Discriminative Validity of the General Behavior Inventory Using Youth Report

Carla Kmett Danielson,[1] **Eric A. Youngstrom,**[1,2,3] **Robert L. Findling,**[1,2] **and Joseph R. Calabrese**[1,2]

The present study investigated the ability of the General Behavior Inventory (GBI) to discriminate between diagnostic groups using youth self-report. One hundred and ninety-seven youths ages 10–17 years presenting at a midwestern urban outpatient clinic specializing in mood disorders completed the GBI as part of the intake process. Diagnoses were determined by a structured clinical interview (K-SADS) administered by either a child and adolescent psychiatrist or a research assistant trained to a high level of interrater reliability ($\kappa > .85$). Games–Howell post hoc tests showed that the diagnostic groups significantly differed on the GBI's 2 subscales, Depression and Hypomanic–Biphasic. Logistic regression demonstrated that the scales discriminated between bipolar and disruptive behavior disorders, unipolar and bipolar depression, and mood and disruptive behavior disorders or no diagnosis. Receiver Operating Characteristic (ROC) curves further indicated the good diagnostic efficiency of the scales. Results indicate that the GBI's subscales might aid in making traditionally difficult differential diagnoses, such as between bipolar disorder and Attention Deficit Hyperactivity Disorder (ADHD) and between unipolar and bipolar depression.

KEY WORDS: differential diagnosis; child psychiatry; clinical assessment; bipolar disorder; depression.

Accurately diagnosing psychiatric disorders in youth has proven to be difficult as a result of overlapping symptoms and high rates of comorbidity. Specifically, recent attention in the literature has focused on clinicians' difficulty differentiating among bipolar disorders and Attention Deficit Hyperactivity Disorder (ADHD; e.g., Biederman, 1998; Geller, Warner, Williams, Zimerman, et al., 1998), as mania often presents similarly to hyperactivity and impulsivity (Biederman, 1995; Biederman, Russell, Soriano, Wozniak, & Faraone, 1998; Carlson, 1998; Weller, Weller, & Fristad, 1995). Likewise, hypomania can be quite difficult to differentiate from hyperactivity. Further, the differential diagnosis between bipolar disorders and other disruptive behavior disorders, such as conduct disorder (CD), also can prove to be a difficult task. Bowring and Kovacs (1992) explain that although, "the overlap between mania and CD may not be readily apparent.... Children with either of these diagnoses may present with a negative, irritable mood, including belligerence and hostility" (p. 613). Despite these similarities in presentation, manic disorders and CD demonstrate differences in the quality of disturbed mood, the syndromatic nature of the manic disorders, age of onset, and course (Bowring & Kovacs, 1992). This is even further complicated in that youths with ADHD often have other comorbid disruptive behavior disorders (Giedd, 2000), so diagnosis involves differentiating among bipolar disorders versus multiple disruptive behavior disorders, such as ADHD and CD.

The distinction between bipolar versus unipolar depression also has been recognized as difficult in both adults and youth. For example, Baldessarini and Tondo (2000) report that bipolar disorder in adults is frequently mistaken for major depression and is often inappropriately treated. In children and adolescents, irritability is considered a

[1]Department of Psychology, Case Western Reserve University, Cleveland, Ohio.
[2]Department of Psychiatry, University Hospitals of Cleveland/Case Western Reserve University, Cleveland, Ohio.
[3]Address all correspondence to Eric A. Youngstrom, PhD, Department of Psychology, Case Western Reserve University, 10900 Euclid Avenue, Cleveland, Ohio 44106-7123; e-mail: eay@po.cwru.edu.

hallmark symptom of unipolar depression. Irritability is not only common to the depressive episodes of bipolar disorder, but it is common to manic episodes as well (Biederman, 1995). Further, when a youth presents with unipolar depressive symptoms, even if the youth has never been treated for a psychiatric disorder before, a unipolar depressive disorder cannot be diagnosed unless a history of manic or hypomanic symptoms can be ruled out for certain (i.e., the youth could be experiencing a major depressive episode in and of itself, or the youth could be experiencing a depressive episode of bipolar disorder). Attempting to make distinctions between the two disorders has been the subject of several discussions and investigations (e.g., Kovacs, 1987; Winokur, Coryell, Endicott, & Akiskal, 1993).

Accurate diagnostic decision making is imperative to ensure provision of the most appropriate interventions. The consequences of an incorrect diagnostic decision can be costly for both the child and the clinician, particularly if the wrong psychopharmacological agent is administered (Brown & Sawyer, 1998). Although lengthy assessments including structured clinical interviews and naturalistic behavioral observations over time would be ideal, modern day clinical practice favors accuracy, brevity, and cost-efficiency (Starfield et al., 1994). Thus, as the demand for clinical services for youth rises, the importance of time and cost efficiency in diagnostic assessment rises as well.

As a result of the push to increase the efficiency of these diagnostic decisions in order to ensure the most cost-effective treatment (Groth-Marnat, 1999), measures that may help focus assessment of child and adolescent psychopathology and facilitate these difficult differential diagnoses have begun to receive much attention in the literature (e.g., Geller, Warner, Williams, & Zimerman, 1998). For example, the Child Behavior Checklist's (CBCL; Achenbach, 1991) ability to identify mean differences between bipolarity and ADHD (Geller, Warner, Williams, & Zimerman, 1998); the CBCL's discriminative ability between bipolarity and ADHD (Biederman et al., 1995); and the CBCL's overall diagnostic accuracy (Chen, Faraone, Biederman, & Tsuang, 1994) have been evaluated. Although the CBCL appears to perform fairly well in distinguishing bipolar from nonbipolar youths (Biederman et al., 1995), bipolar patient profiles showed elevations on several CBCL scales. Thus, it is not clear whether the CBCL was identifying bipolar disorder per se, or simply indicating that bipolar youths present with greater levels of impairment in general.

The primary purpose of this investigation is to evaluate another such measure on its ability to accurately distinguish among disorders in adolescents: The General Be-

havior Inventory (GBI; Depue, 1987; Depue et al., 1981). The GBI is a self-report inventory with items focusing on mood-related behaviors, including depressive, hypomanic, and biphasic symptoms. Specifically, the GBI is composed of two subscales, the Depressive scale and the Hypomanic–Biphasic scale, which were rationally derived on the basis of a two-dimensional model of these symptoms proposed by the instrument authors (Depue, 1987; Depue, Krauss, Spoont, & Arbisi, 1989). Two exploratory factor analyses (EFA) have been published on the GBI. The first EFA, conducted by Depue et al. (1981), resulted in a single, dominant factor underlying the data. However, evidence of a two-factor structure of the GBI was reported in the second EFA, published by the authors of the present study (Youngstrom, Findling, Danielson, & Calabrese, 2001). Excellent reliability and good discriminative validity of the Depressive and Hypomanic scales have been demonstrated in their use with adult populations (e.g., Depue et al., 1989; Klein, Dickstein, Taylor, & Harding, 1989; Mallon, Klein, Bornstein, & Slater, 1986). In addition, recent investigations have demonstrated that a parents' report of their child's symptoms on the GBI (where items on the GBI were modified to read "your child" instead of "you") provides clinically useful information about the child and can be used to help a clinician make differential diagnoses among bipolar versus disruptive disorders and unipolar versus bipolar depression (Youngstrom et al., in press). Despite the GBI's usefulness with adult self-report about own behavior and with parent report about child behavior, minimal information is known about its adequacy with youth when the youths themselves are the reporters. To the authors' knowledge, one study to date has reported on the GBI's utility in an older adolescent sample (ages 15–21; Klein, Depue, & Slater, 1986), and no studies have reported on the GBI's validity in a sample that includes youth of younger ages as in the present study (ages 10–17). Specifically, Klein et al. (1986) administered the GBI to 37 offspring (15–21 years) of Bipolar I patients and 21 offspring (15–21 years) of 15 psychiatric patients without affective disorders, which along with demonstrating a familial relationship between cyclothymia and bipolar disorder as identified by the GBI, also supported the clinical specificity and sensitivity of the GBI.

Two arguments can be made in support of establishing valid youth self-report measures of behavior and psychopathology. First, although parents traditionally have been the most popular source of information regarding young children's behavior, some research has demonstrated that clinicians actually prefer accurate youth report when assessing internalizing disorders (Loeber, Green, & Lahey, 1990). This preference has been attributed to the

youth's direct access to feelings, as opposed to the parent's inference of these feelings. Second, biases and other problems, such as depression in the parent, can adversely affect the accuracy of a parental report regarding a child's behavior (Youngstrom, Izard, & Ackerman, 1999; Youngstrom, Loeber, & Stouthamer-Loeber, 2000; cf. Richters, 1992). Hence, it is necessary to determine which instruments are the most psychometrically sound and suitable for youths to complete when assessing psychopathology and which provide the most useful information to the clinician.

Despite the evidence that youth report can be important and useful, minimal studies have looked at a youth self-report measure in terms of its ability to help with differential diagnoses. Among the studies examining the use of youth report in differential diagnosis, the majority have focused on determining the severity of a specific disorder or distinguishing among disorders within the same spectrum (e.g., Clark & Bolton, 1985; Fine, Moretti, Haley, & Marriage, 1985; Steinhausen & Metzke, 2000). In addition, we know of only one study looking at the use of adolescent report in distinguishing between different families of symptoms. That study, which involved the development of a child report version of an emotion scale, investigated the scale's ability to differentiate between symptoms of anxiety and depression in children (Laurent et al., 1999). It is important to note that none of these studies included the GBI, and only one of the studies tested a measure's ability to differentiate among actual *DSM* diagnoses determined by a structured clinical interview (Fine et al., 1985).

Thus, the present study focused on investigating the youth's report on the GBI. Our first hypothesis was that the measure would conform to the two-dimensional model proposed by Depue, as was found when using parent report about the youth's behavior on the GBI. That is, a Depression factor and a Hypomanic-Biphasic factor would be indicated by the youths' responses on the GBI.

The second hypothesis was that the two subscales of the GBI, the Depression and the Hypomanic-Biphasic scales, would statistically discriminate among different diagnostic groups. It would be most useful to consider how the GBI could be used to make common clinical decisions and distinctions. Because the GBI is specific to mood-related behaviors, it follows that it should be useful in making distinctions between youths with mood disorders and youths without mood disorders. In settings where youths often present with symptoms similar to mood disorders (e.g., parents often report observing irritability in their adolescents), a clinician might want to use the measure to help rule out a mood disorder. A clinician also might want to make a more specific differential diagnosis; for example, does this youth have an internalizing (i.e.,

mood) or externalizing (i.e., disruptive behavior) disorder? Thus, we investigated whether the GBI scales could discern: (1) any mood disorders versus no mood disorders, and (2) any mood disorder versus disruptive behavior disorders. Additionally, as indicated earlier, bipolar disorder is particularly difficult to differentiate from disorders that share similar symptoms, including ADHD and CD (i.e., disruptive behavior disorders) and unipolar depression. As such, the second two comparisons on which the GBI scales were tested included (3) bipolar disorders versus disruptive behavior disorders, and (4) bipolar disorders versus unipolar depressive disorders. With regard to the latter comparison, the researchers hypothesized that only one of the two subscales of the GBI would be useful in differentiating among the disorders. That is, because higher levels of dysphoric mood are a component of both unipolar and bipolar depressive disorders, it was expected that the Depression scale would not significantly differentiate between the two disorder domains, but the Hypomanic–Biphasic Scale would.

The third hypothesis posed by the researchers is related to how current research in assessment is emphasizing the importance of quantifying the information value of each instrument using techniques such as Receiver Operating Characteristic (ROC) Curves (McFall & Treat, 1999). Thus, in evaluating the accuracy of the GBI scales as an assessment device by quantifying the information value in terms of diagnostic efficiency, it was hypothesized that the scales would have at least moderate accuracy. Moderate accuracy was predicted because the scales were found to have medium accuracy when making similar comparisons with the parent report GBI (Youngstrom et al., in press). It is important to note that moderate accuracy in conjunction with low cost and low demand could still yield a clinically useful instrument (Kraemer, 2000).

METHOD

Participants

Participants were 197 youths ages 10–17 years ($M = 13.54$, $SD = 1.72$) presenting at a midwestern urban outpatient clinic specializing in the treatment of mood disorders and psychopharmacology. This age range was selected because youths younger than 10 may not be able to complete and fully comprehend the self-report items and individuals older than 17 begin to fall in the adult range. A trained member of the research team read the items of the GBI aloud to those youth whose reading skills were questionable. Fifty-two percent ($n = 103$) were male and 79.2% ($n = 156$) were Caucasian, 16.8% African American, 2% Latino, and 2% of other ethnicity.

Procedure

Assent was provided by all youths who participated in the study, and written informed consent was obtained from their legal guardians. All participants and their families completed an intake assessment, which involved a structured clinical interview of the youth by a research assistant using the Schedule for Affective Disorders and Schizophrenia for School-Age Children—Present and Lifetime (K-SADS-PL or K-SADS-E; Kaufman et al., 1997) and trained to a high level of interrater reliability ($\kappa > .85$). In addition, a child and adolescent psychiatrist reviewed the results of all K-SADS interviews. Youths completed the GBI, describing their own behavior as part of the assessment process. As determined by the structured interview, current diagnoses of the youths were as follows: 67 youths (34%) met the criteria for a unipolar mood disorder (nine with Major Depression and one with Depression Not Otherwise Specified); 39 youths (19.8%) met the criteria for Bipolar I; 34 youths (17.3%) met the criteria for Bipolar II, Cyclothymia, or Bipolar Not Otherwise Specified Disorder; 35 youths (17.8%) for disruptive behavior disorders (19 with ADHD, 1 with Oppositional Defiant Disorder, and 1 with CD); 6 youths (3%) for "other diagnoses" (i.e., 1 with Posttraumatic Stress Disorder, 1 with Generalized Anxiety Disorder, 2 with Schizophrenia, and 2 with Schizoaffective Disorder); and 16 youths (8.1%) did not meet the criteria for any Axis I disorder. Youths with primary Axis I disruptive behavior disorders did not have any comorbid mood disorders, whereas 41.4% of the youths diagnosed with a mood disorder also were diagnosed with a disruptive behavior disorder, with the majority ($n = 28$) having a secondary diagnosis of ADHD. In addition, 42.9% of the youths with a primary disruptive behavior diagnosis also had a comorbid secondary disruptive behavior disorder diagnosis.

The General Behavior Inventory (GBI)

The GBI is a 73-item instrument that asks the reporter to rate unipolar depression, hypomania, and biphasic symptoms over the past week. Responses are given on a 4-point Likert scale, with 0 being *never or hardly ever* and 3 being *very often or almost constantly*. Thus, high scores represent greater psychopathology. As recommended by Depue (1987), items can be scored using a dichotomous model, dividing a population into cases and noncases, where those individuals responding 0 or 1 to an item receive 0 points and those responding 2 or 3 to an item receive 1 point. The scale can also be scored in the original Likert fashion, where the subject's responses are merely added. The authors chose to use the original Likert scaling because the Likert scale preserves more in-

formation. For example, it may be useful to know whether a youth never experiences feelings of mania (rating of 0) or whether the youth sometimes experiences feelings on mania (rating of 1). Also, Likert scaling increases the variability in observed scores and increases the reliability of the obtained scales (Pedhazur & Schmelkin, 1991). The scoring procedure recommended by Depue (1987) yields a Depression score based on 46 items and a Hypomanic–Biphasic scale based on 28 items (item 44 is included on both scales).

RESULTS

Missing Data

One hundred and forty-four cases had complete data, 35 were missing only 1 of the 73 items, 9 were missing 2 items, 1 was missing 3 items, and 8 were missing 5 or more items. There did not seem to be a systematic pattern to which particular items were skipped, although item 70 (an item about loss of sexual interest) was skipped more than any other item (i.e., was skipped by 15 youths). In order to reduce the effects of loss of power as a result of missing data points, cases were included in the analyses of the two subscales of the GBI if 95% of the items in that scale were completed.

Exploratory Factor Analysis

The items were grouped into 20 homogeneous parcels in order to maintain an adequate subject to variable ratio and improve the reliability of the factor indicators. Three to four homogeneous items were assigned to each parcel by three independent raters (CKD, EAY, RLF) with 98% agreement. The parcels demonstrated good internal consistency, with alphas ranging from .60 to .89. Corrected item–total correlations ranged from .21 (item 70) to .79 (item 56). Of the 20 parcels, 8 parcels contained items with hypomanic–biphasic content and 12 parcels contained items with depressive content. The parceling procedure has been recommended by many methodologists for several reasons, including reducing the number of variables, improving the subject-to-variable ratio, improving the reliability of the input indicators (i.e., Spearman–Brown prophecy), and combining clinical/rational and empirical standards in developing the parcels (Fabrigar, Wegener, MacCallum, & Strahan, 1999).

In order to explicitly test the number of factors required, two statistical decision rule techniques, Horn's Parallel Analysis (HPA; Horn, 1965) and Minimum Average Partial method (MAP; Velicer, 1976), both of which have demonstrated better accuracy than all other

Table I. Items, Internal Consistencies, and Factor Loadings of the Parcels on the GBI ($N = 197$)

		Factor loading	
Parcel	α	Depression	Hypomanic–Biphasic
Biphasic			
Parcel 1 (Items 2, 24, 35, 48)	.78	.63	.36
Parcel 2 (Items 19, 40, 53)	.85	.44	.45
Hypomanic			
Parcel 3 (Items 4, 7, 15)	.71	.00	.76
Parcel 4 (Items 22, 30, 31, 66)	.73	.01	.80
Parcel 5 (Items 11, 17, 42, 51)	.64	.27	.61
Parcel 6 (Items 27, 44, 54)	.70	.55	.35
Parcel 7 (Items 8, 57, 64)	.64	.48	.42
Parcel 8 (Items 38, 43, 46, 61)	.69	.00	.76
Depression			
Parcel 9 (Items 3, 23, 45, 63)	.88	1.00	−.13
Parcel 10 (Items 47, 56, 62, 73)	.89	.99	−.25
Parcel 11 (Items 9, 10, 13, 70)	.64	.68	.17
Parcel 12 (Items 21, 33, 49, 59)	.80	.75	.19
Parcel 13 (Items 1, 12, 41)	.70	.75	.19
Parcel 14 (Items 5, 25, 37, 52)	.79	.53	.35
Parcel 15 (Items 14, 39, 55)	.77	.83	.00
Parcel 16 (Items 29, 36, 50, 71)	.74	.69	.25
Parcel 17 (Items 16, 60, 65, 67)	.60	.69	.17
Parcel 18 (Items 20, 32, 34, 72)	.86	.97	−.01
Parcel 19 (Items 18, 26, 58, 68)	.71	.65	.30
Parcel 20 (Items 6, 28, 69)	.73	.87	−.12

Note. Factor loadings based on PAF extracting two factors. Copies of the GBI are available from Richard Depue (rad5@cornell.edu) or Eric Youngstrom (eay@po.cwru.edu).

commonly used retention decision rules (Velicer, Eaton, & Fava, 2000; Zwick & Velicer, 1986), were applied to the data. These procedures are described in greater detail elsewhere (see Youngstrom et al., 2001). A two-factor solution was indicated using both decision rules. Then, in order to determine the congruence between the two factors that had resulted from the data and the two scales rationally constructed by Depue, a principal axis factor analysis of the 20 parcels was conducted using an oblique rotation (called "oblimin" in SPSS). The solution was similar to the expected Depressive and Hypomanic format, with the exception of two hypomanic–biphasic parcels (parcels 1 and 6), which loaded on the Depression scale (factor 1). Two other hypomanic–biphasic parcels (parcels 2 and 7) also loaded on both factors. Table I presents the items, internal consistencies, and factor loadings from the pattern matrix for each parcel. Internal consistency was high for both the Depression scale ($\alpha = .97$) and the Hypomanic–Biphasic scale ($\alpha = .94$). The correlation between the two factors was high ($r = .62$).

Discriminative Validity

On the basis of their Axis I primary diagnoses, the subjects were divided into five diagnostic groups: (1) any

mood disorders (youths with any bipolar disorder or unipolar depressive disorder primary diagnosis); (2) no mood disorders (youths with no mood disorder diagnosis whatsoever); (3) bipolar disorders; (4) disruptive behavior disorders (youths with ADHD, CD, or Oppositional Defiant Disorder as a primary diagnosis); or (5) unipolar depression (youths with Major Depression, Dysthymia, or Depression Not Otherwise Specified as a primary diagnosis). Table II presents the means and standard deviations on the Depression and the Hypomanic–Biphasic scales for each group.

One-way analyses of variance and Games–Howell post hoc tests (which do not assume equal variances or sample sizes across groups) were computed to determine reliable post hoc group differences among the different diagnostic categories. Significant mean differences found among the groups on the Depression and the Hypomanic–Biphasic scales are noted on Table II. Specifically, significant mean differences were found between the any mood and the no mood and the disruptive behavior disorder groups on the Depression scale, with youths with any mood disroder scoring on average over 25 points higher than the no mood and the disruptive behavior disorders group. In addition, the bipolar spectrum group and unipolar group each scored significantly higher than the

Table II. Effect Sizes (Cohen's d) for Group Comparisons

	Any mood (n = 138)		No mood (n = 57)		Bipolar (n = 71)		Disruptive behavior disorder (n = 35)		Unipolar (n = 67)	
	D	HB	D	HB	D	HB	D	HB	D	HB
Any mood	**46.68** **(29.57)**	**26.32** **(15.48)**	−0.95**	−0.79**	0.07	0.23	−0.94**	−0.74**	0.07	0.26
No mood			**21.22** **(19.06)**	**15.00** **(11.69)**	0.96**	1.02**	−0.03	0.04	1.08**	0.61*
Bipolar					**44.70** **(28.36)**	**29.90** **(16.82)**	−0.94**	−0.96**	0.14	−0.49**
Disruptive behavior disorder							**20.74** **(18.84)**	**15.46** **(10.94)**	1.06**	0.58*
Unipolar									**48.77** **(30.89)**	**22.57** **(13.04)**

Note. D = Depression scale; HB = Hypomanic–Biphasic scale. Means and standard deviations in bold; standard deviations in parentheses.
$^*p < .05$. $^{**}p < .01$. Denotes significant mean differences per Games–Howell post hoc test.

disruptive behavior group and the no mood disorder group on the Depression scale. On the Hypomanic-Biphasic scale, the bipolar group scored significantly higher than all other groups, with the exception of those youths with any mood disorder, which included the youths in the bipolar group. Further, youths with unipolar depression scored significantly higher than those youths in the no mood disorders group and those youths in the disruptive behavior group on the Hypomanic-Biphasic scale. In order to illustrate the magnitude of the group differences more clearly, effect sizes as calculated by Cohen's d for the mean differences among the diagnostic groups on the GBI scales are presented in Table II.

Although mean differences and effect sizes provide information about *group* categorization, other statistical procedures are necessary to assess how well a measure

can provide information with regard to *individual* classification. Logistic regression and ROC Curve analyses were conducted for these purposes. Logistic regression analyses tested the extent to which the two subscales could differentiate between clinical diagnostic categories based on the structured clinical interview and successfully classify each case in its respective category (see Table III). Results demonstrated that the GBI did the best at separating those youths with any mood disorder as the primary diagnosis from those with no mood disorders. In comparing youths with any mood disorder with youths that have a disruptive behavior disorder, the two groups were well differentiated by the GBI. Also, youths with a bipolar disorder versus those with a disruptive behavior disorder were well discriminated by the GBI, as were those youths with unipolar depression versus those with bipolar

Table III. Performance of the GBI Separating Among Disorders Using Logistic Regression

			Regression weights		
Comparison groups	$\chi^2(2\,df)$	Nagelkerke R^2	Depression	Hypomanic– Biphasic	Constant
Any mood (n = 137) vs. no mood disorders (n = 57)	36.29***	.24	.039**	.005	−2.40***
Any mood disorders (n = 137) vs. disruptive behavior disorders (n = 35)	26.15***	.22	.044*	−.003	−1.92*
Bipolar disorders (n = 73) vs. disruptive behavior disorders (n = 35)	22.01***	.26	.017	.048	−2.98**
Bipolar disorders (n = 73) vs. unipolar depression (n = 67)	33.41***	.29	−.056***	.123***	−1.42*

$^*p < .05$. $^{**}p < .005$. $^{***}p < .001$, two-tailed.

disorders. All four comparisons accounted for adequate variability, as indicated by Nagelkerke R^2 estimates ranging from .22 to .29. In addition, one advantage of logistic regression is that researchers can use the procedure to determine whether separate scales of a measure are most useful in isolation or in combination. Unique contributions of the Depression scale and the Hypomanic–Biphasic scale in making the discriminations among the different diagnostic groups are indicated by the individual regression weights computed in the logistic regression analyses. The only comparison of diagnostic groups in which unique, meaningful information was provided by both scales was in discriminating bipolar from unipolar depression. The regression weights for the two scales for each of the four diagnostic group comparisons are provided in Table III.

Receiver Operating Characteristic (ROC) curves determined the relative value of the two GBI subscales in their ability to make the diagnostic group distinctions. ROC curves entail plotting the balance between the sensitivity (the probability of a positive test result, given that the adolescent has the target disorder) and specificity (the probability of a negative test result, given that the adolescent does not have the target disorder) of a diagnostic test while systematically moving the cut score across its full range of values. In an ROC Curve plot, the diagonal line demonstrates the "random ROC," which reflects a test with zero discriminating power. The accuracy of the ROC curve is quantified by calculating the Area Under the Curve (AUC). An AUC of .50 represents when a test's diagnostic performance is equal to chance, whereas an AUC of 1.0 indicates perfect diagnostic performance. The closer the AUC value to 1.0, the better the diagnostic accuracy of the test.

In the present study, ROC curves resulted in areas under the curve (AUC) significantly better than .50 in several comparisons. The Depression scale resulted in significant areas under the curve in comparing any mood disorder versus no mood disorder, any mood versus a disruptive behavior disorder, and bipolar versus disruptive behavior disorders, with AUCs ranging from .76 to .77. As hypothesized, the AUC in comparing unipolar versus bipolar depression on the Depression scale was not significant. It was not expected that youths in the unipolar and bipolar depression groups would be well discriminated by the Depression scale, because both disorders include the symptoms of depression. The Hypomanic–Biphasic scale also had good diagnostic performance across all four comparisons: Any mood disorder versus no mood disorder, any mood versus a disruptive behavior disorder, bipolar versus disruptive behavior disorder, and unipolar versus bipolar depression, with AUCs ranging from .63 to .76. Figure 1 presents the ROC curves with the Area Under the Curve (AUC) values.

DISCUSSION

The purpose of this study was to investigate the GBI, using youth self-report, as an instrument to potentially aid diagnostic assessment of child and adolescent psychopathology concerning mood and behavioral problems. This study differs in three ways from prior studies that have looked at use of youth report in differential diagnosis. First, the focus of this study is on the GBI, and the GBI only has been applied to one older adolescent sample (Klein et al., 1986). Second, the youths were placed into diagnostic categories based on a structured clinical interview administered by highly trained research assistants and the interview results were reviewed by child and adolescent psychiatrists. To the author's knowledge, this rigorous method of diagnosis, which increases the reliability of the diagnostic categorization, has only been done in one prior published study looking at differential diagnosis of unipolar depressive disorders in youths using self-report instruments (Fine et al., 1985). Third, the GBI was examined through statistical procedures that provide useful information about diagnostic accuracy and efficiency. These techniques, including logistic regression and ROC analyses, have yet to achieve common usage in evaluation of youth report measures of psychopathology.

Post hoc mean comparisons demonstrated that the diagnostic groups significantly differed on the GBI's subscales, Depression and Hypomanic–Biphasic symptoms. The differences between the groups were in an expected direction on the Hypomanic–Biphasic scale, with the bipolar group scoring significantly higher than the no mood disorder group, the disruptive behavior disorder group, and the unipolar group. Significant differences also resulted among the groups with regard to the four predicted comparisons on the Depressive scale, with the exception of the unipolar versus bipolar groups, in which no significant difference was found. This indicates that youths with bipolar disorders experience substantial depressive symptomatology as well.

As indicated by the logistic regression and ROC analyses, the GBI may be useful for individual classification of the youths with regard to difficult diagnostic comparisons, such as bipolar disorders versus disruptive behavior disorders. Meaningful information was provided by the logistic regression with regard to whether unique information could be ascertained from the Depression and the Hypomanic–Biphasic scales individually, or if the two GBI scales were most useful in combination. In this study, the Depression scale in isolation provided unique information that may be useful in separating youths with mood disorders from youths with no mood disorders, including those youths with disruptive behavior disorders. Although
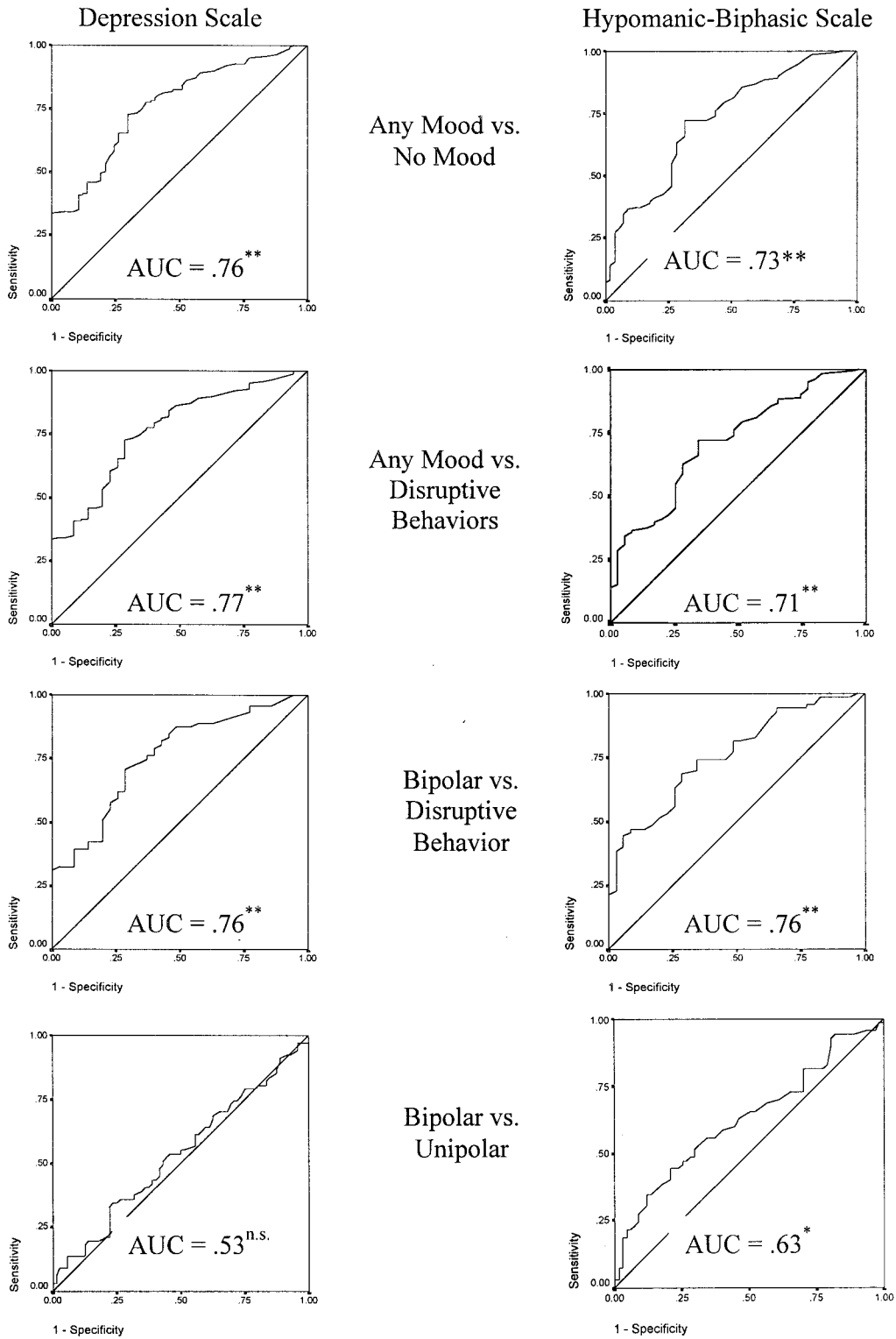
Fig. 1. ROC curves for the GBI subscales for comparisons between diagnostic groups. *Note.* $*p < .01$. $**p < .001$, two-tailed.

useful information for the discrimination of youths with bipolar disorders from youths with disruptive behavior disorders can be gathered from either the Depression or the Hypomanic–Biphasic scale, the information provided by the scales is virtually redundant with each other. In other words, neither scale seems to provide information that the other scale does not provide. However, the combination of both scales was most useful in discriminating bipolar disorders from unipolar depressive disorders. That is, both scales contribute unique information is the predication, indicating that even though depression is a symptom of both disorders, the Depression scale provides additional information about potential unipolar depression after controlling for an individual's level of Hypomanic–Biphasic symptoms.

Additionally, ROC curves showed that moderate information value is provided by the GBI in making these discriminations that often have proved difficult for clinicians using other sources of data. The several significant AUCs indicated that both subscales of the GBI perform substantially better than chance at ranking the youth in their appropriate group. In general, by observing the AUCs, it appeared that youths who are diagnosed with a particular type of disorder usually will score in the direction on the GBI subscales that is expected for that disorder. For example, the AUC of the ROC Curve on the Depression subscale in comparing any mood disorder versus no mood disorder was .76. This means that in random pairings, a youth diagnosed with a mood disorder will score higher than a youth with no mood disorder 76% of the time. The one comparison that resulted in a nonsignificant AUC involved comparing unipolar depression versus bipolar disorders on the Depression subscale. It is important to note that this was expected, as depression is a core feature of both disorders.

Three limitations are salient. First, the great majority of the participants in the study had presented at an outpatient clinic for evaluation. This study was done under the auspices of a research program in the Division that either (1) emphasized psychopharmacological studies or (2) assessed and treated youths with bipolar spectrum disorders. Therefore this is not an epidemiological sample and may not be representative of a clinical population at large. Thus, the sample is likely biased to an unknown extent.

Second, the authors chose to use Likert-type scoring instead of the dichotomous "case scoring," which is more commonly used, to construct the GBI scales. The decision to use the Likert-type scoring was partially based on Depue and colleagues' recommendation to use Likert-type scores in research applications (Depue et al., 1981). Also, established psychometric principles demonstrate that Likert-based scores would be as or more reliable than dichotomized items. Because the GBI can be used to make decisions about individuals, it is important to preserve as much reliable information as possible in the scores.[4]

Third, this study relied on youth report, without similar measures from parents, teachers, and staff being presented. Many have found that youth can be reliable reporters of their own behavior and perceptions in a variety of contexts (Herjanic, Herjanic, Brown, & Wheatt, 1975; Witt, Cavell, Heffer, Carey, & Martens, 1988). Others, however, question the disparity between a child's report and his or her parent's or teacher's report, as a child's perception of himself or herself does not consistently demonstrate optimal levels of convergence with views of the other reporters who know the child well. Nevertheless, this disparity does not seem to be a function of the child's inability to be a reliable reporter of his or her characteristics; the internal consistency and short-term stability of a child's self-rating has been reported to approach the levels found for most adults (Weinberger, 1996). Future research should address identifying optimal strategies for combining youth report with other sources of information to maximize overall accuracy in the assessment and diagnosis of depression, bipolar disorder, and disruptive behavior disorders. Also, the results of the present study should be replicated in order to cross-validate the findings.

In sum, the GBI appears to be a promising instrument in the assessment of psychopathology in youths in both an empirical and a practical sense. Analyses demonstrated that the GBI did well at differentiating among disorders, and thus can be helpful in making differential diagnoses in youths (see Findling et al., 2002 for a development of potential cut scores and applications of logistic regression results). The GBI is quick, typically requiring 5–10 min to complete. The instrument is also relatively easy to administer, either as an independent self-report, or, in instances where a youth's literacy skills are of concern, as a measure that is read to the youths by staff. Additionally, the GBI, using youth report, provides something unique that cannot be gained from measures where the parent, teacher, or clinician is the reporter: The youth's own perspective. Youth report can be useful when dealing with a mood disorder (i.e., when the clinician would prefer to hear the child's direct report of their inner feelings and experiences as opposed to through an indirect source, such as a parent) or when a parent's report is suspected to be biased; such can be the case when he or she is depressed. The present study offers one instrument, the GBI, as a means by which the clinician can efficiently gather information from a child to help make difficult diagnostic decisions.

---

[4]Analyses also were performed using case scoring. The results using the case scoring did not change in any meaningful way.

REFERENCES

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

Baldessarini, R. J., & Tondo, L. (2000). Treatment of bipolar disorder. *Clinical Reviews and Updates in Neurology and Psychiatry, 2,* 9–12.

Biederman, J. (1995). Developmental subtypes of juvenile bipolar disorder. *Harvard Review of Psychiatry, 3,* 227–230.

Biederman, J. (1998). Resolved: Mania is mistaken for ADHD in prepubertal children. *Journal of the American Academy of Child and Adolescent Psychiatry, 37,* 1091–1093.

Biederman, J., Russell, R., Soriano, J., Wozniak, J., & Faraone, S. V. (1998). Clinical features of children with both ADHD and mania: Does ascertainment source make a difference? *Journal of Affective Disorders, 51,* 101–112.

Biederman, J., Wozniak, J., Kiely, K., Ablon, S., Faraone, S., Mick, E., et al. (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 and Adolescent Psychiatry, 34,* 464–471.

Bowring, M. A., & Kovacs, M. (1992). Difficulties in diagnosing manic disorders among children. *Journal of the American Academy of Child and Adolescent Psychiatry, 31,* 611–614.

Brown, R. T., & Sawyer, M. G. (1998). *Medications for school-age children: Effects on learning and behavior.* New York: Guilford Press.

Carlson, G. A. (1998). Mania and ADHD: Comorbidity or confusion. *Journal of Affective Disorders, 51,* 177–187.

Chen, W. J., Faraone, S. V., Biederman, J., & Tsuang, M. (1994). Diagnostic accuracy of the Child Behavior Checklist Scales for attention-deficit hyperactivity disorder: A receiver-operating characteristic analysis. *Journal of Consulting and Clinical Psychology, 62,* 1017–1025.

Clark, D. A., & Bolton, D. (1985). An investigation of two self-report measures of obsessional phenomena in obsessive–compulsive adolescents: Research note. *Journal of Child Psychology and Psychiatry and Allied Disciplines, 26,* 429–437.

Depue, R. A. (1987). *General Behavior Inventory* (manual). Minneapolis: University of Minnesota.

Depue, R. A., Krauss, S., Spoont, M. R., & Arbisi, P. (1989). General Behavior Inventory identification of unipolar and bipolar affective conditions in a nonclinical university population. *Journal of Abnormal Psychology, 98,* 117–126.

Depue, R. A., Slater, J. F., Wolfstetter-Kausch, H., Klein, D., Goplerud, E., & Farr, D. (1981). A behavioral paradigm for identifying persons at risk for bipolar depressive disorder: A conceptual framework and five validation studies. *Journal of Abnormal Psychology Monograph, 90,* 381–487.

Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods, 4,* 272–299.

Findling, R. L., Youngstrom, E. A., Danielson, C. K., DelPorto, D., Papish David, R., Townsend, L., & Calabrese, J. R. (2002). Clinical decision-making using the General Behavior Inventory in juvenile bipolarity. *Bipolar Disorders, 4,* 34–42.

Fine, S., Moretti, M., Haley, G., & Marriage, K. (1985). Affective disorders in children and adolescents: The dysthymic disorder dilemma. *Canadian Journal of Psychiatry, 30,* 173–177.

Geller, B., Warner, K., Williams, M., & Zimerman, 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,* 93–100.

Geller, B., Warner, K., Williams, M., Zimerman, B., Frazier, Z., Beringer, L., et al. (1998). Prepubertal and early adolescent bipolarity differentiate from ADHD by manic symptoms, grandiose delusions, ultra-rapid or ultradian cycling. *Journal of Affective Disorders, 51,* 81–91.

Giedd, J. N. (2000). Bipolar disorder and attention-deficith/hyperactivity disorder in children and adolescents. *Journal of Clinical Psychiatry, 61,* 31–34.

Grom-Marnat, G. (1999). *Handbook of psychological assessment* (3rd ed.). New York: Wiley.

Herjanic, B., Herjanic, M., Brown, F., & Wheatt, T. (1975). Are children reliable reporters? *Journal of Abnormal Child Psychology, 3,* 41–48.

Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika, 30,* 179–185.

Kaufman, J., Birmaher, B., Brent, D., Rao, U., Flynn, C., Moreci, P., et al. (1997). Schedule for affective disorders and schizophrenia for school-age children-present and lifetime version (K-SADS-PL): Initial reliability and validity data. *Journal of the American Academy of Child and Adolescent Psychiatry, 36,* 980–988.

Klein, D. N., Depue, R. A., & Slater, J. F. (1986). Inventory identification of cyclothymia IX. Validation in offspring of bipolar I patients. *Archives of General Psychiatry, 43,* 441–445.

Klein, D. N., Dickstein, S., Taylor, E. B., & Harding, K. (1989). Identifying chronic affective disorders in outpatients: Validation of the General Behavior Inventory. *Journal of Consulting and Clinical Psychology, 57,* 106–111.

Kovacs, M. (1987). Diagnosis of depressive disorders in children: An interim appraisal of the pertinent *DSM-III* categories. In G. L. Tischler (Ed.), *Diagnosis and classification in psychiatry: A critical appraisal of DSM-III* (pp. 369–383). New York: Cambridge University Press.

Kraemer, H. C. (2000). Statistical analysis to settle ethical issues? *Archives of General Psychiatry, 57,* 327–328.

Loeber, R., Green, S. M., & Lahey, B. B. (1990). Mental health professionals' perception of the utility of children, mothers, and teachers as informants on childhood psychopathology. *Journal of Clinical Child Psychology, 19,* 136–143.

Mallon, J. C., Klein, D. N., Bornstein, R. F., & Slater, J. F. (1986). Discriminant validity of the General Behavior Inventory: An outpatient study. *Journal of Personality Assessment, 50,* 568–577.

McFall, R. M., & Treat, T. A. (1999). Quantifying the information value of clinical assessment with signal detection theory. *Annual Review of Psychology, 50,* 215–241.

Pedhazur, E., & Schmelkin, L. P. (1991). *Measurement, design, and analysis: An integrated approach.* Hillsdale, NJ: Erlbaum.

Richters, J. E. (1992). Depressed mothers as informants about their children: A critical review of the evidence for distortion. *Psychological Bulletin, 112,* 485–499.

Starfield, B., Powe, N. R., Weiner, J. R., Stuart, M., Steinwachs, D., Scholle, S. H., et al. (1994). Costs vs. quality in different types of primary care settings. *JAMA, 272,* 1903–1908.

Steinhausen, H., & Metzke, C. W. (2000). Adolescent self-rated depressive symptoms in a Swiss epidemiological study. *Journal of Youth and Adolescence, 29,* 427–440.

Velicer, W. F. (1976). Determining the number of components from the matrix of partial correlations. *Psychometrika, 41,* 321–327.

Velicer, W. F., Eaton, C. A., & Fava, J. L. (2000). Construct explication through factor or component analysis. A review and evaluation of

alternative procedures for determining the number of factors or components. In R. O. Goffin & E. Helmes (Eds.), *Problems and solutions in human assessment: A festschrift to Douglas Jackson at seventy* (pp. 41–71). Hillsdale, NJ: Erlbaum.

Weinberger, D. A. (1996). Distorted self-perceptions: Divergent self-reports as statistical outliers in the multimethod assessment of children's social–emotional adjustment. *Journal of Personality Assessment*, *66,* 126–143.

Weller, E. B., Weller, R. A., & Fristad, M. A. (1995). Bipolar disorder in children: Misdiagnosis, underdiagnosis, and future directions. *Journal of the American Academy of Child and Adolescent Psychiatry*, *34,* 709–714.

Winokur, G., Coryel, W., Keller, M., Endicott, J., & Akishal, H. (1993). A prospective follow-up of patients with bipolar and primary unipolar affective disorder. *Archives of General Psychiatry*, *50,* 457–465.

Witt, J., Cavell, T., Heffer, R., Carey, M., & Martens, B. (1988). Child self-report: Interviewing techniques and rating scales. In E. Shapiro & T. Kratochwill (Eds.), *Behavioral assessment in schools: Conceptual foundations and practical applications* (pp. 384–454). New York: Guilford Press.

Youngstrom, E. A., Findling, R. L., Danielson, C. K., & Calabrese, J. (2001). Discriminant validity of parent report of hypomanic and depressive symptoms on the General Behavior Inventory. *Psychological Assessment*, *13,* 267–276.

Youngstrom, E. A., Izard, C., & Ackerman, B. (1999). Dysphoria-related bias in maternal ratings of children. *Journal of Consulting and Clinical Psychology*, *67,* 1–12.

Youngstrom, E. A., Loeber, R., & Stouthamer-Loeber, M. (2000). Patterns and correlates of agreement between parent, teacher, and male youth behavior ratings. *Journal of Consulting and Clinical Psychology*, *68,* 1038–1050.