

Evaluation and Comparison of Psychometric Instruments for Pediatric Bipolar Spectrum Disorders in Four Age Groups

Thomas W. Frazier, Ph.D.,^{1,3} Christine A. Demeter, M.A.,¹ Eric A. Youngstrom, Ph.D.,⁴
Joseph R. Calabrese, M.D.,¹ Robert J. Stansbrey, M.D.,¹ Nora K. McNamara, M.D.,¹
and Robert L. Findling, M.D.^{1,2}

ABSTRACT

The primary objective of this study was to evaluate the psychometric characteristics of the Young Mania Rating Scale (YMRS), the K-SADS Mania Rating Scale (KMRS), and the Children's Depression Rating Scale-Revised (CDRS-R) across four age groups (4–7, 8–10, 11–13, and 14–17 years). The interrater reliability of K-SADS diagnoses was also examined. Participants included 1,014 youths (62.1% male) presenting to an outpatient clinical research center. Diagnoses were based upon semistructured K-SADS interviews. Symptomatic assessments and ratings of psychosocial functioning were completed following the diagnostic interview. Mania measures showed unifactorial structure and good internal consistency reliability ($\alpha = 0.79$ – 0.95) across all ages groups. The CDRS-R factor structure shifted from one to two factors in adolescents. For all ages and symptom measures, reliability was excellent in the range where differential diagnosis is most difficult. Efficiencies in discriminating bipolar spectrum disorders from other disorders were excellent (areas under the curve, AUCs = 0.92 – 0.99) for mania measures, with comparable discrimination across age groups. Interrater reliability of K-SADS diagnoses was excellent across age groups (smallest $\kappa = 0.95$). Results indicate that mania measures are useful for assessing symptoms across a wide range of ages. The CDRS-R may be better conceptualized as a two-factor measure in older adolescents. The semistructured K-SADS interview can be used to generate reliable diagnoses across a broad age range.

INTRODUCTION

PEDIATRIC BIPOLAR SPECTRUM DISORDERS (BPSDs; Bipolar I and II, Cyclothymia, and

Bipolar not otherwise specified, NOS) are debilitating conditions frequently associated with substantial psychosocial dysfunction and human suffering (Wozniak et al. 1995; Geller et al.

¹Departments of Psychiatry and ²Pediatrics, Case Western Reserve University, University Hospitals Case Medical Center, Cleveland, Ohio.

³Department of Medical Subspecialty Pediatrics, The Cleveland Clinic, Cleveland, Ohio.

⁴Department of Psychology, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina.

The Stanley Medical Research Institute primarily supported this study. It was also supported in part by an National Institute of Mental Health (NIMH) Developing Centers for Interventions and Services Research Grant (P 20 MH-66054).

2000; Pliszka et al. 2000; Findling et al. 2001). Many youths experience symptoms for years prior to formal diagnosis (Geller and Luby 1997; Findling et al. 2001) and recurrences of manic episodes may become more severe and treatment resistant over time. Thus, it is extremely important to identify BPSD promptly and accurately.

Unfortunately, identification of BPSD in youths is often quite difficult. Many symptoms of bipolarity (such as hyperactivity, aggression, and irritability) co-occur in other common childhood disorders, such as attention-deficit/hyperactivity disorder (ADHD), depression, and conduct disorder (Bowring and Kovacs 1992; Carlson and Kelly 1998; Geller et al. 1998; Kim and Miklowitz 2002). Also, youths with BPSD often have co-morbid conditions that complicate symptom presentation (Lewinsohn et al. 1995; Geller et al. 2000; Kowatch et al. 2005). Furthermore, the course of BPSD is often different in children than adults (Findling et al. 2001; Geller et al. 2002). This may be particularly true in young children (ages 4–7), where the diagnosis of BPSD has become increasingly more common in clinical settings (Tumuluru et al. 2003; Youngstrom et al. 2005; Blader and Carlson 2006).

Semistructured diagnostic instruments, such as the Schedule for Affective Disorder and Schizophrenia for Children (K-SADS) (Orvaschel 1994; Kaufman et al. 1997) are particularly well-suited to making accurate differential diagnoses. These measures gather large amounts of information regarding the presence, history, and time course of bipolar symptoms (Kaufman et al. 1997). However, these instruments are impractical in many clinical settings where clinicians do not have several hours to spend with patients to generate a differential diagnosis. In contrast, clinician-completed rating scales may be more useful to screen for bipolar symptoms, assess the severity of symptoms, and track symptoms over time.

Several clinician-rated measures of mood symptoms have been investigated in youth referred for evaluation of bipolar disorder, including the commonly used Young Mania Rating Scale (YMRS; Young et al. 1978), the

K-SADS Mania Rating Scale (KMRS) (Axelson et al. 2003), and the Children's Depression Rating Scale-Revised (CDRS-R) (Poznanski et al. 1985). Several studies have examined basic psychometric properties of these measures in referred children and adolescents (Poznanski et al. 1985; Fristad et al. 1992; Fristad et al. 1995; Geller et al. 1998). However, the authors are aware of only one study that directly compared the psychometric characteristics of clinician-rated measures between children and adolescents referred for evaluation of BPSDs. Youngstrom and colleagues (2002) directly compared the factor structure of the YMRS between children and adolescents, finding unifactorial and equivalent structure across age groups. However, this study combined younger children into a single age group (age 5–11) and focused exclusively on the factor structure of one instrument. To the author's knowledge, no previous studies have comprehensively compared the psychometric characteristics of these measures across young children (ages 4–7), older children (ages 8–10), and adolescents (ages 11–13 and 14–17).

The primary goal of the present study was to evaluate and describe the psychometric characteristics (factor structure, internal consistency reliability, item-response theory-derived test information, and diagnostic efficiency) of the YMRS, KMRS, and CDRS-R across the first two decades of life. Understanding the psychometric characteristics of these instruments will assist clinicians in determining the utility of briefer, less costly measures in obtaining mood symptom information and choosing the most useful measure for the intended clinical purpose (screening, evaluating symptom severity, measuring changes in mood symptoms over time). Although it is a measure of depression symptoms, the CDRS-R was included because this measure is commonly used with other mania measures to characterize low mood and depressive symptoms associated with bipolar disorder. Additionally, the incremental utility of this measure in screening for bipolar spectrum disorders is an important empirical question. A second goal of the present study was to examine the interrater reliability

of K-SADS-derived diagnoses across childhood and adolescence.

prior to any study-related procedures being performed.

METHOD

Participants

Participants were ascertained during two separate cohorts of data collection (Cohort 1 1996–2003, $n = 769$; Cohort 2 2003–present, $n = 245$). Participants were recruited primarily as part of the screening procedures for various treatment studies across different diagnostic entities, including mood disorders, ADHD, and disruptive behavior disorders. These samples were enriched by referrals of children who had at least one parent diagnosed with a bipolar disorder and were participating in treatment or research at the University Hospitals Case Medical Center Mood Disorders Program. Youths were also recruited via fliers and word of mouth. More information about the ascertainment of these subjects has been described elsewhere (Findling et al. 2001; Findling et al. 2005).

To be included in the present study, youths had to be ages 4–17 years and both the youth and primary caregiver must have been present for the assessment. Exclusion criteria included: (1) difficulties in oral communication invalidating completion of the diagnostic measures; (2) evidence of a pervasive developmental disorder as determined by psychiatric history, interview, or Autism Screening Questionnaire score >15 (Berument et al. 1999); (3) suspected mental retardation documented by educational history, standardized cognitive ability test scores <70 , or a Peabody Picture Vocabulary Test–Third Edition score <70 (Dunn and Dunn 1997); or (4) if the child/adolescent had a current or past medical or neurological history that could significantly affect the youngster's current mood or behavior.

The University Hospitals of Cleveland and Case Western Reserve University Institutional Review Boards for Human Investigation approved all procedures in this study. Written informed consent was obtained from each subject's guardian, and written assent, as age appropriate, was obtained from each subject

Diagnostic and mood symptom severity measures

Eligible subjects and their parents/guardians were both interviewed by a highly trained rater. The semistructured diagnostic instrument used in the interview was either the Schedule for Affective Disorders and Schizophrenia for School-Age Children–Epidemiologic Version (K-SADS-E; Orvaschel 1994), the Schedule for Affective Disorders and Schizophrenia for School-Age Children–Present and Lifetime version (K-SADS-PL; Kaufman et al. 1997), a slightly modified WASH-U-K-SADS (Geller et al. 1998; Geller et al. 2001) or a slightly modified K-SADS-PL that included additional information from the WASH-U K-SADS. All versions of the K-SADS assess for the history of psychiatric symptomatology. The first group of subjects was interviewed with the K-SADS-E. To decrease the interview burden, when the K-SADS-PL became available, it was used instead of the K-SADS-E. The WASH-U K-SADS and/or items from the WASH-U-K-SADS were added to the K-SADS-PL in the interviews administered to some subjects as part of collaborative efforts with other research groups.

The basic training procedures, item interrater reliability, and K-SADS interview procedures used in this study have been described in more detail in a prior publication (Findling et al. 2001). Briefly, interviewers (ranging from bachelor's degree to a doctoral degree) were trained to criterion by completing ratings independently of an experienced rater over the course of three to five interviews. Then, new raters led three to five K-SADS interviews with an experienced rater. For young children, mood symptom ratings were determined by the parent interview and observations of the child's behavior. Consistent with the procedures of all K-SADS interviews, best estimates of diagnoses were made on the basis of all available information. Discrepancies between parent and child were resolved using best clinical judgment. All cases were also reviewed by a licensed clinician who discussed content with the interviewer.

For the first cohort of data collection (Cohort 1), the majority of participants received a diagnostic assessment using the K-SADS-E version ($n = 172$), the Present and Lifetime version ($n = 571$), or an unmodified version of the WASH-U K-SADS ($n = 25$). For the second cohort of data collection (Cohort 2), 245 subjects completed the modified K-SADS-PL, which includes inquiries about mood symptomatology from the WASH-U-K-SADS.

Overall, 631 (62.3%) of the diagnoses were based upon the results of both a K-SADS interview and a clinical interview with a child and adolescent psychiatrist. In a subset ($n = 401$) of these 631 consensus diagnosis sessions, K-SADS diagnosis and child and adolescent psychiatrist were tracked separately to determine reliability between the K-SADS interview and the clinical interview with the child and adolescent psychiatrist. These clinical diagnoses confirmed the K-SADS diagnoses in 93.7% of cases. Also, a total of 59 participant interviews were independently confirmed by a second interviewer, who participated in the initial interview and made independent ratings. These data are used to compute interrater reliability of K-SADS diagnoses.

The Clinician's Global Assessment of Functioning (CGAS; Schaffer et al. 1983) was used to assess child and adolescent overall functioning. This clinician-rated instrument has score ranging from 0 to 100, with 100 being superior functioning at home, school, and with peers. Only Cohort 2 participants were rated on this measure. The Clinical Global Impressions-Severity Scale (CGI-S; National Institute of Mental Health 1985) was used to assess the severity of illness. This clinician-rated instrument has scores ranging from 1 (normal) to 7 (severely ill). Only Cohort 2 participants were rated on this measure.

The YMRS (Young et al. 1978) is an 11-item clinician-rated measure of manic symptoms. Item scores range from 0 to 4 (three items range from 0 to 8). The YMRS yields a total score that ranges from 0 to 56, with higher scores indicating greater manic symptoms. Ratings were based upon reports of symptoms experienced over the past 1–2 weeks. The KMRS (Axelson et al. 2003) is a 13-item clinician-rated measure of mania that is adapted from questions in the

K-SADS mania section. For the present study, items were rated based upon report from the parent and youth. Items were rated on a 1–6 scale (except item 10, rated 1–5) representing increasing levels of symptom severity. Items were summed and 13 points were subtracted from the total, yielding a total score that ranged from 0 to 64, with higher scores indicating greater psychopathology. The CDRS-R (Poznanski et al. 1985) is a 17-item clinician-rated measure of depressive symptoms. For the present study, ratings were based upon reports of symptoms over the past 1–2 weeks. Items were rated on a 1–7 scale with the exception of items 4, 5, and 16, which were rated on a 1–5 scale. Items were summed, and 17 points were subtracted from the total. Subtraction was done for interpretive purposes so that scores of zero indicated the absence of any depressive symptoms. Total scores ranged from 0 to 96, with higher scores indicating greater depressive symptoms. The CDRS-R was included in the present study because it evaluates the symptoms of depression often seen in bipolar disorder and to examine the validity of the CDRS-R in predicting diagnosis relative to the validity of mania measures.

Procedure

Participants completed the intake assessment, which involved the K-SADS diagnostic interview of the youth and primary caregiver and/or a clinical assessment by a child and adolescent psychiatrist. The clinician who completed the K-SADS diagnostic interview completed the YMRS, KMRS, and CDRS-R. Cohort 1 participants were not rated on the KMRS because this was not part of the study related procedures for participants in Cohort 1.

Statistical methods

To examine potential age-related differences in the psychometric properties of interview-rated measures, children and adolescents were placed into four age comparison groups that included young children (ages 4–7; $n = 220$), older children (ages 8–10; $n = 311$), and two groups of adolescents (ages 11–13, $n = 222$ and 14–17, $n = 261$). These groups were chosen to provide adequate sample sizes for each sub-

group and still provide a more refined look at age effects. Preliminary descriptive analyses examined whether demographic and clinician-report data differed across subsamples (Cohorts 1 and 2). Due to the differences in recruitment criteria for treatment studies across ascertainment cohorts, we anticipated significant demographic and diagnostic differences. It should be noted, however, that any observed differences increase the heterogeneity of the sample, thereby enhancing the generalizability of psychometric findings.

Determining whether age groups show equivalent factor structures is important for determining whether clinical measures are functioning consistently across age and to determine whether modifications to the scoring system are needed. The first step in determining whether age groups demonstrated equivalent factor structures was to determine the number of factors measured by each instrument, separately by age group. To do this, principal components analyses were computed and the resulting eigenvalues compared to randomly generated eigenvalues obtained from Glorfeld's modification of Horn's parallel analysis (Horn 1965; Glorfeld 1995). This rule has been shown to be highly accurate in Monte Carlo studies (Zwick and Velicer 1986). Once the number of factors was determined, multi-group confirmatory factor analyses (CFAs) were used to test the invariance of factor structures for each measure across age groups (Reise et al. 1993; Arbuckle 2003).

Multigroup CFAs compared two models: Unconstrained and equal factor loadings. The unconstrained model allows factor loadings to vary across groups. The equal factor loadings model assumes that factor loadings are equal across groups. The difference in fit between these models is evaluated using the chi-square difference test. For analyses of the KMRS, frequency counts for item 11 (Hallucinations) were extremely low and therefore this item was removed from multigroup CFA analyses to facilitate estimation.

Examination of the internal consistency reliability of a measure across age groups is important for determining whether the error variance for an instrument is sufficiently small and remains modest across different ages. Reliabil-

ity of index tests was examined in two ways. First, Cronbach's alpha was computed for each measure to estimate internal consistency. These analyses were computed separately for each age group and ascertainment cohort, as well as pooled together. Age group and test comparisons of reliability coefficients were examined by converting reliability coefficients to Fisher's z' . An analysis of variance (ANOVA) was then performed with Age Group as the between-subjects independent variable, Rating Scale as the within-subjects independent variable and z' -transformed reliability coefficients as the dependent variable.

Second, item response theory (IRT) analyses, using Samejima's Graded Response Model (Samejima 1969; Reise and Yu 1990), were computed for each measure across each age group. These analyses are crucial for determining whether the error variance associated with a measure is consistent across different score ranges (mild, moderate, or severe bipolar symptoms), ages, or their interaction (Age \times symptom level). Test information was computed at six points along the latent trait for each measure ($\theta = -2.0$, $\theta = -1.0$, $\theta = 0.0$, $\theta = 1.0$, $\theta = 2.0$, and $\theta = 3.0$) to examine whether scales showed different patterns of performance across levels of symptom severity. These points of the latent trait roughly correspond with population z -scores reflecting an individual's score's standing relative to the population mean. Theta equal to 3.0 was included to examine whether measures would differ in estimated reliability across the range of the latent trait where only moderate to severely impaired bipolar (YMRS and KMRS) or depressed (CDRS-R) patients would be expected to score. This provides some information about the relative utility of measures for grading the severity of mood disturbance.

Test information refers to the fidelity of measurement at each point on the latent trait, with larger values indicative of greater measurement precision. Reliability estimates were computed by transforming test information to the standard error of measurement (Hambleton et al. 1991). Then, standard errors of measurement were converted to reliability coefficients (Nunnally and Bernstein 1994). To maximize sample size for these analyses, data were col-

lapsed across site. The KMRS sample size was limited because this measure was only collected in ascertainment Cohort 2. To maximize sample size for this measure, the two youngest and oldest age groups were combined. Data were presented as reliability coefficients because these coefficients are more clinically useful and easily interpreted. However, analyses examining age differences in measurement fidelity focused on test information estimates. To examine age and latent trait level differences, a 4 (Age Group) \times 6 (Latent Trait Level: $\theta = -2, -1, 0, +1, +2, +3$) mixed factorial ANOVA was performed with age group as the between-subjects independent variable and latent trait as within-subjects independent variables. These analyses were performed separately for each measure because of the different age groupings used for the KMRS. To examine differences in item information across measures, item information values were averaged across age groups and a repeated measures ANOVA was computed with measure as the within-subjects variable.

Because K-SADS diagnoses were used as the diagnostic criterion, secondary analyses examined the interrater reliability of these diagnoses. Interrater reliability was examined by first transforming primary diagnoses into five categories for each rater: syndromal bipolar disorder (bipolar I or II disorder), subsyndromal bipolar disorder (bipolar disorder NOS or cyclothymia), unipolar mood disorder, disruptive behavior disorders (ADHD, oppositional defiant disorder, or conduct disorder), and a residual category (other diagnoses or no axis I diagnosis). Due to the reduced number of cases with two independent K-SADS raters, kappa coefficients were computed for child and adolescent age groups (ages 4–10 and 11–17 years).

Diagnostic efficiency analyses were computed to compare the relative utility of mood symptom measures in discriminating BPSD from other clinical conditions. The term diagnostic efficiency is used because the measures examined were not intended to be screening measures but fall much lower on the cost/effort continuum relative to more labor intensive measures often used to make diagnoses in research, such as the K-SADS. The primary criterion measure for all diagnostic efficiency

analyses grouped youths into two categories: (0) those with no diagnosis of a BPSD, although multiple other Axis I diagnoses may be present and (1) those with any BPSD (i.e. bipolar I, bipolar II, cyclothymia, or bipolar NOS) present regardless of whether other Axis I conditions might also be co-morbid. Diagnostic efficiency was quantified using nonparametric estimates of the area under the curve (AUC) from receiver operating characteristic analyses. Because the same clinicians completed the diagnostic interview and symptom measures, these analyses were strictly focused on within study comparisons of diagnostic efficiency across measures and age groups.

Total scores on the three measures were used as predictors. Age differences in diagnostic efficiency were compared using the *z* test of independent AUCs (Hanley and McNeil 1982). Test differences in diagnostic efficiency were compared using the *z* test of dependent AUCs (Hanley and McNeil 1983). Logistic regression analyses determined whether CDRS-R provides incremental validity after interpreting the YMRS and/or KMRS (Hosmer and Lemeshow 1989).

SPSS (2004) was used to compute internal consistency reliability, principal components analyses for determining the number of factors measured, and diagnostic efficiency analyses. Multi-group confirmatory factor analyses were performed using AMOS (Arbuckle 2003). Item-response theory analyses were performed using Multilog v7.03 (du Toit 2003). Glorfeld's modification of Horn's parallel analysis and Minimum Average Partial Analysis were performed using SPSS syntax from O'Connor (2000). The significance level (α) of 0.05 was used for all analyses.

RESULTS

Sample size and basic demographics

Table 1 contains overall and subsample descriptive statistics for demographic, diagnostic, and interviewer-rated measures. Ascertainment cohorts differed significantly in age, broad diagnostic distributions (bipolar disorder vs. other Axis I diagnoses vs. no Axis I diagnosis), and bipolar disorder diagnoses (syn-

TABLE 1. DESCRIPTIVE STATISTICS FOR DEMOGRAPHIC, DIAGNOSTIC, AND CLINICIAN-RATED MEASURES, OVERALL AND ACROSS SUBSAMPLES

n	Cohort 1, 769	Cohort 2, 245	Overall, 1,014	F/X ² (p)
Age (SD)	11.39 (3.43)	10.26 (3.62)	11.11 (3.51)	19.34 (<0.001)**
Gender (%)				0.07 (0.788)
Males	476 (61.9)	154 (62.9)	630 (62.1)	
Females	293 (38.1)	91 (37.1)	384 (37.9)	
Bipolar disease diagnoses (%)	389 (50.6)	146 (60.6)	535 (53.0)	7.37 (0.025)*
Syndromal (BP-1/BP-2)	249 (64.0)	58 (39.7)	307 (57.4)	25.60 (<0.001)**
Subsyndromal (cyclo/BP-NOS)	140 (36.0)	88 (60.3)	228 (42.6)	
Other axis I diagnoses (%)	322 (41.9)	81 (33.6)	403 (39.9)	
Unipolar mood disorders	159 (49.4)	32 (39.5)	191 (47.4)	2.62 (0.270)
Disruptive behavior disorders	144 (44.7)	44 (54.3)	188 (46.7)	
Other disorders	19 (5.9)	5 (6.2)	24 (6.0)	
No axis I diagnosis (%)	58 (7.5)	14 (5.8)	72 (7.1)	
Bipolar disease course modifiers (%)				
Rapid cycling	127 (50.8)	26 (44.8)	153 (49.7)	0.67 (0.412)
Psychosis	17 (6.8)	16 (27.6)	33 (10.7)	21.26 (<0.001)**
Mixed states	24 (9.6)	2 (3.4)	26 (8.4)	2.29 (0.131)
Mean YMRS Score (SD)	11.74 (12.22)	17.14 (10.98)	13.05 (12.16)	38.01 (<0.001)**
Mean KMRS Score (SD)		16.80 (11.78)		
Mean CDRS-R Score (SD)	17.20 (17.18)	13.58 (11.25)	16.32 (16.02)	9.53 (0.002)*
Mean CGAS Score (SD)		55.39 (7.93)		
Mean CGI Score (SD)		3.62 (0.93)		

Abbreviations: SD = standard deviation; YMRS = Young Mania Rating Scale; KMRS = K-SADS Mania Rating Scale; CDRS-R = Children's Depression Rating Scale-Revised; CGAS = Children's Global Assessment Scale; CGI = Clinical Global Impressions Severity Scale.

* $p < 0.05$; ** $p < 0.001$.

dromal bipolar vs. subsyndromal bipolar). The Cohort 1 subsample was significantly older, had a greater proportion of bipolar diagnoses relative to other diagnoses or no diagnosis, and had a greater proportion of syndromal bipolar relative to subsyndromal bipolar. Ascertainment cohorts also differed significantly in the proportion of bipolar cases with psychotic features and showed significant differences in mood symptom severity. The Cohort 1 subsample had significantly fewer bipolar cases with psychosis, lower mania symptom severity, and higher depression symptom severity. There were no significant differences in gender distributions or in the distribution of non-bipolar diagnoses.

Factor structure

Table 2 presents the results of principal components analyses and Horn's parallel analyses (HPA). These analyses indicated unifactorial structure for all measures across all age groups. The only exception to this was

suggestion of a minor second factor for the CDRS-R in the oldest age group. In this age group, oblimin rotation of the factors indicated separation of the cognitive and somatic/affective items. However, in spite of this minor second factor in the oldest age group, all analyses involving CDRS-R items indicated a large first factor. Therefore, multi-group CFAs were computed using unifactorial structure for each measure.

Table 3 contains the results of multigroup CFAs. Results indicated factorial invariance for the KMRS ($p = 0.650$). However, results for the YMRS and CDRS-R indicated significantly different factor loadings across age groups (p values < 0.001). For the YMRS, items 8 (Thought Content) and 9 (Disruptive-Aggressive Behavior) had the largest differences in factor loadings across ages. Allowing these loadings to vary across age groups resulted in a nonsignificant model comparison ($\Delta X^2 = 36.68$, $p = 0.101$). However, over half of the CDRS-R items showed substantial changes across age.

TABLE 2. EIGENVALUES DERIVED FROM PRINCIPAL COMPONENTS AND HORN'S PARALLEL ANALYSES FOR EACH MEASURE ACROSS AGE GROUPS

Component		Ages 4-7	Ages 8-10	Ages 11-13	Ages 14-17
KMRS	1	5.99 (1.86)	6.46 (1.78)	6.88 (1.80)	6.28 (1.84)
	2	1.42 (1.68)	1.44 (1.65)	1.36 (1.66)	1.45 (1.67)
	3	1.18 (1.54)	1.05 (1.48)	0.98 (1.51)	1.11 (1.52)
YMRS	1	5.37 (1.37)	5.83 (1.30)	6.22 (1.31)	6.73 (1.31)
	2	1.14 (1.27)	1.08 (1.18)	1.06 (1.20)	0.99 (1.20)
	3	0.98 (1.19)	0.84 (1.16)	0.93 (1.17)	0.82 (1.18)
CDRS-R	1	6.77 (1.51)	8.65 (1.45)	8.04 (1.47)	6.73 (1.46)
	2	1.37 (1.41)	1.29 (1.38)	1.37 (1.40)	1.57 (1.41)
	3	1.20 (1.32)	1.02 (1.28)	0.98 (1.30)	1.16 (1.30)

Abbreviations: KMRS = K-SADS Mania Rating Scale; YMRS = Young Mania Rating Scale; CDRS-R = Children's Depression Rating Scale-Revised.

Note: Parentheses indicate the upper end of the 95% confidence interval of the random eigenvalues derived from Horn's Parallel Analysis.

Reliability of mood severity rating scales

Table 4 contains internal consistency reliability for each measure, overall and separately by age group and subsample. Inspection of this table reveals very good reliability for all measures across all age groups ($\alpha = 0.79-0.95$). However, internal consistency increased significantly with age (linear contrast $p = 0.041$). There were also significant differences in the reliability of mood measures ($F(2,8) = 7.32, p = 0.016$). The YMRS was significantly less reliable than the KMRS and CDRS-R (p values = 0.016

and 0.011). However, the KMRS and CDRS-R did not significantly differ in internal consistency reliability ($p = 0.501$).

Table 5 contains item-response theory-derived reliability estimates for six points along the latent score range of each measure. There were no age differences in reliability for any measure (largest $F(1,10) = 0.39, p = 0.546$). For each measure, reliability estimates increased as scores increased until the population mean ($\theta = 0$ or $+1$) and then began decreasing at approximately one standard deviation above the population mean (quadratic contrast $F(1,9) =$

TABLE 3. RESULTS OF MULTIGROUP CFA FOR EACH CLINICIAN-RATED MEASURE

	Model	χ^2 (df)	$\Delta\chi^2$ (p)
YMRS	Unconstrained	883.11 (176)	
	Factor loadings equal	955.77 (209)	72.66 (<0.001)
KMRS	Unconstrained	440.06 (216)	
	Factor loadings equal	472.27 (252)	32.21 (0.650)
CDRS-R	Unconstrained	1995.02 (476)	
	Factor loadings equal	2442.89 (527)	447.87 (<0.001)

Abbreviations: df = degrees of freedom; YMRS = Young Mania Rating Scale; KMRS = K-SADS Mania Rating Scale; CDRS-R = Children's Depression Rating Scale-Revised. CFA = confirmatory factor analysis.

TABLE 4. INTERNAL CONSISTENCY RELIABILITY FOR EACH CLINICIAN-RATED MEASURE, SEPARATELY BY AGE GROUP AND SUBSAMPLE

	Ages 4-7	Ages 8-10	Ages 11-13	Ages 14-17	All ages
YMRS					
Cohort 1	0.89	0.91	0.92	0.92	0.91
Cohort 2	0.84	0.87	0.87	0.93	0.88
Overall	0.88	0.90	0.91	0.92	0.91
KMRS					
Overall	0.90	0.91	0.92	0.95	0.92
CDRS-R					
Cohort 1	0.93	0.94	0.93	0.93	0.94
Cohort 2	0.79	0.85	0.87	0.89	0.89
Overall	0.90	0.94	0.93	0.93	0.93

Abbreviations: YMRS = Young Mania Rating Scale; KMRS = K-SADS Mania Rating Scale; CDRS-R = Children’s Depression Rating Scale-Revised.

83.66, $p < 0.011$). Measures did not significantly differ in test information ($F(2,10) = 1.14$, $p = 0.358$).

Qualitatively, reliability estimates were quite good for all measures from approximately 1 SD below the population mean to approximately 2 SD above the population mean. However, for extremely high scores, YMRS reliability estimates were somewhat weaker than KMRS and CDRS-R estimates, indicating that the YMRS may be less useful for grading symptom severity in bipolar cases than the other measures.

The interrater reliability of K-SADS diagnoses was high for both age groups (ages 4-10 $\kappa = 1.00$; ages 11-17 $\kappa = 0.95$). The only misclassification observed was in the adolescent age group and involved disagreement as to

syndromal versus subsyndromal bipolar disorder.

Diagnostic efficiency

Table 6 contains AUCs representing diagnostic efficiency (bipolar disorder vs. no bipolar disorder) for each measure, separately for each age group and subsample and overall. There were no significant age differences in diagnostic efficiency for the YMRS or KMRS (largest $z = 1.16$, $p = 0.244$). The YMRS and KMRS possessed excellent diagnostic efficiency at all age levels. Comparison of the KMRS and YMRS did not show any significant test differences in diagnostic efficiency at any age group (largest $z = 1.11$, $p = 0.267$).

TABLE 5. ITEM RESPONSE THEORY RELIABILITY ESTIMATES FOR EACH CLINICIAN-RATED MEASURE, SEPARATELY BY AGE GROUPS

θ	-2	-1	0	+1	+2	+3
YMRS						
Ages 4-7	0.43	0.91	0.95	0.96	0.92	0.65
Ages 8-10	0.23	0.97	0.98	0.96	0.93	0.54
Ages 11-13	0.74	0.93	0.96	0.95	0.92	0.58
Ages 14-17	0.62	0.82	0.98	0.98	0.91	0.56
KMRS						
Ages 4-10	0.79	0.92	0.94	0.93	0.92	0.89
Ages 11-17	0.67	0.91	0.97	0.97	0.91	0.76
CDRS-R						
Ages 4-7	0.72	0.84	0.96	0.97	0.97	0.93
Ages 8-10	0.72	0.89	0.97	0.97	0.96	0.86
Ages 11-13	0.74	0.89	0.96	0.96	0.95	0.84
Ages 14-17	0.80	0.93	0.95	0.95	0.94	0.86

Abbreviations: YMRS = Young Mania Rating Scale; KMRS = K-SADS Mania Rating Scale; CDRS-R = Children’s Depression Rating Scale-Revised.

TABLE 6. DIAGNOSTIC EFFICIENCY (AREAS UNDER THE CURVE) FOR EACH CLINICIAN-RATED MEASURE, SEPARATELY BY AGE GROUP AND SUBSAMPLE

	Age 4-7	Age 8-10	Age 11-13	Age 14-17	All ages
YMRS					
Cohort 1	0.95	0.99	0.98	0.96	0.97
Cohort 2	0.99	0.98	0.92	0.97	0.97
Overall	0.96	0.98	0.97	0.96	0.97
KMRS					
Overall	0.98	0.98	0.94	0.96	0.96
CDRS-R					
Cohort 1	0.68	0.60	0.59	0.42 ^a	0.54
Cohort 2	0.78	0.73	0.70	0.40 ^a	0.57
Overall	0.71	0.63	0.61	0.41 ^a	0.54

Abbreviations: YMRS = Young Mania Rating Scale; KMRS = K-SADS Mania Rating Scale; CDRS-R = Children's Depression Rating Scale-Revised.

^aTest is actually predicting in the opposite direction in the oldest age group.

The CDRS-R had a significant decrease in diagnostic efficiency with increasing age ($z = 4.98, p < 0.001$). In the oldest age group (ages 14-17), the diagnostic efficiency of the CDRS-R was actually less than chance. In the 14-17 age group, low depression scores were more indicative of bipolar status. This effect was driven by the increase in the number of unipolar depressed cases in the older age brackets, increasing the average CDRS-R score for the non-bipolar comparison group. The diagnostic efficiency of the CDRS-R was substantially weaker for discriminating bipolar from non-bipolar cases than the YMRS and KMRS (smallest $z = 2.97, p = 0.003$).

Results of the hierarchical logistic regression analysis indicated that both the YMRS and KMRS, entered simultaneously in step 1, contributed significant unique variance to the prediction of bipolar disorder (YMRS $\beta = 0.023, t = 4.72, p < 0.001$; KMRS $\beta = 0.013, t = 2.84, p = 0.005$). However, the CDRS-R did not contribute significant incremental variance to this prediction ($F(1,237) = 1.57, p = 0.212$).

DISCUSSION

The present study demonstrates that three mood symptom measures commonly used in the evaluation of pediatric bipolar disorder have good to excellent psychometric properties across a wide range of ages and mood symptom levels. The clinician-rated mania measures (YMRS, KMRS) were found to be reliable for

measuring mania symptoms across a wide range of ages. Characterization of the CDRS-R was more variable depending upon the psychometric characteristic examined. More detailed discussion of the performance of these measures is provided below.

Factor structure

The YMRS and KMRS had strong and consistent one-factor solutions across age groups. This finding supports the use of a single total score on these measures in predictions regarding the presence of a bipolar spectrum disorder and in tracking the severity of mania symptoms over time.

Analyses of the CDRS-R indicated that the instrument appears to measure a single factor in children and younger adolescents (ages 4-13 years) and two factors in older adolescents (ages 14-17 years). This suggests that in older adolescents the CDRS-R resembles adult measures of depression, which consistently show both somatic-affective and cognitive factors (Steer et al. 1999). Recent research in children and adolescents has suggested that the CDRS-R measures five factors (Guo et al. 2006); however, this work did not use empirically validated methods for determining the number of factors measured. Regardless, the strong likelihood of multiple factors, at least in older adolescents, suggests that future studies examine scores on these factors as well as a total score. Examination of CDRS-R factors in previous clinical depression treatment trials that used

this outcome measure (Emslie et al. 2006; Wagner et al. 2006) may be useful for determining whether treatment effects were specific to cognitive or somatic-affective depressive symptoms.

Multigroup analyses indicated that the KMRS had the same factor structure across all age groups, whereas the YMRS showed changing factor loadings across age groups. However, inspection of the YMRS factor loadings at each age group indicated relatively minor differences across age groups. The largest factor loading difference across ages was $\Delta = 0.23$ for item 8, Thought Content. Thus, it is likely that the mania factor, measured by the YMRS items, undergoes only minor changes in meaning across age groups, consistent with previous findings (Youngstrom et al. 2002).

The CDRS-R also showed significant factor loading differences across age groups and the changes were much more substantial than for the YMRS. For example, the standardized factor loading for CDRS-R item 17 (Hypoactivity) changes from 0.29, a fairly weak loading, in the youngest age group to 0.67, a moderately strong loading, in the oldest age group. In general, these results suggest that the structure of the CDRS-R is changing substantially from early childhood to late adolescence.

Reliability

Total scores on all three measures had good to excellent internal consistency reliability across ages. However, reliability increased slightly with age. This may suggest increasing clarity of symptom presentation with age. The KMRS also showed slightly better internal consistency than the YMRS, again suggesting that the KMRS may be more useful in pediatric populations.

The IRT estimated reliability of all three measures was excellent in symptom score ranges where discrimination of bipolar disorders from other disorders would be most difficult (i.e., Bipolar vs. ADHD or oppositional defiant disorder, ODD). Additionally, all three measures showed good levels of reliability 1 SD below the population mean on mania and depression symptoms, indicating that these measures are reliable for evaluating even minor disturbances in mood.

In addition, the KMRS and CDRS-R had good estimated reliability in the range of moderate to extreme scores. This indicates that these measures are quite reliable for grading the severity of mood disturbance in individuals with substantial mood symptoms. In contrast, the YMRS appears to be less useful for grading the severity of mania symptoms in individuals with significant illness.

The strong interrater reliability of K-SADS diagnoses in very young children indicates that dysfunctional mood symptoms are readily identified and distinguished in this population. This finding strongly supports the use of this diagnostic measure in future phenomenology and treatment studies focused on bipolar disorder in preschool children. Future studies are needed that further examine the interrater reliability of K-SADS diagnoses in young children using separate raters and separate interviews, because this will provide a more stringent test of the reliability of the instrument.

Diagnostic efficiency

The YMRS and KMRS showed excellent ability to separate bipolar from non-bipolar cases across all age groups. Additionally, each measure contributed unique variance to the prediction of bipolar illness. Concurrent use of these measures in evaluating the symptoms of bipolar disorder in children is recommended.

The CDRS-R was less effective in discriminating bipolar from non-bipolar cases and did not add significantly to the discrimination of bipolar disorder from other conditions. This is not surprising because many cases of pediatric bipolarity present with minimal depressive symptoms (Findling et al. 2001). However, although not the focus of the present study, this measure is likely to be useful for distinguishing mood disorder from other conditions. The CDRS-R also showed an interesting pattern of decreasing diagnostic efficiency with increasing age. In the oldest age group, the instrument predicted in the opposite direction; low scores indicated greater probability of bipolar illness. This was driven by the increasing rate of unipolar depression in the sample, a pattern that is also likely to be true in many clinical settings

given the increase in the occurrence of depression after the onset of puberty (Cyranowski et al. 2000).

Strengths and limitations

The present study had several strengths, including a large sample size, a fairly heterogeneous population derived from two demographically and diagnostically different subsamples, and evaluation of multiple psychometric characteristics using several different analytic techniques. Limitations included reduced sample size and sample heterogeneity for the KMRS and very little variability for the KMRS item 11 (Hallucinations). Unfortunately, the KMRS was only administered during the second ascertainment cohort, limiting the examination of age subgroups for IRT analyses and the generalizability of the KMRS findings. An additional limitation of the present study was the overrepresentation of bipolar disorders in the present sample and lack of information about subtypes of bipolar disorders beyond individuals with rapid cycling. A final consideration is that the mood severity ratings and the criterion diagnoses were based on the same interviewer working with the family. This may have inflated the diagnostic efficiencies of mania measures. For the purposes of the present study, comparing tests to each other and evaluating changes in performance across age groups, is actually a design strength because the same methodology is being used across all comparisons. However, the estimates of diagnostic efficiency reported here should not be compared to published values for other tests where the diagnosis was made independently of the rating scale.

Clinical and research implications

The psychometric characteristics of the KMRS, YMRS, and CDRS-R are generally stable across age, facilitating their use as symptomatic assessments in clinical practice and future phenomenological and treatment studies. However, in older adolescents, the CDRS-R should be viewed as a two-factor instrument. The KMRS appears superior to the YMRS at grading mania symptom severity in individu-

als with more significant bipolar illness. For this reason, clinicians who use only one mania measure for ongoing symptomatic assessment in bipolar patients would be advised to use the KMRS. However, clinicians would be advised to include both mania measures in the context of a broader diagnostic assessment, particularly given that the YMRS has been the most widely used instrument in clinical trials to date.

Future research should examine the KMRS in more heterogeneous samples, including unselected community samples. Diagnostic efficiencies for the KMRS and other measures should be examined in more typical clinical settings where the base rates of bipolar disorders are likely to be substantially smaller. The inclusion of inpatients may also be helpful because this population is likely to have larger variability for items tapping more severe illness, such as items covering hallucinations and delusions. Targeted sampling of the youngest ages may also permit a more fine-grained analysis of psychometric properties in these children where bipolar diagnoses remain controversial. Researchers measuring depressive symptoms in older adolescents should examine subscales of the CDRS-R in addition to the total score when examining phenomenology or treatment effects.

DISCLOSURES

Dr. Findling receives or has received research support, acted as a consultant or served on a speaker's bureau for Abbott, AstraZeneca, Bristol-Myers Squibb, Celltech-Medeva, Forest, GlaxoSmithKline, Johnson & Johnson, Lilly, New River, Novartis, Otsuka, Pfizer, Shire, Solvay, and Wyeth. Drs. McNamara and Stansbrey have no financial ties to disclose. Dr. Youngstrom has received research support from Abbott. Dr. Calabrese has received research support, acted as a consultant, or served on a speaker's bureau for Abbott, AstraZeneca, Bristol Myers Squibb, Eli Lilly, GlaxoSmithKline, Janssen, Merck, Otsuka, Pfizer, and Teva. Drs. Frazier, Demeter, Stansbrey, and McNamara have no conflicts of interest or financial ties to disclose.

REFERENCES

- Arbuckle JL: Amos, 5.0 edition. Chicago, Smallwaters, 2003.
- Axelson D, Birmaher B, Brent D, Wassick S, Hoover C, Bridge J, Ryan N: 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, 2003.
- Berument J, Rutter M, Lord C, Pickles A, Bailey A: Autism screening questionnaire: Diagnostic validity. *Brit J Psychiat* 175:444–451, 1999.
- Blader JC, Carlson G: BPD diagnosis among child and adolescent U.S. psychiatric inpatients, 1996–2003. Chicago (Illinois), Paper presented at the NIMH Pediatric Bipolar Disorder Conference, 2006.
- Bowring MA, Kovacs M: Difficulties in diagnosing manic disorders among children and adolescents. *J Am Acad Child Adolesc Psychiatry* 31:611–614, 1992.
- Carlson GA, Kelly KI: Manic symptoms in psychiatrically hospitalized children: What do they mean? *J Affect Disorders* 51:123–135, 1998.
- Cyranowski JM, Frank E, Young E, Shear K: Adolescent onset of the gender difference in lifetime rates of major depression. *Arch Gen Psychiat* 57:21–27, 2000.
- du Toit M: Irt from ssi: Bilog-mg, multilog, parscale, testfact. Lincolnwood (Illinois), Scientific Software, Inc., 2003.
- Dunn LM, Dunn LM: Examiner's manual for the Peabody Picture Vocabulary Test, 3rd edition. Circle Pines (Minnesota), American Guidance Service, 1997.
- Emslie GJ, Wagner KD, Kutcher S, Krulewicz S, Fong R, Carpenter DJ, Lipschitz A, Machin A, Wilkinson C: Paroxetine treatment in children and adolescents with major depressive disorder: A randomized, multicenter, double-blind, placebo-controlled trial. *J Am Acad Child Adolesc Psychiatry* 45:709–719, 2006.
- Findling RL, Gracious BL, McNamara NK, Youngstrom EA, Demeter C, Calabrese JR: Rapid, continuous cycling and psychiatric co-morbidity in pediatric bipolar I disorder. *Bipolar Disord* 3:202–210, 2001.
- Findling RL, Youngstrom EA, McNamara NK, Stansbrey RJ, Demeter C, Bedoya D, Kahana SY, Calabrese JR: Early symptoms of mania and the role of parental risk. *Bipolar Disord* 7:623–634, 2005.
- Fristad MA, Weller EB, Weller RA: The Mania Rating Scale (MRS): Can it be used with children? A preliminary report. *J Am Acad Child Adolesc Psychiatry* 31:252–257, 1992.
- Fristad MA, Weller RA, Weller EB: The Mania Rating Scale (MRS): Further reliability and validity studies with children. *Ann Clin Psychiatry* 7:127–132, 1995.
- Geller B, Warner K, Williams M, Zimmerman B: Prepubertal and young adolescent bipolarity versus ADHD: Assessment and validity using the WASH-U-KSADS, CBCL, and TRF. *J Affect Disorders* 51:93–100, 1998.
- Geller B, Bolhofner K, Craney JL, Williams M, DelBello MP, Gundersen K: Psychosocial functioning in a prepubertal and early adolescent bipolar disorder phenotype. *J Am Acad Child Adolesc Psychiatry* 39:1543–1548, 2000.
- Geller B, Chaney JL, Bolhofner K, Nickelsburg MJ, Williams M, Zimmerman B: Two-year prospective follow-up of children with a prepubertal and early adolescent bipolar disorder phenotype. *Am J Psychiatry* 159:927–933, 2002.
- Geller B, Luby J: Child and adolescent bipolar disorder: A review of the past 10 years. *J Am Acad Child Adolesc Psychiatry* 36:1168–1176, 1997.
- Geller B, Zimmerman B, Williams M, Bolhofner K, Craney JL, DelBello MP, Soutullo C: Reliability of the Washington University in St. Louis Kiddie Schedule for Affective Disorders and Schizophrenia (WASH-U-KSADS) mania and rapid cycling sections. *J Am Acad Child Adolesc Psychiatry* 40:450–455, 2001.
- Glorfeld LW: An improvement on Horn's parallel analysis methodology for selecting the correct number of factors to retain. *Educational and Psychological Measurement* 55:377–393, 1995.
- Guo Y, Nilsson ME, Heiligenstein J, Wilson MG, Emslie G: An exploratory factor analysis of the Children's Depression Rating Scale-Revised. *J Child Adolesc Psychopharm* 16:482–491, 2006.
- Hambleton RK, Swaminathan H, Rogers HJ: Fundamentals of item response theory. Newbury Park (California), Sage, 1991.
- Hanley JA, McNeil BJ: The meaning and use of the area under the Receiver Operating Characteristic (ROC) curve. *Radiology* 143:29–36, 1982.
- Hanley JA, McNeil BJ: A method of comparing the areas under receiver operating characteristic curves derived from the same cases. *Radiology* 148:839–843, 1983.
- Horn JL: A rationale and test for the number of factors in factor analysis. *Psychometrika* 30:179–185, 1965.
- Hosmer DW, Lemeshow S: Applied Logistic Regression. New York, Wiley, 1989.
- Kaufman J, Birmaher B, Brent D, Rao U, Flynn C, Moreci P, Williamson D, Ryan N: Schedule for Affective Disorders and Schizophrenia for School-Age Children—Present and Lifetime Version (K-SADS-PL): Initial reliability and validity data. *J Am Acad Child Adolesc Psychiatry* 36:980–988, 1997.
- Kim EY, Miklowitz DJ: Childhood mania, attention deficit hyperactivity disorder and conduct disorder: A critical review of diagnostic dilemmas. *Bipolar Disord* 4:215–225, 2002.
- Kowatch RA, Youngstrom EA, Danielyan A, Findling RL: Review and meta-analysis of the phenomenology and clinical characteristics of mania in children and adolescents. *Bipolar Disord* 7:483–496, 2005.
- Lewinsohn PM, Klein DN, Seeley JR: Bipolar disorders in a community sample of older adolescents: Prevalence, phenomenology, comorbidity, and course. *J Am Acad Child Adolesc Psychiatry* 34:454–463, 1995.
- Nunnally JC, Bernstein IH: Psychometric theory. New York, McGraw-Hill, Inc., 1994.
- O'Connor BP: SPSS and SAS programs for determining the number of components using parallel analysis and

- Velicer's MAP test. *Behavior Research Methods Instruments & Computers* 32:396–402, 2000.
- Orvaschel H: Schedule for Affective Disorders and Schizophrenia for School-Age Children-Epidemiologic Version, Fifth Revision. Fort Lauderdale, FL, Nova Southeastern University, 1994.
- Pliszka S, Sherman JO, Barrow MV, Irick S: Affective disorder in juvenile offenders: A preliminary study. *Am J Psychiatry* 157:130–132, 2000.
- Poznanski EO, Freeman LN, Mokros HB: Children's Depression Rating Scale-Revised. *Psychopharmacol Bull* 21:979–989, 1985.
- Reise SP, Yu J: Parameter recovery in the graded response model using MULTILOG. *J Educ Meas* 27:133–144, 1990.
- Reise SP, Widaman KF, Pugh RH: Confirmatory factor analysis and item response theory: Two approaches for exploring measurement invariance. *Psycholog Bull* 114:552–566, 1993.
- Samejima F: Estimation of latent ability using a response pattern of graded scores. *Psychometrika Monogr* 34:100, 1969.
- Schaffer D, Gould MS, Brasic J, Ambrosini P, Fisher P, Bird H, Aluwahlia S: A Children's Global Assessment Scale (CGAS). *Arch Gen Psychiat* 40:1228–1231, 1983.
- SPSS Professional Statistical Package. 13.0 edition. Chicago (Illinois), SPSS, Inc., 2004.
- Steer RA, Ball R, Ranieri WF, Beck AT: Dimensions of the Beck Depression Inventory-II in clinically depressed outpatients. *J Clin Psychol* 55:117–128, 1999.
- Tumuluru RV, Weller EB, Fristad MA, Weller RA: Mania in six preschool children. *J Child Adolesc Psychopharmacol* 13:489–494, 2003.
- Wagner KD, Jonas J, Findling RL, Ventura D, Saikali K: A double-blind, randomized, placebo-controlled trial of escitalopram in the treatment of pediatric depression. *J Am Acad Child Adolesc Psychiatry* 45:280–288, 2006.
- Wozniak J, Biederman J, Kiely K, Ablon JS, Faraone SV, Mundy E, Mennin D: Mania-like symptoms suggestive of childhood-onset bipolar disorder in clinically referred children. *J Am Acad Child Adolesc Psychiatry* 34:867–876, 1995.
- Young RC, Biggs JT, Ziegler MG, Meyer DA: A rating scale for mania: Reliability, validity, and sensitivity. *Brit J Psychiat* 133:429–435, 1978.
- Youngstrom EA, Danielson CK, Findling RL, Gracious BL, Calabrese JR: Factor structure of the Young Mania Rating Scale for use with youths ages 5 to 17 years. *J Clin Child Adolesc Psychol* 31:567–572, 2002.
- Youngstrom EA, Youngstrom JK, Starr M: Bipolar diagnoses in community mental health: Achenbach CBCL profiles and patterns of comorbidity. *Biol Psychiatry* 58:569–575, 2005.
- Zwick WR, Velicer WF: Comparison of five rules for determining the number of components to retain. *Psycholog Bull* 99:432–442, 1986.

Address reprint requests to:

*Dr. Thomas W. Frazier
Department of Pediatrics-Section
of Behavioral Medicine
The Cleveland Clinic
9500 Euclid Avenue
Cleveland, OH 44109*

E-mail: fraziet2@ccf.org

This article has been cited by:

1. Stephanie Danner, Mary A. Fristad, L. Eugene Arnold, Eric A. Youngstrom, Boris Birmaher, Sarah M. Horwitz, Christine Demeter, Robert L. Findling, Robert A. Kowatch. 2009. Early-Onset Bipolar Spectrum Disorders: Diagnostic Issues. *Clinical Child and Family Psychology Review* 12:3, 271-293. [[CrossRef](#)]
2. Lakshmi N Yatham, Sidney H Kennedy, Ayal Schaffer, Sagar V Parikh, Serge Beaulieu, Claire O'Donovan, Glenda MacQueen, Roger S McIntyre, Verinder Sharma, Arun Ravindran, L Trevor Young, Allan H Young, Martin Alda, Roumen Milev, Eduard Vieta, Joseph R Calabrese, Michael Berk, Kyooseob Ha, Flávio Kapczinski. 2009. Canadian Network for Mood and Anxiety Treatments (CANMAT) and International Society for Bipolar Disorders (ISBD) collaborative update of CANMAT guidelines for the management of patients with bipolar disorder: update 2009. *Bipolar Disorders* 11:3, 225-255. [[CrossRef](#)]
3. Boris Birmaher, David Axelson, Michael Strober, Mary Kay Gill, Mei Yang, Neal Ryan, Benjamin Goldstein, Jeffrey Hunt, Christianne Esposito-Smythers, Satish Iyengar, Tina Goldstein, Laurel Chiapetta, Martin Keller, Henrietta Leonard. 2009. Comparison of manic and depressive symptoms between children and adolescents with bipolar spectrum disorders. *Bipolar Disorders* 11:1, 52-62. [[CrossRef](#)]