < .05$) of the P-YMRS (table 15) indicating that this item is not equally related to mania between the two groups. The item is more strongly related and does a better job discriminating mania in females than in males.

There was no DIF detected in the slope for any items of the P-GBI (table 16), the P-CMRS (table 17), or the P-MDQ (Table 18), which indicates that the validity of the item as an indicator of the construct (mania) is not significantly different across the two groups.

Gender Nonuniform DIF

DIF was detected on 1 item of the P-YMRS (table 15) between the male and females (with equal levels of mania). Males endorsed symptoms: 1) *less often* on item 3 (**sexual interest**) compared to females which increased by 0.02 points ($p < .05$).

DIF was not detected on any items of the P-GBI (table 16) between genders.

DIF was detected on 1 item of the P-CMRS (table 17) between males and females (with equal levels of mania), indicating the males endorsed symptoms: 1) *less often* on item 20 (**auditory hallucinations**) which increased by 0.01 points ($p < .05$).

DIF was detected on item 9 of the P-MDQ (table 18) between groups (with equal levels of mania), indicating that the odds of those in the female group: 1) answering no on item 9 (**more active**) was 0.8 times *higher* ($p < .05$) than those in the male group.

Given equal levels of mania, males report symptoms of sexual interest and hallucinations less often than females, and symptoms of being more active more often than females.

CHAPTER 5

DISCUSSION

This study was undertaken to assess the psychometric properties of several psychiatric screening instruments for urban youth with bipolar disorder and evaluate potential racial, gender and age differences using Differential Item Functioning (DIF). The instruments included the Young Mania Rating Scale – Parent Version (P-YMRS), General Behavior Inventory – Parent Version (P-GBI), Child Mania Rating Scale-Parent Version (P-CMRS), Mood Disorders Questionnaire-Parent Version (P-MDQ) and the Child Behavior Checklist (CBCL).

Race

The instrument which revealed the greatest statistically significant differences between African Americans and Whites, using DIF, was the P-CMRS. Parents of AA endorsed greater scores on abilities/powers and parents of Whites endorsed greater scores (a total of 0.7 points) on the cranky/irritable or mad and the rapid mood swings items. Although very modest differences were detected, (a total of 0.06 points) emerged on other items (e.g. the super happy, unrealistic abilities or powers, hypersexuality, and the visual hallucinations items) of the P-CMRS, these are unlikely to be clinically meaningful.

DIF was also detected on the racing thoughts item and the more energetic item of the P-MDQ. The odds of a parent of a White child answering yes to these items were 3.3 and 1.3 times higher, respectively, than an African American. Using the P-YMRS, parents of African American youth endorsed greater symptoms on increased activity/energy. Using the P-GBI, African American

parents reported that their child had more symptoms of being unusually happy/energetic, while White parents reported more rapid mood shifts.

While statistically significant, the DIF detected resulted in a total difference of 0.4 points in the score on the P-YMRS, of 0.6 points on the P-GBI, of 0.96 on the P-CMRS and 4.6 times higher odds of answering yes to 2 items on the P-MDQ, which is not likely to be clinically meaningful or to influence diagnostic decisions made by clinicians or researchers.

DIF against African Americans in this study indicates that they are more likely to endorse certain symptoms of mania than Whites. Overall, the AA parents reported higher rates of increased activity and energy, feeling unusually happy or energetic, having special abilities or powers and more visual hallucinations. Taken together, these symptoms suggest that AA youth may have a different presentation of mania, with more grandiosity and psychosis. These findings are consistent with previous studies of AA adolescents who were much more likely to present or be identified with psychotic features and have higher ratings for auditory hallucinations during their first psychiatric hospitalization than White adolescents were (Patel, et al., 2006).

The difference in the presentation of symptoms found in this study between AA and White youth may be one of the underlying factors contributing to differences in diagnosis found in previous studies of AA and White youth where: 1) AA youth were more likely to receive a diagnosis of psychosis and/or a behavioral disorder, and to be hospitalized and less likely to receive a diagnosis of bipolar disorder, major depression disorder and/or a substance abuse disorder than White

youth were (Muroff, et al., 2008); and 2) AA adolescents with PBD were nearly twice as likely to receive treatment with an antipsychotic agent as White adolescents were (DeBello, et al., 2000).

Conversely, Whites reported higher rates of rapid mood shifts, irritable/cranky moods, racing thoughts and hypersexuality. These findings are consistent with the atypical presentation of PBD in youth described by researchers as presenting with severe impairment and mood disturbances that are characterized by irritability and aggression (Biederman, et al., 2000; Carlson, et al., 2002; Geller, Craney, et al., 2002; Geller, Zimmerman, et al., 2002; Pavuluri, et al., 2005; Wozniak, et al., 1995), and having less symptom-free periods, briefer episodes and higher rates of irritability (Findling, et al., 2001; Geller, Zimmerman, et al., 2002).

There is limited data examining differences in parental report of symptoms which could account for the differences found in this study between African American and White children and adolescents. More research is needed to understand the underlying issues responsible for these differences.

While the results suggest that the differences between groups could lead to an overestimation of mania in AA youth, the total point difference on each scale is unlikely to be clinically meaningful or to influence diagnostic decisions made by clinicians or researchers. These instruments, which rely on parental report may in fact, help to reduce disparities in diagnosis due to clinician bias and perception. Overall, these instruments appear to be functioning well between AA and White youth in detecting PBD symptoms.

AGE

Using DIF, younger children (less than ten years old) in this study were more likely to demonstrate certain symptoms of mania than adolescents, such as talking loud or too fast, having racing thoughts, changes in thought patterns, being off-track, easily distracted and having trouble concentrating, rushing around nonstop, being more active, productive or creative, more social, experiencing increased activity and more energy, and feeling unusually happy, and energetic (except in the context of irritability and anger).

Specifically, parents of younger children endorsed more items related to increased motor activity/energy on the P-YMRS, the P-GBI, the P-CMRS and the P-MDQ. Younger subjects were more likely to be off track, talking louder or faster, rushing around non-stop and being more productive/creative on the P-CMRS and the odds of parents answering yes to the child being easily distracted/having trouble concentrating were 0.4 times higher on the P-MDQ than the adolescents. Additionally, younger subjects demonstrated more changes in thought patterns on the P-YMRS, racing thoughts on the P-GBI and being more social/outgoing on the P-MDQ.

Taken together, these symptoms suggest that younger children may have a more similar presentation of mania to adults, than adolescents do, with an increase in energy and motor activity, euphoric moods, racing thoughts, distractibility, rapid and loud speech patterns, an increase in creativity and productivity and more social/outgoing behavior. However, because youth have briefer episodes, higher rates of irritability, less symptom-free periods, (Findling, et

al., 2001; Geller et al., 2002) and higher rates of comorbidity, (Findling et al., 2001; Geller et al., 2000a; Wozniak, Biederman, Monuteaux, Richards, & Faraone, 2002) their overall presentation of mania differs greatly from that of adults.

Parents of adolescents reported greater symptoms on sleep patterns and irritability on the P-YMRS, increased happy/energetic/irritability/anger on the P-GBI, and more mood swings, being irritable/cranky/mad, suspiciousness, having an increased interest in sex, being off-track and having auditory hallucinations on the P-CMRS. The total odds of an a parent of an adolescent answering yes to 3 items on the P-MDQ (more interested in sex, unusual foolish/risky behavior, and spending more money), unrelated to mania, was 15 times higher than that of younger children. The hypersexuality item was not equally related to mania between children and adolescents on the P-YMRS, P-CMRS or the P-MDQ. It consistently did a better job discriminating mania in adolescents than in children using these instruments.

Other than the auditory hallucinations, the increase in symptoms reported by parents of adolescents are consistent with the expected changes experienced during this stage of development. Hormone fluctuations in adolescence contribute to physiologically-driven changes in energy levels, sleep patterns, mood and sexual interest (Gemelli, 1996). Severity of symptoms, in the context of the overall clinical picture, can be helpful in distinguishing between symptoms of psychopathology and the expected physiologically-driven changes in adolescence.

Previous studies of mania have documented psychotic symptoms, labile moods, and/or mixed manic and depressive features in adolescents (Pavuluri, et

al., 2005). In comparison to children, adolescents with bipolar disorder have higher rates of mood symptoms and lower quality of life scores (Freeman et al., 2009). An inpatient chart review comparing manic patients under the age of 21 (N=9) with those over the age of 30 (N=12) reported that adolescents had more psychotic symptoms, including delusions and ideas of references than adults did and that their atypical presentation may increase the risk of being misdiagnosed with schizophrenia (Ballenger, Reus & Post, 1982).

An outpatient study of youth with Bipolar I disorder, divided into 2 groups of 6-9 years of age and 10-16 years of age documented rates of psychosis similar to that found in adults. The lifetime prevalence of delusions and pathological hallucinations was 76.3% (Tillman et al., 2008). The most common type of delusion reported was grandiosity and the most common type of hallucination was visual. There were no differences found between the two age groups in this study. Previous studies of adolescents with Bipolar I or Bipolar II disorder versus Bipolar disorder NOS have reported increased rates of inflated self esteem, decreased need for sleep and increased goal directed activity in the BP I and BP II groups (Lewinsohn et al., 1995).

The range of comorbidity and the quality of presenting symptoms can vary greatly according to age and duration of illness (Post & Kowatch, 2006). It is important to note that the majority of children and adolescents with PBD do not meet the full DSM-IV-TR category for Bipolar I or Bipolar II disorder and it can be difficult to compare findings between samples of different age and diagnostic groups. Subthreshold forms of bipolar disorder are subsumed under the category

of “bipolar spectrum disorders” and a diagnosis of bipolar disorder not otherwise specified (NOS) is used. Only 3% of children and 4% of adolescents in the current study met the diagnostic criteria for Bipolar I disorder. The remainder of children and adolescents with bipolar disorder in this study met the diagnostic criteria for either bipolar II disorder, bipolar disorder NOS or cyclothymia.

A meta-analysis of studies of PBD youth, 5 to 18 years of age, suggested that younger age and parental report contribute to higher rates of distractibility, irritability, poor judgment and aggression (Kowatch, et al., 2005). The most commonly reported symptoms in this meta-analysis were increased energy, distractibility and pressured speech. Four out of five cases reported significant irritability and grandiosity, greater than 70% elated or euphoric mood, decreased need for sleep or racing thoughts, 69% poor judgment, 50% flight of ideas and slightly greater than one third of the subjects reported hypersexuality (Kowatch, et al., 2005). Many of the differences in descriptions of pediatric bipolar phenomenology found in this study were attributable to methodological differences between studies.

Previous studies of youth with PBD suggest that chronic mania in youth is accompanied by severe impairment with mood disturbances characterized by irritability and aggressiveness (Biederman, et al., 2000; Carlson, et al., 2002; Geller, Craney, et al., 2002; Geller, Zimmerman, et al., 2002; Pavuluri, et al., 2005; Wozniak, et al., 1995). However, irritability and aggression, some of the most impairing features of mania, are not specific to mania. Irritability can be a feature

of depression, generalized anxiety disorder, ODD, or PTSD (Kowatch, et al., 2005).

The most controversial symptoms of mania in children and adolescents are elation and grandiosity. Elated or euphoric mood and grandiosity are considered the cardinal features of mania in youth by some researchers (Craney and Geller, 2003). However, grandiosity and excessive involvement in pleasurable activities can vary as a function of age and development (Bowring & Kovacs, 1992; Geller & Luby, 1997). It is important to consider each symptom in the context of the larger clinical picture as no single symptom of mania is itself diagnostic. The total clinical picture, rather than the symptoms alone, is what predicts a diagnosis of PBD (Kowatch, et al., 2005).

DIF against children in this study indicates that parents of children are more or less likely to endorse specific symptoms of mania than parents of adolescents, which can lead to an overestimation or underestimation of manic symptoms as noted above for children relative to adolescents using these instruments.

While statistically significant, a total difference of 0.86 points in the score on the P-YMRS and of 0.6 points on the P-GBI is unlikely to be clinically meaningful or influence diagnostic decisions made by clinicians or researchers. The total difference of 2 points on the P-CMRS is not likely to be clinically meaningful in practice. However, a 2 point difference in the score of an instrument used in a research study with stringent cut off criteria for diagnosis could affect diagnostic and severity decisions and influence the results of a research study. Unrelated to mania, the total odds (15 times higher) of a parent of an adolescent

answering yes on 3 items of the P-MDQ (more interested in sex, unusual foolish/risky behavior and spending more money) than parents of younger children is likely to be clinically meaningful and influence diagnostic decisions in practice and research studies.

In the context of normal development, we expect children to behave and interact differently than adolescents do. Inflated self appraisals in young children, a part of normative development, is not considered to be indicative of mania as it might be in an adolescent or adult. An increase in sexual interest is associated with normal adolescent development, while it is indicative of psychopathology in a child. Changes in sleep patterns are normal in adolescents as they undergo sleep phase delay, with the typical time to fall asleep being 11pm or later due to a change in their internal clocks (Wolfson & Carskadon, 2001).

Children cannot spend money that they do not have, which may account for the increase reports of spending in adolescents, when compared to children. The developmental equivalent of spending in children might be better described by children perseverating on buying toys and items, hounding their parents incessantly and having meltdowns when their parents won't take them to the store. To better capture spending in children, these items might be reworded in a more developmentally-sensitive manner.

Adolescence is a period of turbulent emotions and mood swings, combined with rapid intellectual development (Viner & Christie, 2005), which may contribute to the differences in mood symptoms between the children and adolescents in this study. Erickson characterizes adolescence as a period in which "identity" is the

primary psychosocial crisis (Erickson, 1998). Identity development is critical to how the adolescent perceives oneself and to how they function in social interactions (Garcia, 2010). Critical thinking and information processing develop during adolescence. Older adolescents have greater abilities to remember and to reason than younger adolescents do (Garcia, 2010), which may affect symptoms between the younger and older adolescents within this age group. The adolescent's quest for emotional autonomy from parents and their need for connectedness run parallel (Yeh & Yang, 2006) complicating the parent/child relationship which may affect exchange of information and ultimately the parental report.

In the context of normal development, some of the differences identified in this study between the children and adolescents, unrelated to mania, are expected. However, in the absence of age-specific diagnostic criteria, developmentally-appropriate wording on the items of screening instruments and literature which distinguishes between symptom expression in two very different stages of development, it is difficult to truly understand what is underlying these differences. Future research studies of DIF might be improved by tailoring the wording of the items in screening instruments to the developmental stage or chronological age of the child or adolescent.

GENDER

Using DIF, parents of males in this study were more likely to report certain symptoms of mania, such as disruptive/aggressive behavior, being off track, distracted and having trouble concentrating, and being more active than parents of

females. Conversely, parents of females were more likely to report symptoms of being more social or outgoing, having more sexual interest and experiencing more hallucinations than parents of males.

Specifically, parents of males endorsed more items related to disruptive/aggressive behavior on the P-YMRS, being off track on the P-CMRS and being easily distracted/trouble concentrating and being more active on the P-MDQ. Parents of females were more likely to report symptoms of hypersexuality on the P-YMRS, and of auditory hallucinations on the P-CMRS, than parents of males. The odds of parents answering yes to females being more social/outgoing on the P-MDQ were 3.1 times higher than parents of males.

DIF against males in this study indicated that parents of males are more or less likely to endorse a symptom of mania than parents of females, which can lead to an overestimation or underestimation of mania for males relative to females using these instruments. While statistically significant, the DIF detected resulted in a total difference of 0.4 points in the score on the P-YMRS, of 0.3 on the P-CMRS, and a combined total of up to 5 times the odds of answering differently to 3 items on the P-MDQ. These differences are not likely to be clinically meaningful or to influence diagnostic decisions made by clinicians or researchers. The scales appear to be functioning well between males and females, despite the minor differences reported between groups.

While rates of bipolar disorder are equal among males and females later in life, early onset cases, especially prior to the age of 13 years of age, are predominantly male (McClellan, et al., 2007). Symptoms manifest themselves

differently between boys and girls. Prior to puberty, boys and girls have equal rates of depression. However, during adolescence, the rate of depression among girls doubles (Coyle et al., 2003). Boys present with more manic symptoms and girls present with more depressive symptoms and lower quality of life scores than boys with bipolar disorder do (Freeman, et al., 2009). This is an important consideration, given that by the age of 25, females with BD who have gone untreated are estimated to lose 14 years of effective functioning in work and school and with family, 12 years of normal health, and have a life expectancy shortened by 9 years, compared to healthy females (DHEW, 1979).

Findings from the current study suggest that unrelated to mania, boys are more active, disruptive, aggressive, distracted, off-track and have more trouble focusing than girls. Findings related to gender differences in aggression (Burton, Hafetz, & Henninger, 2007), socialization and sexuality (Metts, Sprecher, & Regan, 1998) are logical and well documented in the literature. In general, research studies have documented that males are more aggressive than females (Coie & Dodge 1997, Maccoby & Jacklin 1974) and given gender stereotypes, aggression is generally considered more tolerable and socially acceptable in males versus females (Eisenberg et al., 1993; 1995; Gjerde, 1995). Findings from previous gender studies suggest that more attentional control, negative, acting out behaviors and emotionality are associated with better social functioning in males than in females (Eisenberg et al., 1995). Findings from the current study suggest that by parent-report males are more off track, distractible and have trouble focusing, which would predict worse social functioning according to previous

research findings. Females in this study were found to be more social and outgoing with is consistent with the literature on females being more likely to invest time in talking, establishing emotional connections, and forming friendships characterized by physical closeness, laughter, and reciprocal loyalty and support (Frosh et al., 2002).

Males had higher rates of activity in the current study which is consistent with gender studies. Relative to females, males are perceived as higher in exhibited activity levels (Walker et al., 2001) and lower in attentional skills (Eisenberg et al., 1997; Eisenberg et al., 2003; Murphy et al., 2004; Walker et al., 2001); and during adolescence, girls are more inactive and have a steeper decline in physical activity than boys do (Caspersen, Periera & Curran, 2000).

An increase in sexual interest among females in the current study was a surprising finding. However, given that males are more likely to express their sexual interest and less likely to be inhibited by social norms in expressing their desires than females are (Metts, Sprecher, & Regan, 1998), parents may view their son's expression of sexual interest as normal and their daughter's expression of sexual interest as pathological.

There is no literature to support an increase in reports of auditory hallucinations in females. However, societal expectations may play an important role in the report and recognition of symptoms across the board between genders. A study of youth, ages 10-15, examined differences in the expression and report of symptoms between boys and girls found that, generally, girls have an easier time reporting symptoms and seeking help for illness than boys do. Younger

participants reported that pressures on boys to behave in masculine or “tough” ways increase with age and that psychological symptoms such as crying were not acceptable and outlined most explicitly and emphatically as against the rules. The main difference was identified in the extent to which boys and girls’ social and gender identities are threatened by the act of seeking help. Boys are fearful that reporting symptoms might directly challenge the most valued aspect of their identity: their masculinity (MacLean, Sweeting & Hunt, 2010). With age, boys are less likely to report symptoms due to fear that it could have a detrimental impact on their perceived masculinity. Participants of both genders revealed that stereotypical masculine responses of stoicism, independence, control and strength are expected from boys in response to illness. Boys report conforming to masculine ideals by attempting to conceal, disguise, ignore and overcome symptoms on their own rather than seeking help which could impact self-report and parent-report of symptoms (MacLean, Sweeting & Hunt, 2010).

According to the Gateway Provider Model (GPM), the key role of the gateway provider is to initiate and direct treatment of youth. Educating primary care providers and providing them with education, information and valid screening instruments is important to facilitate the recognition of pathological symptoms which may ultimately increase timely diagnosis and treatment of PBD in youth regardless of race, age or gender. The assessment for DIF in the most-commonly used screening instruments for PBD in this study revealed statistically significant differences between groups in race, gender and age, as outlined below. However, with a few exceptions, the differences identified in this study are unlikely to be

clinically meaningful and influence diagnostic decisions in practice and research studies. Overall, the instruments are functioning well between different racial and age groups and between genders and may be valuable tools for gateway providers, clinicians and researchers in screening for PBD.

CONCLUSION

Measurement invariance of 5 commonly-used screening instruments was examined using differential item functioning (DIF). Measurement invariance indicates that the parameters characterizing an item (discrimination and difficulty) are comparable across groups (Raju, Laffitte, & Byrne, 2002; Reise, Widaman, & Pugh, 1993). The DIF methods (Zumbo, 2007) used in this study allowed the researcher to evaluate the items and ultimately the instrument they constitute, to determine if they were functioning in the same manner between different racial and age groups and between genders.

As noted in the results section, measurement variance was found in the instruments between the groups indicating that the parameters, characterizing the items, where DIF was detected, are not comparable across groups. What this means, is that the probability of endorsing an item differs between equal groups due to group membership, indicating a lack of measurement invariance (Messick, 1989, 1995).

The assessment for DIF using the P-YMRS, P-GBI, P-CMRS, P-MDQ and CBCL between different racial and age groups and between genders in a sample of urban youth revealed statistically significant differences between groups.

All of the instruments, except for the CBCL, purport to measure the same or similar constructs of mania. The CBCL, a screening instrument used in multiple PBD research studies and recommended by the National Institute of Mental Health Roundtable on PBD as a method to benchmark similarities across samples (Nottelmann, 2001), does not contain a mania or hypomania scale nor does it

include items based on the DSM-IV-TR diagnostic criteria of mania (Youngstrom et al., 2005) which may limit its diagnostic efficiency in PBD.

Previous research studies have reported that the CBCL is helpful in identifying children at high risk to develop bipolar disorder. Some studies suggest that children with a deviant profile on the CBCL's attention problems, aggressive behavior, and anxious-depressed subscales are likely to meet the criteria for DSM bipolar I disorder in epidemiological and clinical samples (Achenbach, 1991; Hazell, Lewin, & Carr, 1999; Mick, Biederman, Pandina, & Faraone, 2003) and that the CBCL may provide markers for psychopathology in youth (Biederman, Wozniak, Kiely, & Ablon, 1995). Other studies reported that the CBCL statistically discriminated between PBD and other psychiatric disorders (Biederman, et al., 1995; Carlson, 1998; Carlson & Kelly, 1998; Dienes, Chang, Blasey, Adleman, & Steiner, 2002; Faraone, Althoff, Hudziak, Monuteaux, & Biederman, 2005; Geller, Warner, Williams, & Zimmerman, 1998; Hazell, et al., 1999).

However, the findings from this study suggest that the diagnostic efficiency of the CBCL is questionable. Previous studies revealed that the sensitivity of the CBCL in identifying mania, is significantly lower than that of other instruments, which are specific to bipolar disorder (Pavuluri, Henry, Devineni, Carbray, & Birmaher, 2006; Youngstrom, et al., 2004) such as the P-YMRS, P-GBI, P-CMRS and the P-MDQ. The CBCL performed poorly as an indicator of PBD in a sample from an urban community mental health center with high rates of comorbidity. Youth with ADHD and ODD generated similar scores on the CBCL as youth with PBD in this sample (Youngstrom, et al., 2005). Elevated scores on the CBCL

were reported in both the bipolar and non-bipolar groups suggesting that the elevation on this scale is not specific to mania. The CBCL does poorly discriminating clinical diagnosis of PBD due to pervasive externalizing behavior problems in some samples and the variable presentation of PBD in youth (Youngstrom, Youngstrom, & Starr, 2005).

While children with PBD display higher scores on many of the informant scales, the ASEBA behavior checklists do not possess strong diagnostic efficiency statistics in discrimination PBD from non-PBD individuals (Kahana, Youngstrom, Findling, & Calabrese, 2003). A long term outcome study of youth characterized with severe aggression, inattention and mood instability, coined as the CBCL-PBD phenotype in this study, suggested that the CBCL may be a better tool for identifying children with severe functional impairment and broad-ranging psychiatric comorbidities, rather than bipolar disorder itself. The participants in this study exhibited marked social impairment, increased rates of suicidal thoughts and behaviors and a heightened risk for comorbid anxiety, bipolar disorder, cluster B personality disorders and ADHD in young adulthood, compared to participants without this presentation. However, diagnostic accuracy for any one particular disorder was found to be low using the CBCL (Meyer et al., 2009). Elevations on specific behavior checklists may be more indicative of greater symptoms severity, functional impairment, and overall psychopathology or comorbid conditions associated with PBD, rather than providing specific diagnostic information about PBD (Kahana, et al., 2003).

CBCL

EFA and CFA were used to provide validity evidence concerning the structure of the instruments used in this study of urban youth. Unidimensionality was established for all of the screening instruments except for the CBCL. Using EFA, the CBCL produced more than 1 factor underlying the items, 8 factors according to the Scree plot, and 3 factors using a more stringent cutoff score of eigenvalues > 2 . Both methods indicate that the CBCL is multidimensional, rather than unidimensional. Therefore, it could not be included in further analyses in this study. The findings from this study are consistent with previous studies suggesting that variations in scores on the CBCL may be due to factors unrelated to mania which limit its usefulness in assessing PBD in research studies and in clinical settings.

P-YMRS

The P-YMRS, previously validated by Gracious and colleagues (Gracious, et al., 2002), has reasonable psychometric properties but includes several items with poor factor loadings. Axelson and colleagues suggest that the item content may not be developmentally appropriate for children particularly in the area of insight and appearance. Questions have been raised on psychometric grounds about the inclusion of item 11 by researchers because it is much less related to the construct of mania than the other items in the set are (Gracious 2002; Youngstrom et al., 2002). Consistent with previous findings, a factor loading of 0.08 indicates that item 11 (lack of insight) is not correlated with the other items in the set and is a weak measure of mania. Therefore, it was excluded from further

analyses, leaving only 10 items on the P-YMRS. After item 11 was excluded, correlations among the remaining 10 items reflected a single underlying dimension with a total score showing: acceptable internal consistency ($\alpha = .75$) replicating previous studies (Gracious 2002; Youngstrom et al., 2002).

Other Instruments

Good internal consistency emerged for the 13 items on the P-MDQ ($\alpha = .83$); the ten items on the P-GBI ($\alpha = .88$); and the 21 items on the P-CMRS ($\alpha = .90$). The assumption of a unidimensional latent trait (mania) showed to be sufficient to explain the actual response of the respondents on these scales. Findings from previous research studies of the P-YMRS, P-GBI and the P-MDQ between African Americans and European-Americans revealed no statistically significant differences. However, the reliability of these instruments was lower in the sample from the urban community mental health clinic than in the sample from the academic center clinic (Youngstrom et al., 2005).

Suggestions for Future Research

These data reveal statistically significant DIF meaning that the probability of endorsing an item, that exhibits DIF, differs between groups due to group membership (race, age or sex) indicating a lack of measurement invariance. The items exhibiting DIF are not functioning in the same manner between two groups.

When an item shows DIF, test developers need to decide whether this item should be retained or eliminated. The results of this study do not provide sufficient evidence to distinguish between problems in the wording of the question and an underlying, clinically-meaningful variation in the manifestation of mania between

groups. Statistical procedures that are beyond the scope of this study can be used in future research studies to determine if changes in the items are needed. More research is needed to decipher the instruments at a level so that IRT and CFA can be used to cleanly test this hypothesis in the future.

Limitations

Item Response Theory (IRT) was used to guide these analyses. IRT assumes unidimensionality but because unidimensionality and local independence cannot be strictly met, sets of items with one dominant dimension, in the presence of one or more minor dimensions is typical. A set of items may be unidimensional but still contain pairs of items that are correlated. Local independence requires that any two items be uncorrelated if the latent variable is fixed (Lord, 1980). All of the measures showed some local dependence, meaning that a secondary non-essential factor (e.g., a hyperactivity factor in the mood items) is present. DIF is the process of looking for extra factors influencing response patterns. Local dependence could falsely cover true DIF and falsely indicate DIF exists when it does not. Therefore, regression techniques, rather than CFA, using IRT were used to assess for DIF in the data.

The characteristics of this sample may have influenced the results. This study was conducted in urban youth, primarily from families with low SES, which may limit its generalizability to other populations. Comorbidity, which is common in PBD, may also vary in different samples contributing to unique differences between groups. The wide variety of levels of development and maturity inherent

to age, both within and between the groups (5-10 years old and 11-18 years old), may also be a confounding factor in these data.

The design of this study, a secondary analysis of previously collected data is not as strong as a prospective sample specifically collected to examine DIF and may be a limitation. This is the first known investigation of DIF between racial and age groups and between genders in PBD sample. Further research using advanced methodology is needed to better understand how these instruments function between different groups.

Clinical Implications

According to the Standards for Education and Psychological Testing, evidence to support that test scores measure a latent trait similarly across populations subgroups such as different racial, and age groups and between genders is needed (American Educational Research Association, et al., 1999). A lack of measurement invariance suggests systematic measurement bias, indicating that observed scores do not represent the measured trait equally for different groups, which can result in unintended social consequences (Messick, 1989,1995) for individuals and groups.

Observed scores from screening instruments are used by researchers and clinicians to assess PBD and to aide in clinical diagnosis, as well as to assess for improvement in PBD symptoms in clinical trials of medications so it is crucial that they are measuring what they are supposed to. Observed score differences should be interpreted cautiously unless measurement model parameters are invariant (similar) across compared groups (Borsboom, 2006). Using different

measures in heterogenous samples of PBD can produce results that are difficult to compare and may not be generalizable across studies and in different populations.

REFERENCES

- Achenbach, T. M. (1991). *Manual for the child behavior checklist/4-18 and profile*. Burlington: University of Vermont, Department of Psychiatry.
- Achenbach, T. M., & Rescorla, L. A. (2001). *Manual for the ASEBA School-Age Forms & Profiles*. Burlington, VT: University of Vermont, Research Center for Children, Youth, & Families.
- Althoff, R. R., Rettew, D. C., Faraone, S. V., Boomsma, D. I., & Hudziak, J. J. (2006). Latent class analysis shows strong heritability of the child behavior checklist-juvenile bipolar phenotype. *Biological Psychiatry*, *60*(9), 903-911.
- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington, DC, US: American Educational Research Association.
- American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders* (4th text revision ed.). Washington, DC: American Psychiatric Association.
- Anastasi, A., & Urbina, S. (1997). *Psychological Testing 7/E*. New Jersey: Prentice Hall.
- Andersen, R. (1995). Revisiting the behavioral model and access to medical care: Does it matter? *Journal of Health and Social Behavior*, *36*(1), 1-10.
- Andersen, R., & Aday, L. (1978). Access to Medical Care in the U.S.: Realized and Potential. *Medical Care*, *XVI*(7), 533-546.

- Angst, J. (2004). Bipolar disorder--a seriously underestimated health burden. *European Archives Of Psychiatry And Clinical Neuroscience*, 254(2), 59-60.
- Angst, J., & Cassano, G. (2005). The mood spectrum: improving the diagnosis of bipolar disorder. *Bipolar Disorders*, 7 Suppl 4, 4-12.
- Angst, J., Gamma, A., & Lewinsohn, P. (2002). The evolving epidemiology of bipolar disorder. *World Psychiatry: Official Journal Of The World Psychiatric Association (WPA)*, 1(3), 146-148.
- APA. (2000). *Diagnostic and statistical manual of mental disorders:DSM-IV-TR*. Washington: American Psychiatric Association
- Axelson, D., Birmaher, B., Brent, D., Wassick, S., Hoover, C., Bridge, J., et al. (2003). A preliminary study of the kiddie schedule for affective disorders and schizophrenia for school-age children mania rating scale for children and adolescents. *Journal of Child and Adolescent Psychopharmacology*, 13, 463-470.
- Ayalon, L., & Alvidrez, J. (2007). The experience of Black Consumers in the Mental Health System-Identifying Barriers to and Facilitators of Mental Health Treatment Using the Consumer's Perspective. *Issues in Mental Health Nursing*, 28, 1323-1340.
- Baker, F. M., & Bell, C. C. (1999). African American Treatment Concerns. *Psychiatric Serv.*, 50, 362-368.
- Ballenger, J., Reus, V. & Post, R. (1982). The atypical clinical picture of adolescent mania. *American Journal of Psychiatry*, 139, 602-606.

- Barnes, A. (2008). Race and hospital diagnoses of schizophrenia and mood disorders. *Social Work, 53*(1), 77-83.
- Barney, L. J., Griffiths, K. M., Jorm, A. F., & Christensen, H. (2006). Stigma about depression and its impact on help-seeking intentions. *The Australian And New Zealand Journal Of Psychiatry, 40*(1), 51-54.
- Bauer, M., Juckel, G., Correll, C., Leopold, K., & Pfennig, A. (2008). Diagnosis and treatment in the early illness phase of bipolar disorders. *European Archives Of Psychiatry And Clinical Neuroscience, 258*(5), 50-54.
- Bauer, M., & Pfennig, A. (2005). Epidemiology of bipolar disorders. *Epilepsia, 46*(suppl 4), 8-13.
- Berument, S. K., Rutter, M., Lord, C., Pickles, A., & Bailey, A. (1999). Autism screening questionnaire: Diagnostic validity. *British Journal of Psychiatry, 175*, 444-451.
- Biederman, J., Mick, E., Faraone, S., Spencer, T., Wilens, T., & Wozniak, J. (2000). Pediatric mania: A developmental subtype of bipolar disorder? . *Biological Psychiatry, 48*(6), 458-466.
- Biederman, J., Wozniak, J., Kiely, K., & Ablon, S. (1995). CBCL clinical scales discriminate prepubertal children with structured interview-derived diagnosis of mania from those with ADHD. *Journal of the American Academy of Child & Adolescent Psychiatry, 34*(4), 464-471.
- Birmaher, B., Arbelaez, C., & Brent, D. (2002). Course and outcome of child and adolescent major depressive disorder. *Child and Adolescent Psychiatric Clinics of North America, 11*(3), 619-638.

- Birmaher, B., Axelson, D., Monk, K., Kalas, C., Goldstein, B., Hickey, M. B., et al. (2009). Lifetime psychiatric disorders in school-aged offspring of parents with bipolar disorder: the Pittsburgh Bipolar Offspring study. *Archives Of General Psychiatry*, 66(3), 287-296.
- Birmaher, B., Axelson, D., Strober, M., Gill, M., Valeri, S., Chiapetta, L., et al. (2006). Clinical course of children and adolescents with bipolar spectrum disorders. [research]. *Archives of General Psychiatry*, 63(2), 175-183.
- Blader, J. C., & Carlson, G. (2007). Increased rates of bipolar disorder diagnosis in child, adolescent, and adult inpatients. *Biological Psychiatry*, 62(2), 107-114.
- Bollen, K., & Curran, P. (2005). *Latent Curve Models: A Structural Equation Perspective*. Hoboken: Wiley Interscience.
- Borsboom, D. (2006). When does measurement invariance matter? *Medical Care*, 44, S176-S181.
- Bossuyt, P., Reitsma, J., & Bruns, D. (1992). Towards complete and accurate reporting of studies of diagnostic accuracy: the STARD initiative. *Annals of Internal Medicine*, 138, 40-44.
- Bowring, M. A., & Kovacs, M. (1992). Difficulties in diagnosing manic disorders among children and adolescents. *Journal of American Academy of Child and Adolescent Psychiatry*, 31, 611-614.
- Brodie, R. E. (2004). Race, sex, and class bias in the diagnosis of dsm-iv disorders. *Dissertation Abstracts International*, 64(11B).

- Browne, M., & Cudeck, R. (1993). Alternative ways of assessing fit *Testing Structural Equation Models*. Newbury Park: Sage.
- Burton, L., Hafetz, J., & Henninger, D. (2007). Gender Differences in relational and physical aggression. *Social Behavior and Personality, 35*(1), 41-50.
- Carlson, G. (1995). Identifying prepubertal mania. *Journal of American Academy of Child and Adolescent Psychiatry, 34*(6), 750-754.
- Carlson, G. (1998). Mania and ADHD: Comorbidity or Confusion. *Journal of Affective Disorders, 51*, 177-187.
- Carlson, G. (2005). Early onset bipolar disorder: clinical and research considerations. *Journal Of Clinical Child And Adolescent Psychology: The Official Journal For The Society Of Clinical Child And Adolescent Psychology, American Psychological Association, Division 53, 34*, 333-343.
- Carlson, G., Bromer, E., Driessens, C., Mojtabai, R., & Schwartz, J. (2002). Age of Onset, childhood psychopathology, and 2-year outcome in psychotic bipolar disorder. *American Journal of Psychiatry, 159*, 307-309.
- Carlson, G., Findling, R., Post, R., Birmaher, B., Blumberg, H., Correll, C., et al. (2009). AACAP 2006 Research Forum--Advancing research in early-onset bipolar disorder: barriers and suggestions. *Journal Of Child And Adolescent Psychopharmacology, 19*(1), 3-12.
- Carlson, G., & Kelly, K. (1998). Manic symptoms in psychiatrically hospitalized children-what do they mean? *Journal of Affective Disorders, 51*, 123-135.
- Carter, T. D. (2003). Early age of onset as a risk factor for poor outcome of bipolar disorder. *Journal of Psychiatric Research, 37*, 297-303.

- Casper, R., Belanoff, J., & Offer, D. (1996). Gender Differences, But No Racial Group Differences, in Self-Reported Psychiatric Symptoms in Adolescents. *Journal of Child and Adolescent Psychiatry, 35*(4), 500-508.
- Caspersen, C., Periera, M., Curran, K. (2002). Changes in physical activity patterns in the United States, by sex and cross-sectional age. *Medical Science Sports Exercise, 32*, 1601-1609.
- Chang, K., Howe, M., Gallelli, K., & Miklowitz, D. (2006). Prevention of Pediatric Bipolar Disorder: Integration of neurobiological and psychosocial processes. *Annals of the New York Academy of Sciences, 1094*(1), 235-247.
- Chang, K., Steiner, H., & Ketter, T. (2000). Psychiatric phenomenology of child and adolescent bipolar offspring. *Journal Of The American Academy Of Child And Adolescent Psychiatry, 39*(4), 453-460.
- Chen, Y. W., & Dilsaver, S. C. (1996). Lifetime rates of suicide attempts among subjects with bipolar and unipolar disorders relative to subjects with other Axis I disorders. *Biological Psychiatry, 39*, 896-899.
- Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2nd ed.). Hillsdale: Lawrence Erlbaum.
- Coie, J., & Dodge, K. (1997). Aggression and antisocial behavior. In W. Damon & N. Eisenberg (Eds). *Handbook of Child Psychology, Vol. 3: Social, emotional and personality development*.
- Constant, K., Williams, F., Youngstrom, E., Duax, J., Scovil, K., Carpenter Song, E., et al. (2007). Effects of Ethnicity on Diagnostic Rates of Bipolar Spectrum Disorder and Manic Symptom Expression in Youth Ages 5-17.

- Cook, T., & Campbell, D. (1979). *Quasi-Experimentation: Design and Analysis for Field Settings*. Chicago: Rand McNally.
- Coyle, J. T., Pine, D. S., Charney, D. S., Lewis, L., Nemeroff, C. B., Carlson, G. A., et al. (2003). Depression and Bipolar Support Alliance Consensus Statement on the Unmet Needs in Diagnosis and Treatment of Mood Disorders in Children and Adolescents. [literature review]. *Journal of American Academy of Child and Adolescent Psychiatry*, 42(12), 1494-1503.
- Craney, J., & Geller, B. (2003). A prepubertal and early adolescent bipolar disorder-I phenotype: review of phenomenology and longitudinal course. *Bipolar Disorders*, 5(4), 243-256.
- Danielson, C. K., Youngstrom, E. A., Findling, R. L., Gracious, B. L., & Calabrese, J. R. (2001). Addressing the diagnostic problem of bipolar disorder in youths: Can the Young Mania Rating Scale help?
- DelBello, M. O., Soutullo, C. A., & Strakowski, S. M. (2000). Racial differences in treatment of adolescents with bipolar disorder. *American Journal of Psychiatry*, 157(5), 837-838.
- Dell'Osso, L., Pini, S., Cassano, G., Mastrocinque, C., Seckinger, R., Sacttoni, M., et al. (2002). Insight into illness in patients with mania, mixed mania, bipolar depression and major depression with psychotic features. *Bipolar Disorders*, 4, 315-322.
- Depue, R., Krauss, S., Spont, M., & Arbisi, P. (1989). General behavior inventory identificaion of unipolar and bipolar affective conditions in a nonclinical university population. *Journal of Abnormal Psychology*, 98, 117-126.

DHEW, U. S. (1979). *US DHEW Medical Practice Project 1979. A State of the Service Report for the Office of the Assistant Secretary for the US Dept of Health, Education and Welfare.*

Dienes, K. A., Chang, K. D., Blasey, C. M., Adleman, N. E., & Steiner, H. (2002). Characterization of children of bipolar parents by parent report CBCL. *Journal of Psychiatric Research, 36*(5), 337-346.

Diler, R. S., Birmaher, B., Axelson, D., Goldstein, B., Gill, M., Strober, M., et al. (2009). The Child Behavior Checklist (CBCL) and the CBCL-bipolar phenotype are not useful in diagnosing pediatric bipolar disorder. *Journal Of Child And Adolescent Psychopharmacology, 19*(1), 23-30.

du Toit, M. (Ed.). (2003). *IRT from SSI: BILOG-MG, MULTILOG, PARSCALE, TESTFACT.* Lincolnwood, IL: Scientific Software, Inc.

Duax, J., Youngstrom, E., Calabrese, J., & Findling, R. (2007). Sex Differences in Pediatric Bipolar Disorder. *Journal of Clinical Psychiatry, 68*(10), 1565-1573.

Dunn, L. M., & Dunn, L. M. (1997). *Examiner's Manual for the Peabody Picture Vocabulary Test-Third edition* (3rd ed.). Circle Pines: American Guidance Service.

Eisenberg, N., Fabes, R., Bernzweig, J., Karbon, M., Poulin, R., & Hanish, L. (1993). The relations of emotionality and regulation to preschoolers' social skills and sociometric status. *Child Development, 64*, 1418-1438.

- Eisenberg, N., Fabes, R., Murphy, B., Maszk, P., Smith, M., & Karbon, M. (1995). The role of emotionality and regulation in children's social functioning: A longitudinal study. *Child Development, 66*, 1360-1384.
- Eisenberg, N., Fabes, R., Shepard, S., Murphy, B., Guthrie, I., Jones, S., et al. (1997). Contemporaneous and longitudinal prediction of children's social functioning from regulation and emotionality. *Child Development, 68*, 642-664.
- Eisenberg, N., Valiente, R., Fabes, R., Smith, C., Reiser, M., Shepard, S., et al. (2003). The relations of effortful control and ego control in children's resiliency and social functioning. *Developmental Psychology, 39*, 761-776.
- Erickson, E. (1998). *Identity, Youth and Crisis*. New York: W. W. Norton.
- Faedda, G., Baldessarini, R., Suppes, T., Tondo, L., Becker, I., & Lipschitz, D. S. (1995). Pediatric-Onset bipolar disorder: A neglected clinical and public health problem. *Harvard Review of Psychiatry, 3*, 171-195.
- Fan, X., & Silvo, S. (2005). Sensitivity of fit indices to misspecified structural or measurement model components: Rationale of two index strategies revisited. *Structural Equation Modeling, 12*, 343-367.
- Fan, X., Thompson, B., & Wang, L. (1999). Effects of sample size, estimation methods, and model specification on structural equation modeling fit indexes. *Structural Equation Modeling, 6*(1), 56-83.
- Faraone, S. V., Althoff, R. R., Hudziak, J. J., Monuteaux, M., & Biederman, J. (2005). The CBCL predicts DSM bipolar disorder in children: A receiver operating characteristic curve analysis. *Bipolar Disorders, 7*(6), 518-524.

- Faul, F., Erdfelder, E., Buchner, A., & Lang, A. (2009). Statistical Power Analyses using G*Power 3.1: Tests for Correlation and Regression Analyses. *Behavior Research Methods, 41*(4), 1149-1160.
- Findling, R., Youngstrom, E., Danielson, C., DelPorto-Bedoya, D., Papish-David, R., Townsend, L., et al. (2002). Clinical decision-making using the General Behavior Inventory in juvenile bipolarity. *Bipolar Disorders, 4*(1), 34-42.
- Findling, R., Youngstrom, E., DelPorto, D., Papish-David, R., Townsend, L., Danielson, C., et al. (2001). Clinical decision-making using the General Behavior Inventory. *Bipolar Disorders*.
- Freeman, A. J., Youngstrom, E. A., Michalak, E., Siegel, R., Meyers, O. I., & Findling, R. L. (2009). Quality of life in pediatric bipolar disorder. *Pediatrics, 123*(3), e446-452.
- Frosh, S., Pheonix, A., & Pattman, R. (2002). *Young Masculinities*. Basingstoke: Palgrave.
- Garcia, C. (2010). Conceptualization and Measurement of Coping During Adolescence: A Review of the Literature. *Journal of Nursing Scholarship, 42*(2), 166-185.
- Geller, B., Craney, J., Bolhofner, K., Nickelsburg, M., Williams, M., & Zimmerman, B. (2002). Two-year prospective follow up of children with a prepubertal and early adolescent bipolar disorder phenotype. *American Journal of Psychiatry, 159*, 927-933.

- Geller, B., Fox, L., & Clarke, K. (1994). Rate and predictors of prepubertal bipolarity during follow-up of 6-12 year old depressed children. *Journal of Child and Adolescent Psychiatry*, 33, 461-468.
- Geller, B., & Luby, J. (1997). Child and adolescent bipolar: A review of the past 10 years. *Journal of the American Academy of Child and Adolescent Psychiatry*, 36, 1168-1176.
- Geller, B., Tillman, R., Bolhofner, K., & Zimmerman, B. (2008). Child bipolar I disorder: prospective continuity with adult bipolar I disorder; characteristics of second and third episodes; predictors of 8-year outcome. *Archives Of General Psychiatry*, 65(10), 1125-1133.
- Geller, B., Tillman, R., Craney, J., & Bolhofner, K. (2004). Four-year prospective outcome and natural history of mania in children with a prepubertal and early adolescent bipolar disorder phenotype. *Archives Of General Psychiatry*, 61(5), 459-467.
- Geller, B., Warner, K., Williams, M., & Zimmerman, B. (1998). Prepubertal and young adolescent bipolarity versus ADHD: Assessment and validity using the WASH-U-KSADS, CBCL and TRF. *Journal of Affective Disorders*, 51(2), 93-100.
- Geller, B., Zimmerman, B., Williams, M., Bolhofner, K., & Craney, J. (2001). Bipolar Disorder at prospective follow-up of adults who had prepubertal major depressive disorder. *American Journal of Psychiatry*, 158, 125-127.
- Geller, B., Zimmerman, B., Williams, M., Bolhofner, K., Craney, J., Delbello, M., et al. (2000a). Diagnostic characteristics of 93 cases of a prepubertal and

early adolescent bipolar disorder phenotype by gender, puberty, and comorbid attention deficit hyperactivity disorder. *Journal of Child and Adolescent Psychopharmacology*, 10(3), 157-164.

Geller, B., Zimmerman, B., Williams, M., Bolhofner, K., Craney, J., Delbello, M., et al. (2000b). Six-month stability and outcome of a prepubertal and early adolescent bipolar disorder phenotype. *Journal of Child and Adolescent Psychopharmacology*, 10(3), 165-173.

Geller, B., Zimmerman, B., Williams, M., Delbello, M., Bolhofner, K., Craney, J., et al. (2002). DSM-IV Mania Symptoms in a Prepubertal and Early Adolescent Bipolar Disorder Phenotype Compared to Attention-Deficit Hyperactive and Normal Controls. *Journal of Child and Adolescent Psychopharmacology*, 12(1), 11-25.

Geller, B., Zimmerman, B., Williams, M., Delbello, M., Frazier, J., & Beringer, L. (2002). Phenomenology of prepubertal and early adolescent bipolar disorder: Examples of elated mood, grandiose behaviors, decreased need for sleep, racing thoughts and hypersexuality. *Journal of Child and Adolescent Psychopharmacology*, 12, 3-9.

Geller, B., Zimmerman, B., Williams, M., Bolhofner, K., Craney, J., Delbello, M., et al. (2001). Reliability of the Washington University in St. Louis Kiddie Schedule for Affective Disorders and Schizophrenia (WASH-U-KSADS) Mania and Rapid Cycling Sections. *Journal of the American Academy of Child and Adolescent Psychiatry*, 40, 450-455.

- Geller B. (2005). Prepubertal and early adolescent bipolar I disorder: Review of diagnostic validation by Robins and Guze criteria. *Journal of Clinical Psychiatry, 66*(7), 21-28.
- Gemelli, R. (1996). *Normal Child and Adolescent Development*. Washington D.C.: American Psychiatric Press.
- General, S. (2001). *Mental Health: Culture, Race and Ethnicity: A supplement to Mental Health*.
- Gjerd, P. F. (1995). Alternative pathways to chronic depressive symptoms in young adults: Gender differences in developmental trajectories. *Child Development, 66*, 1277-1300.
- Glorfeld, L. W. (1995). An improvement on Horn's parallel analysis methodology for selecting the correct number of factors to retain. *Educational and Psychological Measurement, 55*, 377-393.
- Gonzalez, R., & Griffen, D. (2001). Testing parameters in structural equation modeling: Every one matters. *Psychological Methods, 6*(3), 258-269.
- Good, B. J. (1997). Studying Mental Illness in Context: Local, Global, or Universal? *Ethos, 25*(2), 230-248.
- Goodwin, F., & Jamison, K. (Eds.). (2007). *Manic-Depressive Illness: Bipolar Disorders and Recurrent Depression* (2nd ed.). New York: Oxford University Press.
- Gracious, B., Youngstrom, E., Findling, R., & Calabrese, J. (2002). Discriminative validity of a parent version of the Young Mania Rating Scale. *Journal of American Academy of Child and Adolescent Psychiatry, 41*, 1350-1359.

- Grant, B. F., Stinson, F. S., Hasin, D. S., Dawson, D. A., Chou, S. P., Ruan, W. J., et al. (2005). Prevalence, correlates, and comorbidity of bipolar I disorder and axis I and II disorders: results from the National Epidemiologic Survey on Alcohol and Related Conditions. *The Journal of clinical psychiatry*, 66(10), 1205-1215.
- Guadagnoli, E., & Velicer, W. F. (1988). Relation to sample size to the stability of component patterns. *Psychological Bulletin*, 103(2), 265-275.
- Hambleton, R. K., & Swaminathan, H. (1985). *Item Response Theory: Principles and Applications*. Boston: Kluwer Nijhoff Publishing.
- Hazell, P., Lewin, T., & Carr, V. (1999). Confirmation that Child Behavioral Checklist Clinical Scales Discriminate Juvenile Mania from Attention Deficit Hyperactivity Disorder. *Journal of Paediatr Child Health*, 35, 199-203.
- Helm, C., Newport, J. D., Bonsall, R., Miller, A. H., & Nemeroff, C. B. (2001). Altered pituitary-adrenal axis responses to provocative challenge tests in adult survivors of childhood abuse. *American Journal of Psychiatry*, 158, 575-581.
- Henry, D. B., Pavuluri, M. N., Youngstrom, E. A., & Birmaher, B. (2008). Accuracy of Brief and Full Forms of the Child Mania Rating Scale. *Journal of Clinical Psychology*, 64, 368-381.
- Hirschfeld, R. M. A., Williams, J. B. W., Spitzer, R. L., Calabrese, J. R., Flynn, L., Keck, P. E., Jr., et al. (2000). Development and validation of a screening instrument for bipolar spectrum disorder: The Mood Disorder Questionnaire. *American Journal of Psychiatry*, 157(11), 1873-1875.

- Hodgins, S., Faucher, B., Zarac, A., & Ellenbogen, M. (2002). Children of parents with bipolar disorder: A population at high risk for major affective disorders. *Child and Adolescent Psychiatric Clinics of North America*, 11, 533-553.
- Hogan, R., Hogan, J., & Roberts, B. (1996). Personality measurement and employment decisions: Questions and answers. *American Psychologist*, 51, 469-477.
- Horn, J. (1965). A Rationale and Test for the Number of Factors in Factor Analysis. *Psychometrika*, 30, 178-185.
- Horn, J. L., & Engstrom, R. (1979). Cattell's scree test in relation to Bartlett's chi-square test and other observations on the number of factors problem. *Multivariate Behavioral Research*, 14(3), 283-300.
- Hu, L., & Bentler, P. (1998). Fit indexes in covariant structural analysis: Sensitivity to underparameterized model misspecification. *Psychological Methods*, 3, 424-453.
- Hu, L., & Bentler, P. (1999). Cut off criterion for fit indexes in covariant structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, 6, 1-55.
- Jöreskog, K. G. (1993). Testing structural equation models. In K. A. Bollen & J. S. Long (Eds.), *Testing Structural Equation Models* (pp. 294-316). Newbury Park, CA: Sage.
- Judd, L., & Akiskal, H. S. (2003). The prevalence and disability of bipolar spectrum disorders in the US population: re-analysis of the ECA database taking into account subthreshold cases. *Journal of Affective Disorders*, 73, 123-131.

- Kahana, S. Y., Youngstrom, E. A., Findling, R. L., & Calabrese, J. R. (2003). Employing Parent, Teacher, and Youth Self-Report Checklists in Identifying Pediatric Bipolar Spectrum Disorders: An Examination of Diagnostic Accuracy and Clinical Utility. *Journal of Child and Adolescent Psychopharmacology*, 13(4), 471-488.
- Kashani, J. H., Beck, N. C., Hooper, E. W., & al., e. (1987). Psychiatric disorders in a community sample of adolescents. *American Journal of Psychiatry*, 144, 584-589.
- Kazdin, A. E. (1999). The meanings and measurement of clinical significance. *Journal of Clinical Psychology*, 67(3), 332-339.
- Kaymaz, N., Krabbendam, L., de Graaf, R., Nolen, W., ten Have, M., & Van OS, J. (2006). Evidence that the urban environment specifically impacts on the psychotic but not the affective dimension of bipolar disorder. *Social Psychiatry and Psychiatric Epidemiology*, 41, 679-685.
- Kessler, R. (1994). Lifetime and 12-month prevalence of DSM-III-R psychiatric disorders in the United States: results from the National Comorbidity Survey. *Journal of the American Medical Association*, 279(9), 654-655.
- Kessler, R., Berglund, P., Demler, O., Jin, R., & Walters, E. (2005). Lifetime Prevalence and Age-of-Onset Distributions of DSM-IV Disorders in the National Comorbidity Survey Replication. *Archives of General Psychiatry*, 62(6), 593-602.
- Kessler, R., Chiu, W., Demler, O., Merikangas, K., & Walters, E. (2005). Prevalence, severity, and comorbidity of 12-month DSM-IV disorders in the

- National Comorbidity Survey Replication. *Archives of General Psychiatry*, 62, 617-627.
- Kline, R. B. (1998). *Principles and practice of structural equation modeling*. New York, NY, USA: Guilford.
- Kovacs, M. (1996). Presentation and course of major depressive disorder during childhood and later years of the life span. *Journal of Child & Adolescent Psychiatry*, 35, 705-715.
- Kowatch, R., Youngstrom, E., Danielyan, A., & Findling, R. (2005). Review and meta-analysis of the phenomenology and clinical characteristics of mania in children and adolescents. *Bipolar Disorders*, 7(6),483-496.
- Kraemer, H. (1992). *Evaluating Medical Tests: Objective and Quantitative Guidelines*. Washington DC: Sage Publications.
- Leff, J. P., Fischer, M., & Bertelsen, A. (1976). A cross national epidemiological study of mania. *British Journal of Psychiatry*, 129, 428-442.
- Leibenluft, E. (2008). Pediatric bipolar disorder comes of age. *Archives of General Psychiatry*, 65(10), 1122-1124.
- Lewinsohn, P., Klein, D., & Seeley, J. (1995). Bipolar disorder in a community sample of older adolescents: Prevalence, phenomenology, comorbidity, and course. *Journal of American Academy of Child and Adolescent Psychiatry*, 34(4), 454-463.
- Lewinsohn, P., Klein, D., & Seeley, J. (2000). Bipolar disorder during adolescence and young adulthood in a community sample. [research]. *Bipolar Disorders*, 2, 281-293.

- Lin, P. I., McInnis, M. G., Potash, J. B., Willour, V., Mackinnon, D. F., & DePaulo, J. R. (2006). Clinical Correlates of Familial Aggregation of Age of Onset in Bipolar Disorder. *American Journal of Psychiatry*, *163*, 240-246.
- Lish, J. D., Dine-Keenan, S., Whybrow, P. C., Price, R. A., & Hirschfeld, R. M. (1994). The National Depressive and Manic-Depressive Association (National DMDA) survey of bipolar members. *Journal of Affective Disorders*, *31*, 281-294.
- Long, J. (1988). *Confirmatory Factor Analysis. A Preface to LISREL*. Beverly Hills: Sage.
- Lopez, S. R., & Guarnaccia, P. J. (2000). Cultural psychopathology: Uncovering the social world of mental illness. *Annual Review of Psychology*, *51*, 571-598.
- Lord, F. (1980). *Applications of Item Response Theory to Practical Testing Problems*. Hillsdale: Lawrence Erlbaum.
- MacCallum, R. C., Widaman, K. F., Zhang, S., & Hong, S. (1999). Sample size in factor analysis. *Psychological Methods*, *4*(1), 84-99.
- Maccoby, E., & Jacklin, C. (1974). *The Psychology of Sex Differences*, Stanford University Press.
- MacLean, A., Sweeting, H., & Hunt, K. (2010). 'Rules' for boys, 'guidelines' for girls: Gender differences in symptom reporting during childhood and adolescence. *Social Science & Medicine*, *70*, 597-604.
- McClellan, J., Kowatch, R. A., Findling, R. L., & et.al. (2007). Practice Parameter for the Assessment and Treatment of Children and Adolescents with Bipolar

- Disorder. *Journal of American Academy of Child and Adolescent Psychiatry*, 46(1), 107-125.
- Meade, A., Johnson, E., & Braddy, P. (2008). Power and sensitivity of alternative indices in tests of measurement invariance. *Journal of Applied Psychology*, 93, 568-592.
- Merikangas, K. R., Akiskal, H. S., Angst, J., Greenberg, P. E., Hirschfeld, R. M. A., Petukhova, M., et al. (2007). Lifetime and 12-month prevalence of bipolar spectrum disorder in the National Comorbidity Survey replication. *Archives Of General Psychiatry*, 64(5), 543-552.
- Messick, S. (1989). Meaning and values in test validation: The science and ethics of assessment. *Educational Researcher*, 18, 5-11.
- Messick, S. (1995). Validity of psychological assessment: Validation from inferences of person's responses and performances as scientific inquiry into score meaning. *American Psychologist*, 50, 741-759.
- Metts, S., Sprecher, S., & Regan, P. (1998). Communication and sexual desire. In P. Andersen & L. Guerrero (Eds.), *Communication and Emotion* (pp. 343-377). Thousand Oaks, CA: Sage.
- Meyer, S. E., Carlson, G. A., Youngstrom, E., Martinez, P. E., Hakak, R., Radke-Yarrow, M., et al. (2009). Long-term outcomes of youth who manifested the CBCL-Pediatric Bipolar Disorder phenotype during childhood and/or adolescence. *Journal of Affective Disorders*, 113(3), 227-235.

- Michaud, P. A., Narring, F., Dubuis-Arber, F., & Paccaud, F. (1993). Health Survey of 15-20 year old adolescents in French-speaking Switzerland. *Schweiz Med Wochenschr*, *123*, 1883-1895.
- Mick, E., Biederman, J., Pandina, G., & Faraone, S. V. (2003). A preliminary meta-analysis of the Child Behavior Checklist in pediatric bipolar disorder. *Biological Psychiatry*, *53*(11), 1021-1027.
- Mueller, R. O. (1996). Confirmatory Factor Analysis *Basic Principles of structured equation modeling: An introduction to LISREL and EQS* (pp. 62-128). New York: Springer-Verlag.
- Munro, B. H. (Ed.). (2005). *Statistical Methods for Healthcare Research* (5th ed.). New York: Lippincott Williams & Wilkins.
- Muroff, J., Edelsohn, G. A., Joe, S., & Ford, B. C. (2008). The role of race in diagnostic and disposition decision making in a pediatric psychiatric emergency service. *General Hospital Psychiatry*, *30*(3), 269-276.
- Murphey, B., Shepard, S., Eisenberg, N., & Fabes, R. (2004). Concurrent and across time prediction of young adolescent's social functioning: The role of emotionality and regulation. *Social Development*, *13*, 56-86.
- Murray, C. J. L., & Lopez, A. D. (1997). Alternative projections of mortality and disability by cause 1990-2020: Global Burden of Disease Study. *Lancet*, *349*(9064), 1498.
- Neighbors, H. W., Caldwell, C., Williams, D. R., Nesse, R., Taylor, R. J., Bullard, K. K., et al. (2007). Race, Ethnicity, and the Use of Services for Mental Disorders. *Archives of General Psychiatry*, *64*, 485-494.

- Neighbors, H. W., Jackson, J. S., Campbell, L., & al., E. (1989). The influence of racial factors on psychiatric diagnosis: a review and suggestions for research. *Community Mental Health Journal, 25*, 301-311.
- Neighbors, H. W., Trierweiler, S. J., Ford, B. F., & Muroff, J. R. (2003). Racial Differences in DSM Diagnosis Using a Semi-Structured Instrument: The Importance of Clinical Judgment in the Diagnosis of African Americans. *Journal of Health and Social Behavior, 44*(3), 237-256.
- Nottelmann, E., Biederman, J., Birmaher, B., Carlson, G. A., Chang, K. D., Fenton, W. S., et al. (2001). National Institute of Mental Health research roundtable on prepubertal bipolar disorder. *Journal of the American Academy of Child & Adolescent Psychiatry, 40*(8), 871-878.
- Offord, D. R., Boyle, M. H., & Szatmari, P. (1987). Ontario Child Health Study, II. Six Month Prevalence of disorder and rates of service utilization. *Archives of General Psychiatry, 44*, 826-836.
- Patel, N. C., Delbello, M. P., & Strakowski, S. M. (2006). Ethnic differences in symptom presentation of youths with bipolar disorder. *Bipolar Disorders, 8*(1), 95-99.
- Pavuluri, M. N., Birmaher, B., & Naylor, M. W. (2005). Pediatric Bipolar Disorder: A Review of the Past 10 Years. *Journal of American Academy of Child and Adolescent Psychiatry, 44*(9), 135-160.
- Pavuluri, M. N., Henry, D. B., Carbray, J. A., & Birmaher, B. (2004). Child Mania Rating Scale (CMRS): Development, reliability and validity. *Biological Psychiatry, 55* (suppl), 845S.

- Pavuluri, M. N., Henry, D. B., Devineni, B., Carbray, J. A., & Birmaher, B. (2006). Child mania rating scale: development, reliability, and validity. *Journal of the American Academy of Child and Adolescent Psychiatry, 45*(5), 550-560.
- Perlis, R. H., Miyahara, S., Marangell, L. B., Wisniewski, S. R., Ostacher, M., DelBello, M. P., et al. (2004). Long-Term implications of early onset in bipolar disorder: data from the first 1000 participants in the systematic treatment enhancement program for bipolar disorder (STEP-BD). *Biological Psychiatry, 55*(9), 875-881.
- Pescosolido, B. A. (1991). Illness careers and network ties: A conceptual model of utilization and compliance. *Advances in Medical Sociology, 2*, 161-184.
- Pescosolido, B. A., (1992). Beyond Rationale Choice: The Social Dynamics of How People Seek Help. *American Journal of Sociology, 97*, 1096-1138.
- Pescosolido, B. A., Gardner, C. B., & Lubell, K. M. (1998). How people get into mental health services: stories of choice, coercion and "muddling through" from "first-timers". *Social Science & Medicine (1982), 46*(2), 275-286.
- Pescosolido, B. A., Monahan, J., Link, B. G., Stueve, A., & Kikuzawa, S. (1999). The public's view of the competence, dangerousness, and need for legal coercion of persons with mental health problems. *American Journal of Public Health, 89*(9), 1339-1345.
- Pfeffer, C. R. (2001). Diagnosis of child and adolescent suicidal behavior: unmet needs for suicide prevention. *Biological Psychiatry, 49*, 1055-1061.
- Pini, S., de Queiroz, V. R., Pagnin, D., Pezawas, L., Angst, J., Cassano, G. B., et al. (2005). Prevalence and burden of bipolar disorders in European

- countries. *European Neuropsychopharmacology: The Journal Of The European College Of Neuropsychopharmacology*, 15(4), 425-434.
- Post, R., & Kowatch, R. (2006). The Health Care Crisis of Childhood-Onset Bipolar Illness: Some Recommendations for Its Amelioration. *Journal of Clinical Psychiatry*, 67, 115-125.
- Post, R., Leverich, G., Xing, G., & Weiss, R. (2001). Developmental vulnerabilities to the onset and course of bipolar disorder. *Development and Psychopathology*, 13(3), 581-598.
- Raju, N.S., Laffitte, L.J., & Byrne, B.M. (2002). Measurement equivalence: A comparison of methods based on confirmatory factor analysis and item response theory. *Journal of Applied Psychology*, 87, 517-529.
- Rao, U., Ryan, N., Birmaher, B., & al., E. (1995). Unipolar depression in adolescents: clinical outcome in adulthood. *Journal of Child and Adolescent Psychiatry*, 34, 566-578.
- Reise, S. P., Widaman, K.F., & Pugh, R.H. (1993). Confirmatory Factor Analysis and Item Response Theory: Two Approaches for Exploring Measurement Invariance. *Psychological Bulletin*, 114(3), 552-566.
- Rios-Ellis, B. (2005). *Critical disparities in Latino mental health: Transforming research into action*. National Council of La Raza.
- Romeo, R. D., & McEwen, B. S. (2006). Stress and the adolescent brain. *Annals of the New York Academy of Sciences*, 1094, 202-214.
- Sattler, J. M. (2002). *Assessment of children: Behavioral and clinical applications (4th ed.)*. La Mesa, CA, US: Jerome M Sattler Publisher.

- Schurhoff, F., Bellivier, F., Jouvent, R., Simeoni, M., Bouvard, M., Allilaire, J., et al. (2000). Early and late onset bipolar disorders: two different forms of manic-depressive illness? *Journal of Affective Disorders*, *58*, 215-221.
- Shankman, S. A., Klein, D. N., Lewinsohn, P. M., Seeley, J. R., & Small, J. W. (2008). Family study of subthreshold psychopathology in a community sample. *Psychological Medicine*, *38*(2), 187-198.
- Sherazi, R., McKeon, P., McDonough, M., Daly, I., & Kennedy, N. (2006). What's New? The Clinical Epidemiology of Bipolar I Disorder. *Harvard Review of Psychiatry*, *November/December*, 273-284.
- Smith, G. T., & McCarthy, D. M. (1995). Methodological considerations in the refinement of clinical assessment instruments. *Psychological Assessment*, *7*(3), 300-308.
- Smoller, J., & Finn, C. (2003). Family, twin and adoption studies of bipolar disorder. *American Journal of Medical Genetics*, *123C*(1), 48-58.
- Snowden, L. R. (2003). Bias in Mental Health Assessment and Intervention: Theory and Evidence. *American Journal of Public Health*, *93*, 239-243.
- Stiffman, A. R., Pescosolido, B., & Cabassa, L. J. (2004). Building a model to understand youth service access: the gateway provider model. *Mental Health Services Research*, *6*(4), 189-198.
- Strakowski, S. M., Flaum, M., Amador, X., Bracha, A. H., Pandurangi, A. K., & Et.al. (1996). Racial Differences in the Diagnosis of Psychosis. *Scizophrenia Research*, *21*, 117-124.

- Strakowski, S. M., McElroy, S. L., Keck, P. E., & West, S. A. (1996). Racial Influences on diagnosis in psychotic mania. *Journal of Affective Disorders*, 39, 157-162.
- Strober, M., & Carlson, G. (1982). Bipolar illness in adolescents with major depression: clinical, genetic, and psychopharmacologic predictors in a three- to four-year prospective follow-up investigation. *Archives Of General Psychiatry*, 39(5), 549-555.
- SurgeonGeneral. (2001). *Mental Health: Culture, Race and Ethnicity: A supplement to Mental Health*.
- Tabachnick, B. G., & Fidell, L. S. (Eds.). (2007). *Using multivariate statistics*. Boston: Allyn & Bacon.
- Teresi, J.A. (2001). Statistical methods for examination of differential item functioning (DIF) with applications to cross-cultural measurement of functional, physical and mental health. *Journal of Mental Health and Aging*, 7(31), 31-40.
- Tillman, R., Geller, B., Klages, T., Corrigan, M., Bolhofner, K., & Zimmerman, B. (2008). Psychotic phenomena in 257 young children and adolescents with bipolar I disorder: delusions and hallucinations (benign and pathological). *Bipolar Disorders*, 10(1), 45-55.
- Tohen, M., Zarate, C., & Turvey, C. (1995). *The McLean First-Episode Mania Project Proceedings of the 148th Annual Meeting American Psychiatric Association*.

- Tsuchiya, K. J., Byrne, M., & Mortensen, P. B. (2003). Risk factors in relation to an emergence of bipolar disorder:a systematic review. *Bipolar Disorders, 5*, 231-242.
- Van Meter, A., Moreira, A., & Youngstrom, E. (2009). A Meta-Analysis of Epidemiological Studies of Pediatric Bipolar Disorder. University of North Carolina.
- Van Ryn, M. (2002). Research of the provider contribution to race/ethnicity disparities in medical care. *Medical Care, 40*(1), 140-151.
- Van Ryn, M., & Burke, J. (2004). The effect of patient race and socio-economic status on physician's perceptions of patients. *Social Science Medicine, 50*(6), 813-828.
- Velicer, W. F. (1976). Determining the number of components from the matrix of partial correlations. *Psychometrika, 41*, 321-327.
- Viner, R., & Christie, D. (2005). ABC of adolescence: Adolescent development. *British Medical Journal, 330*(7486), 301-304.
- Wagner, K. D., Emslie, G. J., Findling, R. L., Gracious, B., & Reed, M. L. (2004). *Clinic validation of the Adolescent Mood Disorder Questionnaire (A-MDQ)*. Paper presented at the Annual Meeting of the American Psychiatric Association.
- Walker, S., Berthelsen, D., & Irving, K. (2001). Temperament and peer acceptance in early childhood: Sex and social status differences. *Child Study Journal, 31*, 177-192.

- Waraich, P., Goldner, E. M., Somers, J. M., & Hsu, L. (2004). Prevalence and incidence studies of mood disorders: a systematic review of the literature. *Canadian Journal Of Psychiatry. Revue Canadienne De Psychiatrie*, *49*(2), 124-138.
- Weist, M., Myers, C., Danforth, J., McNeil, D., Ollendick, T., & Hawkins, R. (2000). Expanded school mental health services: Assessing needs related to school level and geography. *Community Mental Health Journal*, *36*(3), 259-273.
- Weller, E., Weller, R., & Fristad, M. (1995). Bipolar disorder in children: misdiagnosis, underdiagnosis and future directions. *Journal of the American Academy of Child and Adolescent Psychiatry*, *34*, 709-714.
- Wilens, T. E., Biederman, J., Millstein, R., Wozniak, J., Haehsy, A. L., & Spencer, T. J. (1999). Risk for substance use disorders in youths with child and adolescent-onset bipolar disorder. *Journal of American Academy of Child and Adolescent Psychiatry*, *38*(6), 680-685.
- Williams, D., & Mohammed, S. (2009). Discrimination and racial disparities in health: evidence and needed research. *Journal of Behavioral Medicine*, *32*, 20-47.
- Williams, D., & Williams-Morris, R. (2000). Racism and mental health: the African American experience. *Ethn Health*, *5*(3-4), 243-268.
- Wolfson, A., & Carskadon, M. (2001). Sleep deprivation may be undermining teen health. *Monitor on Psychology*, *32*(9), 42.
- Woods, S. W. (2000). The Economic Burden of Bipolar Disorder. *Journal of Clinical Psychiatry*, *61*, 38-41.

- Wozniak, J., Biederman, J., Kiely, K., Ablon, S., Faraone, S., Mundy, E., et al. (1995). Mania-like symptoms suggestive of childhood-onset bipolar disorder in clinically referred children. *Journal of American Academy of Child and Adolescent Psychiatry, 34*(7), 867-876.
- Wozniak, J., Biederman, J., Monuteaux, M., Richards, J., & Faraone, S. (2002). Parsing the comorbidity between bipolar disorder and anxiety disorders: a familial risk analysis. *Journal Of Child And Adolescent Psychopharmacology, 12*(2), 101-111.
- Yatham, L., Kauer-Sant'Anna, M., Bond, D., Lam, R., & Torres, I. (2009). Course and outcome after the first manic episode in patients with bipolar disorder: prospective 12-month data from the Systematic Treatment Optimization Program For Early Mania project. *Canadian Journal Of Psychiatry. Revue Canadienne De Psychiatrie, 54*(2), 105-112.
- Yeh, K., & Yang, Y. (2006). Construct validation of individuating and relating autonomy in culturally Chinese adolescents. *Asian Journal of Social Psychology, 9*, 148-160.
- Young, R., Biggs, J., Ziegler, V., & Meyer, D. (1978). A rating scale for mania: Reliability, validity and sensitivity. *British Journal of Psychiatry, 133*, 429-435.
- Youngstrom, E., Birmaher, B., & Findling, R. (2008). Pediatric Bipolar Disorder: Validity, phenomenology, and recommendations for diagnosis. *Bipolar Disorders, 10*, 194-214.

Youngstrom, E., Danielson, C., Findling, R., Gracious, B., & Calabrese, J. (2002).

Factor structure of the Young Mania Rating Scale for use with youths ages 5 to 17 years. *Journal Of Clinical Child And Adolescent Psychology: The Official Journal For The Society Of Clinical Child And Adolescent Psychology, American Psychological Association, Division 53, 31(4), 567-572.*

Youngstrom, E., & Duax, J. (2005). Evidence-Based Assessment of Pediatric

Bipolar Disorder, Part I: Base Rate and Family History. *Journal of American Academy of Child and Adolescent Psychiatry, 44(7), 68-85.*

Youngstrom, E., Findling, R., Calabrese, J., Gracious, B., Demeter, C., Bedoya,

D., et al. (2004). Comparing the diagnostic accuracy of six potential screening instruments for bipolar disorder in youths aged 5 to 17 years. [research]. *American Academy of Child and Adolescent Psychiatry, 43(7), 847-858.*

Youngstrom, E., Findling, R., Danielson, C., & Calabrese, J. (2001). Discriminative

validity of parent report of hypomanic and depressive symptoms on the general behavior inventory. [research]. *Psychological Assessments, 13, 267-276.*

Youngstrom, E. A., Findling, R. L., Kogos Youngstrom, J., & Calabrese, J. R.

(2005). Toward an Evidence-Based Assessment of Pediatric Bipolar Disorder. *Journal of Clinical Child & Adolescent Psychology, 34(3), 433-448.*

- Youngstrom, E., Meyers, O., Demeter, C., Youngstrom, J., Morello, L., Piiparinen, R., et al. (2005). Comparing diagnostic checklists for pediatric bipolar disorder in academic and community mental health settings. *Bipolar Disorders*, 7(6), 507-517.
- Youngstrom, E., Meyers, O., Youngstrom, J., Calabrese, J., & Findling, R. (2006). Diagnostic and measurement issues in the assessment of pediatric bipolar disorder: implications for understanding mood disorder across the life cycle. *Development And Psychopathology*, 18(4), 989-1021.
- Youngstrom, E., Youngstrom, J. K., & Starr, M. (2005). Bipolar diagnoses in community mental health: Achenbach child behavior checklist profiles and patterns of comorbidity. *Biological Psychiatry*, 58(7), 569-575.
- Yu, C.-Y. (2002). *Evaluating Cutoff Criteria of Model Fit Indices for Latent Variable Models with Binary and Continuous Outcomes*. University of California, Los Angeles.
- Zumbo. (2007). Three Generations of DIF Analyses: Considering Where It Has Been, Where It Is Now, and Where It Is Going. *Language Assessment Quarterly*, 4(2).

APPENDICES

Table 1: Definitions for the purpose of this study:

P-YMRS:	Young Mania Rating Scale (parent version)
P-GBI	General Behavior Inventory (parent version)
CMRS-P	Child Mania Rating Scale (parent version)
P-MDQ	Mood Disorders Questionnaire (parent version)
CBCL	Child Behavior Checklist

Pediatric Bipolar Disorder (PBD): Pediatric bipolar disorder is also referred to as early-onset bipolar disorder and juvenile-onset bipolar disorder in the literature. These terms are used interchangeably to describe bipolar disorder I and II, as defined in the DSM-IV-TR (APA, 2000) and bipolar disorder NOS or Bipolar Spectrum disorder, which are used as a working diagnosis of children and adolescents who are impaired with symptoms of bipolar disorder but who do not meet the full diagnostic criteria as defined in the DSM-IV-TR. (APA, 2000)

Urban Youth: For the purpose of this study, urban youth are defined as families (N=621) with low SES, as defined by Medicaid status, who lived and received services in an urban area. The remaining (N=209) families included in this sample sought outpatient evaluation at an urban academic medical center.

African American: For the purpose of this study, an African American is defined

in this data set to mean individuals of African heritage, who do not necessarily share cultural, religious or ethical beliefs.

Table 2: Core Features of Bipolar Disorder in Prepubertal and Early-Adolescent

Bipolar Disorder Phenotype

Mood Symptoms: Elevated, Expansive or Irritable
Associated Symptoms: 3 out of 7 or 4 out of 7 if mood is primarily irritable
Inflated Self Esteem/Grandiosity
Decreased Need for Sleep
Flight of Ideas or Racing Thoughts
Poor Judgment and Hypersexuality
Distractibility
Goal Directed Activity
Talkative/Pressured Speech

Table 3: Demographics

Ethnicity	Community	Academic	Total
African American	545	28	573
White, Non-Hispanic	39	146	185
Total	584	174	758

Table 4: Diagnosis Stratified by Group

Group	Bipolar I	Bipolar II, NOS, cyclothymia	Unipolar Depression	Disruptive Disorders	Residual Diagnosis	Total
African American	17 (3%)	57 (10%)	160 (28%)	280 (49%)	54 (9%)	568
White	12 (7%)	44 (24%)	50 (27%)	61 (34%)	15 (8%)	182
Younger	12 (3%)	50 (14%)	60 (17%)	204 (57%)	34 (9%)	360
Older	19 (4%)	71 (16%)	169 (37%)	158 (35%)	37 (8%)	454
Male	12 (2%)	71 (14%)	114 (23%)	260 (53%)	33 (7%)	490
Female	19 (6%)	50 (15%)	115 (35%)	103 (31%)	40 (12%)	327

TABLE 5: EIGENVALUES

Component	<i>Initial Eigenvalues</i>			<i>Total Variance Explained</i>			<i>Rotation Sums of Squared Loadings^a</i>
	<i>Total</i>	<i>% of Variance</i>	<i>Cumulative %</i>	<i>Total</i>	<i>% of Variance</i>	<i>Cumulative %</i>	<i>Total</i>
P-YMRS							
1	3.188	28.98	28.98	3.188	28.98	28.98	2.726
2	1.148	10.435	39.415	1.148	10.435	39.415	2.379
3	1.031	9.369	48.784	1.031	9.369	48.784	1.753
4	1.006	9.146	57.93	1.006	9.146	57.93	1.116
5	0.885	8.047	65.977				
P-GBI							
1	9.842	35.149	35.149	9.842	35.149	35.149	8.184
2	1.626	5.808	40.957	1.626	5.808	40.957	7.735
3	1.31	4.677	45.634	1.31	4.677	45.634	4.793
4	1.154	4.123	49.757	1.154	4.123	49.757	3.309
5	0.983	3.511	53.268				
P-CMRS							
1	6.992	33.296	33.296	6.992	33.296	33.296	5.927
2	1.845	8.786	42.083	1.845	8.786	42.083	4.824
3	1.44	6.855	48.938	1.44	6.855	48.938	4.238
4	1.336	6.364	55.302	1.336	6.364	55.302	2.685

5	1.143	5.443	60.745				
P-MDQ							
1	4.277	32.899	32.899	4.277	32.899	32.899	4.136
2	1.29	9.921	42.82	1.29	9.921	42.82	2.267
3	0.98	7.542	50.362				
CBCL							
1	11.15	25.351	25.351	11.15	25.351	25.351	6.95
				5			
2	3.767	8.562	33.913	3.767	8.562	33.913	6.062
3	2.251	5.117	39.03	2.251	5.117	39.03	7.089
4	1.604	3.646	42.677	1.604	3.646	42.677	6.896
5	1.437	3.266	45.943	1.437	3.266	45.943	4.686
6	1.374	3.122	49.065	1.374	3.122	49.065	3.081
7	1.268	2.882	51.947	1.268	2.882	51.947	3.179
8	1.142	2.595	54.542	1.142	2.595	54.542	3.191
9	0.997	2.265	56.807				

Extraction Method: Principal Component Analysis

When components are correlated, sums of squared loadings cannot be added to obtain a total variance.

Table 6: One Group CFA

Instrument	Chi-Squared	DF	P-Value	CFI	TLI	RMSEA
P-YMRS	1573.940	43	*0.000	0.890	0.917	0.092
P-GBI	407.707	44	*0.000	0.901	0.953	0.15
P-CMRS	751.713	69	*0.000	0.769	0.896	0.144
P-MDQ	206.618	82	*0.000	0.965	0.974	0.068

- **Relative chi-squares ratios of < 3.00 are preferred but some researchers consider ratios between 3.00 and 4.00 and less frequently as high as 5.00 to be a good data-model fit.**
- **Recommended cut off values: CFI (<0.95), TLI (<0.95)**
- **RMSEA values of >0.10 indicate poor fit; 0.05 to 0.08 indicate fair fit; and < 0.05 indicate close fit.**

Table 7: P-YMRS: RACE DIF

Table 7 RACE P-YMRS		<i>Constant</i>	<i>Uniform DIF</i>	<i>Slope</i>	<i>Non-Uniform DIF</i>
YMRS-P: Item 1: Elevated Mood	B	*0.527	-0.111	*0.119	-0.027
YMRS-P: Item 2: Increased Motor Activity/Energy	B	*0.344	*-0.375	*0.137	-0.006
YMRS-P: Item 3: Sexual Interest	B	0.169	0.03	*0.02	0.016
YMRS-P: Item 4: Sleep	B	0.051	0.037	*0.075	-0.008
YMRS-P: Item 5: Irritability	B	*0.295	0.191	*0.099	0.009
YMRS-P: Item 6: Speech (Rate and Amount)	B	0.044	-0.17	*0.073	0.018
YMRS-P: Item 7: Thoughts	B	-0.044	-0.07	*0.11	0.01
YMRS-P: Item 8: Content	B	-0.042	0.079	*0.073	-0.023
YMRS-P: Item 9: Disruptive/Aggressive Behavior	B	0.684	-0.017	0.065	0.031
YMRS-P: Item 10: Appearance	B	0.25	-0.192	0.054	0.025

*(p<0.05)

Table 8: P-GBI RACE DIF

Table 8 RACE P-GBI		<i>Constant</i>	<i>Uniform DIF</i>	<i>Slope</i>	<i>Non- Uniform DIF</i>
P-GBI:53. depressed or irritable, and then extremely high, elated, and overflowing with energy	B	-0.005	-0.119	*0.104	0.008
P-GBI:54. feeling unusually happy and intensely energetic	B	0.099	-0.089	*0.108	0.01
P-GBI:4. unusually happy and intensely energetic (clearly more than your usual self)	B	*0.346	*-0.294	*0.107	-0.002
P-GBI:11. unusually happy or high	B	-0.015	-0.128	*0.066	0.009
P-GBI:22. extreme happiness and intense energy lasting several days	B	-0.087	-0.112	*0.097	0.002
P-GBI:40. Mood swings	B	*0.307	*0.398	*0.114	-0.002
P-GBI:27. feeling unusually happy and intensely energetic /rage	B	0.025	-0.108	*0.103	-0.001
P-GBI:19. Rapid shifts in mood or energy	B	*0.306	*0.28	*0.117	-0.009
P-GBI:64. Racing thoughts	B	-0.003	0.126	*0.068	-0.002
P-GBI:31. extreme happiness and intense energy and trouble falling asleep (over an hour)	B	0.029	-0.045	*0.089	0.017

Table 9 P-CMRS: RACE DIF

Table 9 RACE		<i>Constant</i>	<i>Uniform DIF</i>	<i>Slope</i>	<i>Non- Uniform DIF</i>
P-CMRS					
CMRS-P: 1. super happy	B	0.081	0.015	*0.048	*-0.017
CMRS-P: 2. irritable, cranky, or mad	B	*0.393	*0.444	*0.043	0
CMRS-P: 3. can be or do anything	B	*0.62	*-0.777	*0.036	0.015
CMRS-P: 4. abilities or powers	B	-0.044	*-0.229	*0.02	*0.014
CMRS-P: 5. need less sleep	B	-0.069	-0.039	*0.035	0.011
CMRS-P: 6. too much energy	B	0.033	-0.038	*0.074	-0.003
CMRS-P: 7. talks too much, loud, or fast	B	*0.288	-0.046	*0.073	-0.006
CMRS-P: 8. racing thoughts	B	*-0.228	0.151	*0.066	-0.008
CMRS-P: 9. jumping topics	B	*-0.184	0.102	*0.066	-0.008
CMRS-P: 10. rush nonstop	B	-0.112	-0.084	*0.063	-0.001
CMRS-P: 11. off track	B	*0.764	0.02	*0.053	-0.004
CMRS-P: 12. productive/creative	B	-0.021	-0.201	*0.05	0.006
CMRS-P: 13. hypersexuality	B	0.044	-0.069	*0.008	*0.018
CMRS-P: 14. more social	B	-0.089	-0.069	*0.025	0.004
CMRS-P: 15. foolish or risky	B	-0.023	0.062	*0.032	-0.002
CMRS-P: 16. rages	B	*0.217	0.231	*0.057	0.01
CMRS-P: 17. act silly	B	-0.009	0.129	*0.063	-0.008
CMRS-P: 18. mood swings	B	0.12	*0.331	*0.063	0.007
CMRS-P: 19. suspicious	B	-0.1	0.144	*0.033	-0.008

CMRS-P: 20. auditory halluc.	B	*-0.136	0.151	*0.029	*-0.021
CMRS-P: 21. visual halluc.	B	-0.078	0.089	*0.019	*-0.012

*(p<0.05)

Table 10: P-MDQ RACE DIF

Table 10 RACE P-MDQ		Constant	Uniform DIF	Slope	Nonuniform DIF
1. MDQ-P: felt so good or hyper that other people thought your child was not his/her normal self?	B	*-2.383	-0.526	*0.471	0.094
2. MDQ-P: felt so irritable that he/she shouted at people or started fights or arguments?	B	*-0.732	0.246	*0.349	0.063
3. MDQ-P: felt much more self-confident than usual	B	*-2.033	-0.169	*0.322	-0.008
4. MDQ-P: got much less sleep than usual and found he/she didn't really miss it?	B	*-2.733	0.195	*0.383	-0.055
5. MDQ-P: was much more talkative or spoke much faster than usual	B	*-2.506	-0.234	*0.474	0.147
6. MDQ-P: thoughts raced through his/her head or your child couldn't slow his/her mind down?	B	*-3.295	*1.201	*0.443	0.003
7. MDQ-P: were so easily distracted by things around them?	B	-0.27	-0.215	*0.491	-0.141
8. MDQ-P: had much more energy than usual?	B	*-2.92	-1.101	*0.599	*0.287
9. MDQ-P: was much more active or did many more things than usual?	B	*-3.336	-0.87	*0.627	0.093
10. MDQ-P: was much more social or outgoing than usual?	B	*-2.831	0.058	*0.321	-0.075
11. MDQ-P: was much more interested in sex than usual?	B	*-3.33	0.158	*0.214	0.036
12. MDQ-P: did things that were unusual for him/her or that other people thought were excessive?	B	*-2.65	0.561	*0.414	-0.097
13. MDQ-P: spending money got him/her or your family in trouble?	B	*-3.955	0.399	*0.286	-0.099

*(p<0.05)

Table 11: P-YMRS AGE DIF

Table 11 AGE P-YMRS		<i>Constant</i>	<i>Uniform DIF</i>	<i>Slope</i>	<i>Non-Uniform DIF</i>
Item					
YMRS-P: Item 1: Elevated Mood	B	*0.531	-0.109	*0.114	0.001
YMRS-P: Item 2: Increased Motor Activity/Energy	B	*0.41	*-0.32	*0.155	*-0.029
YMRS-P: Item 3: Sexual Interest	B	0.107	0.122	*0.018	0.014
YMRS-P: Item 4: Sleep	B	-0.084	*0.23	*0.073	0.006
YMRS-P: Item 5: Irritability	B	*0.19	*0.268	*0.107	-0.004
YMRS-P: Item 6: Speech (Rate and Amount)	B	0.002	0.018	*0.086	-0.011
YMRS-P: Item 7: Thoughts	B	-0.068	0.043	*0.125	*-0.026
YMRS-P: Item 8: Content	B	-0.041	0.046	*0.062	0.006
YMRS-P: Item 9: Disruptive/Aggressive Behavior	B	*0.735	-0.092	*0.075	-0.004
YMRS-P: Item 10: Appearance	B	*0.196	0.024	*0.058	0.003

*(p<0.05)

Table 12: P-GBI AGE DIF

Table 12	AGE	P-GBI	<i>Constant</i>	<i>Uniform DIF</i>	<i>Slope</i>	<i>Non-Uniform DIF</i>
		P-GBI:53. depressed or irritable, and then extremely high, elated, and overflowing with energy	B -0.10	0.12	*0.11	-0.01
		P-GBI:54. feeling unusually happy and intensely energetic	B -0.01	*0.166	*0.107	0.007
		P-GBI:4. unusually happy and intensely energetic (clearly more than your usual self)	B 0.44	*-0.249	0.105	-0.007
		P-GBI:11. unusually happy or high	B -0.039	-0.016	*0.069	0.001
		P-GBI:22. extreme happiness and intense energy lasting several days	B -0.051	-0.114	*0.093	0.005
		P-GBI:40. Mood swings	B *0.352	0.097	*0.115	-0.001
		P-GBI:27. feeling unusually happy and intensely energetic /rage	B -0.022	0.035	*0.102	0.001
		P-GBI:19. Rapid shifts in mood or energy	B *0.307	0.108	*0.114	0.007
		P-GBI:64. Racing thoughts	B *0.148	*-0.195	*0.062	0.011
		P-GBI:31. extreme happiness and intense energy and trouble falling asleep (over an hour)	B 0.042	-0.032	*0.098	-0.007

*(p<0.05)

Table 13: P-CMRS AGE DIF

Table 13	AGE	P-CMRS	<i>Constant</i>	<i>Uniform DIF</i>	<i>Slope</i>	<i>Non-Uniform DIF</i>
		ITEM				
		CMRS-P: 1. super happy	B 0.107	-0.006	*0.046	-0.008
		CMRS-P: 2. irritable, cranky, or mad	B *0.281	*0.325	*0.05	-0.005
		CMRS-P: 3. can be or do anything	B *0.384	0.15	*0.04	-0.009
		CMRS-P: 4. abilities or powers	B *-0.207	0.159	*0.027	-0.005
		CMRS-P: 5. need less sleep	B -0.152	0.114	*0.042	-0.006
		CMRS-P: 6. too much energy	B *0.276	*-0.349	*0.074	-0.009
		CMRS-P: 7. talks too much, loud, or fast	B *0.453	*-0.318	*0.066	0.009
		CMRS-P: 8. racing thoughts	B *-0.237	0.083	*0.068	-0.008
		CMRS-P: 9. jumping topics	B *-0.173	0.049	*0.067	-0.006
		CMRS-P: 10. rush nonstop	B 0.01	-0.165	*0.066	*-0.013
		CMRS-P: 11. off track	B *1.029	*-0.473	*0.044	*0.015
		CMRS-P: 12. productive/creative	B -0.092	0.116	*0.056	*-0.012
		CMRS-P: 13. hypersexuality	B 0.101	-0.148	*0.008	*0.011
		CMRS-P: 14. more social	B *-0.126	0.073	*0.028	-0.008
		CMRS-P: 15. foolish or risky	B -0.098	0.137	*0.036	-0.006
		CMRS-P: 16. rages	B *0.343	-0.163	*0.059	0.004
		CMRS-P: 17. act silly	B -0.117	0.189	*0.06	0.005
		CMRS-P: 18. mood swings	B 0.044	*0.281	*0.066	0.001

CMRS-P: 19. suspicious	B	-0.242	*0.251	*0.035	-0.003
CMRS-P: 20. auditory halluc.	B	-0.052	-0.062	*0.017	*0.01
CMRS-P: 21. visual halluc.	B	-0.019	-0.022	*0.013	0.003

*(p<0.05)

Table 14: P-MDQ AGE DIF

Table 14	AGE	P- MDQ	<i>Constant</i>	<i>Uniform DIF</i>	<i>Slope</i>	<i>Non-Uniform DIF</i>
ITEM						
1.	MDQ-P:	felt so good or hyper that other people thought your child was not his/her normal self?	B *-2.501	-0.116	*	-0.13
2.	MDQ-P:	felt so irritable that he/she shouted at people or started fights or arguments?	B *-0.81	0.294	*0.382	-0.067
3.	MDQ-P:	felt much more self-confident than usual	B *-2.355	0.481	*0.331	-0.018
4.	MDQ-P:	got much less sleep than usual and found he/she didn't really miss it?	B *-3.135	0.702	*0.443	-0.118
5.	MDQ-P:	was much more talkative or spoke much faster than usual	B *-2.719	0.282	*0.563	-0.093
6.	MDQ-P:	thoughts raced through his/her head or your child couldn't slow his/her mind down?	B *-3.317	0.679	*0.504	-0.114
7.	MDQ-P:	were so easily distracted by things around them?	B 0.223	*-0.912	*0.434	0.001
8.	MDQ-P:	had much more energy than usual?	B *-3.046	-0.574	*0.84	-0.19
9.	MDQ-P:	was much more active or did many more things than usual?	B *-3.639	0.357	*0.744	*-0.181
10.	MDQ-P:	was much more social or outgoing than usual?	B *-4.665	*2.382	*0.449	*-0.154
11.	MDQ-P:	was much more interested in sex than usual?	B *-4.232	*1.446	*0.269	-0.057
12.	MDQ-P:	did things that were unusual for him/her or that other	B *-2.94	*0.796	*0.451	-0.106

people thought were excessive?

13. MDQ-P: spending money got him/her or your family in trouble?	B	*-5.435	*2.02	*0.349	-0.103
---	---	---------	-------	--------	--------

*(p<0.05)

Table 15: P-YMRS GENDER DIF

Table 15 GENDER P-YMRS		<i>Constant</i>	<i>Uniform DIF</i>	<i>Slope</i>	<i>Non-Uniform DIF</i>
ITEM					
YMRS-P: Item 1: Elevated Mood	B	*0.421	0.125	*0.126	-0.029
YMRS-P: Item 2: Increased Motor Activity/Energy	B	*0.266	-0.077	*0.144	-0.014
YMRS-P: Item 3: Sexual Interest	B	*0.187	-0.017	*0.017	*0.021
YMRS-P: Item 4: Sleep	B	0.053	0.001	*0.072	0.006
YMRS-P: Item 5: Irritability	B	*0.288	0.141	*0.102	0.007
YMRS-P: Item 6: Speech (Rate and Amount)	B	-0.01	0.046	*0.085	-0.012
YMRS-P: Item 7: Thoughts	B	-0.023	-0.064	*0.108	0.009
YMRS-P: Item 8: Content	B	-0.054	0.103	*0.067	-0.004
YMRS-P: Item 9: Disruptive/Aggressive Behavior	B	*0.822	*-0.364	*0.067	0.017
YMRS-P: Item 10: Appearance	B	*0.201	0.02	*0.053	0.016

*(p<0.05)

Table 16: P-GBI GENDER DIF

Table 16 GENDER P-GBI		<i>Constant</i>	<i>Uniform DIF</i>	<i>Slope</i>	<i>Non-Uniform DIF</i>
P-GBI:53. depressed or irritable, and then extremely high, elated, and overflowing with energy	B	-0.057	0.066	*0.108	-0.011
P-GBI:54. feeling unusually happy and intensely energetic	B	0.057	0.088	*0.107	0.009
P-GBI:4. unusually happy and intensely energetic (clearly more than your usual self)	B	*0.362	-0.176	*0.104	-0.003
P-GBI:11. unusually happy or high	B	-0.041	-0.019	*0.067	0.005
P-GBI:22. extreme happiness and intense energy lasting several days	B	*-0.091	-0.061	*0.093	0.006
P-GBI:40. Mood swings	B	*0.378	0.076	*0.114	0.001
P-GBI:27. feeling unusually happy and intensely energetic /rage	B	-0.006	0.011	*0.105	-0.007
P-GBI:19. Rapid shifts in mood or energy	B	*0.337	0.089	*0.114	0.01
P-GBI:64. Racing thoughts	B	0.08	-0.104	*0.065	0.008
P-GBI:31. extreme happiness and intense energy and trouble falling asleep (over an hour)	B	0.03	-0.023	*0.095	-7.70E-05

*(p<0.05)

Table 17: P-CMRS GENDER DIF

Table 17 GENDER P-CMRS		<i>Constant</i>	<i>Uniform DIF</i>	<i>Slope</i>	<i>Non-Uniform DIF</i>
CMRS-P: 1. super happy	B	0.098	-0.028	*0.042	0.001
CMRS-P: 2. irritable, cranky, or mad	B	*0.409	0.163	*0.045	0.005
CMRS-P: 3. can be or do anything	B	*0.385	0.187	*0.041	-0.015
CMRS-P: 4. abilities or powers	B	-0.09	-0.066	*0.024	0.002
CMRS-P: 5. need less sleep	B	-0.043	-0.124	*0.036	0.009
CMRS-P: 6. too much energy	B	0.134	-0.226	*0.076	-0.008
CMRS-P: 7. talks too much, loud, or fast	B	*0.317	-0.109	*0.07	0.003
CMRS-P: 8. racing thoughts	B	*-0.198	-0.002	*0.067	-0.005
CMRS-P: 9. jumping topics	B	*-0.15	-0.02	*0.062	0.007
CMRS-P: 10. rush nonstop	B	*-0.151	0.077	*0.065	-0.009
CMRS-P: 11. off track	B	*0.886	*-0.295	*0.05	0.004
CMRS-P: 12. productive/creative	B	-0.113	0.166	*0.054	-0.006
CMRS-P: 13. hypersexuality	B	-0.008	0.099	*0.015	-0.006
CMRS-P: 14. more social	B	-0.075	-0.054	*0.024	0.002
CMRS-P: 15. foolish or risky	B	-0.03	0.018	*0.035	-0.007
CMRS-P: 16. rages	B	*0.236	0.027	*0.061	0.002
CMRS-P: 17. act silly	B	0.035	-0.04	*0.061	-0.003
CMRS-P: 18. mood swings	B	0.148	0.183	*0.063	0.006

CMRS-P: 19. suspicious	B	-0.046	-0.103	*0.029	0.009
CMRS-P: 20. auditory halluc.	B	-0.047	-0.066	*0.016	*0.013
CMRS-P: 21. visual halluc.	B	-0.018	-0.02	*0.014	0.001

*(p<0.05)

Table 18: P-MDQ GENDER DIF

Table 18 GENDER P-MDQ		<i>Constant</i>	<i>Uniform DIF</i>	<i>Slope</i>	<i>Non-Uniform DIF</i>
1. MDQ-P: felt so good or hyper that other people thought your child was not his/her normal self?	B	*-2.306	-0.536	*0.483	0.043
2. MDQ-P: felt so irritable that he/she shouted at people or started fights or arguments?	B	*-0.856	0.499	*0.375	-0.071
3. MDQ-P: felt much more self-confident than usual	B	*-2.092	0.07	*0.321	-0.006
4. MDQ-P: got much less sleep than usual and found he/she didn't really miss it?	B	*-2.737	0.064	*0.367	0.012
5. MDQ-P: was much more talkative or spoke much faster than usual	B	*-2.795	0.55	*0.534	-0.057
6. MDQ-P: thoughts raced through his/her head or your child couldn't slow his/her mind down?	B	*-2.71	-0.535	*0.425	0.03
7. MDQ-P: were so easily distracted by things around them?	B	-0.074	*-0.556	*0.469	-0.076
8. MDQ-P: had much more energy than usual?	B	*-3.07	-0.231	*0.731	-0.133
9. MDQ-P: was much more active or did many more things than usual?	B	*-3.697	0.703	*0.716	*-0.206
10. MDQ-P: was much more social or outgoing than usual?	B	*-3.489	*1.128	*0.337	-0.013

11. MDQ-P: was much more interested in sex than usual?	B	*-3.413	0.426	*0.208	0.03
12. MDQ-P: did things that were unusual for him/her or that other people thought were excessive?	B	*-2.426	-0.059	*0.388	-0.006
13. MDQ-P: spending money got him/her or your family in trouble?	B	*-4.313	0.817	*0.295	-0.068

*(p<0.05)

Figure 1: Gateway Provider Model

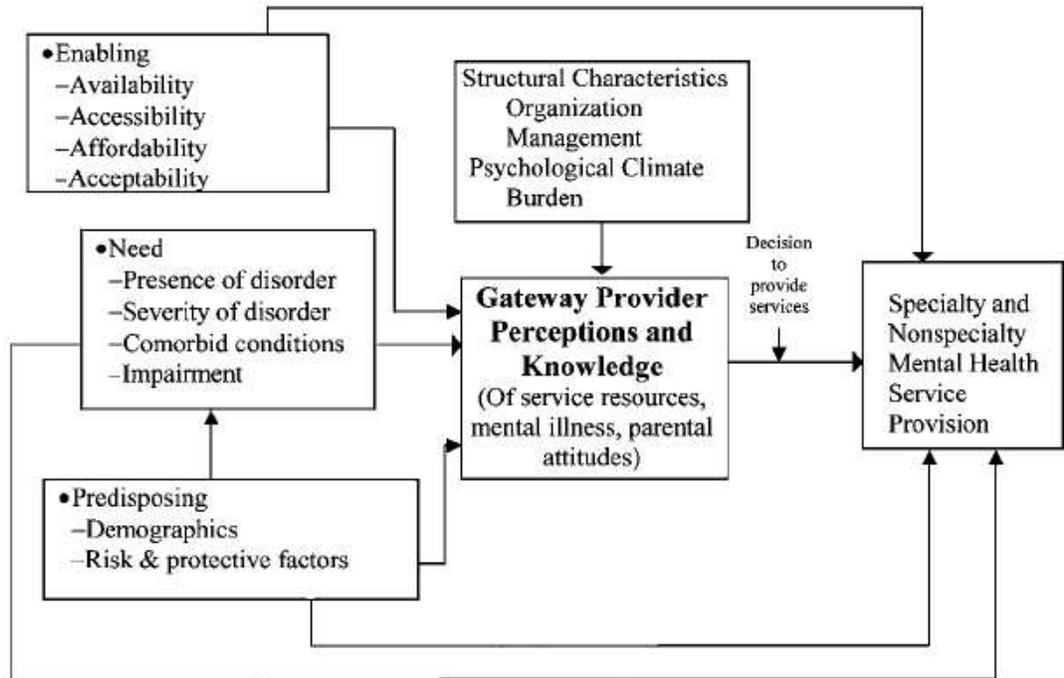


Fig. 1. Gateway provider service framework.

Figure 2: Highest Level of Education

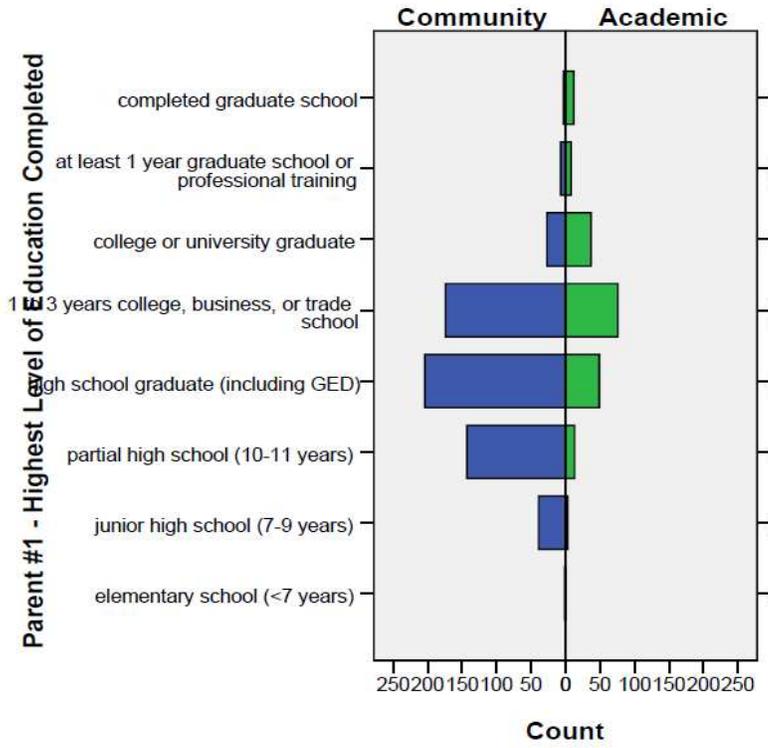


Figure 3: Estimated Annual Income

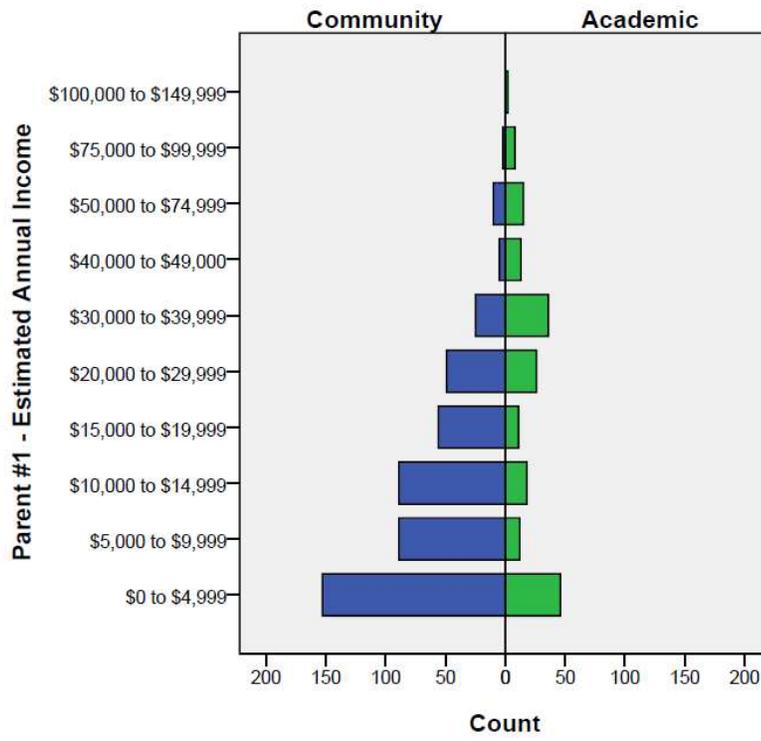


Figure 4: Hollingshead Occupational Scale Score

