



HHS Public Access

Author manuscript

Psychol Assess. Author manuscript; available in PMC 2016 March 01.

Published in final edited form as:

Psychol Assess. 2015 March ; 27(1): 21–30. doi:10.1037/pas0000020.

Structural Invariance of General Behavior Inventory (GBI) Scores in Black and White Young Adults

Laura L. Pendergast,

Temple University

Eric A. Youngstrom,

University of North Carolina – Chapel Hill

Christopher Brown,

Temple University

Dane Jensen,

Temple University

Lyn Y. Abramson, and

University of Wisconsin-Madison

Lauren B. Alloy

Temple University

Abstract

In the United States, Black and White individuals show discrepant rates of diagnosis of bipolar disorder versus schizophrenia and antisocial personality disorder, as well as disparate access to and utilization of treatment for these disorders (e.g., Alegria, Chatterji et al., 2008; Chrishon et al., 2012). Such diagnostic discrepancies might stem from racially-related cognitive biases in clinical judgment or from racial biases in measurements of bipolar disorder. The General Behavior Inventory (GBI) is among the most well-validated and widely used measures of bipolar mood symptoms, but the psychometric properties of the GBI have been examined primarily in predominantly White samples. This study used multi-group confirmatory factor analyses (CFA) to examine the invariance of GBI scores across racial groups with a non-clinical sample. Fit was acceptable for tests of configural invariance, equal factor loadings, and equal intercepts, but not invariance of residuals. Findings indicate that GBI scores provide functionally invariant measurement of mood symptoms in both Black and White samples. The use of GBI scores may contribute consistent information to clinical assessments and could potentially reduce diagnostic discrepancies and associated differences in access to and utilization of mental health services.

Correspondence concerning this article should be sent to: Laura L. Pendergast, Laura.Pendergast@temple.edu.
Laura L. Pendergast, Department of Psychological, Organizational, and Leadership Studies in Education, Temple University; Eric A. Youngstrom, Department of Psychology, University of North Carolina – Chapel Hill; Christopher Brown and Dane Jensen, Department of Psychology, Temple University; Lyn Y. Abramson, Department of Psychology, University of Wisconsin-Madison; and Lauren B. Alloy, Department of Psychology, Temple University.

Keywords

Bipolar disorder; General Behavior Inventory (GBI); Race; Invariance

Bipolar disorders (BD) are recurrent psychiatric conditions that affect approximately 4% of the population and often result in long-term impairment and disability (Merikangas et al., 2007). Pharmacological treatment and psychotherapy have been shown to notably reduce symptoms of and impairment from BD (e.g., Fountoulakis & Vieta, 2008). However, access to appropriate treatment is dependent upon accurate assessment and diagnosis of (hypo)manic and depressive symptoms. The General Behavior Inventory (GBI; Depue, Krauss, Spont, & Arbisi, 1989) is a self-report screening questionnaire tapping symptoms of BD. Relative to other BD screening measures, the GBI has been touted as having “the most robust psychometric properties of the available self-report screeners” (Miller, Johnson, & Eisner, 2009). However, the GBI was developed with a predominantly White sample (Depue et al., 1981; Depue et al., 1989), and the validity of GBI scores with members of other racial groups is unknown. The purpose of this study is to examine the structural invariance of GBI scores in Black and White young adults.

Diagnosis of Bipolar Disorder

Research on assessment of BD is less extensive than that for unipolar mood disorders, but the need for effective, efficient diagnostic tools is no less significant. For example, individuals with BD show increased suicidality (Robins & Regier, 1991), substance abuse (Angst, 1998; Regier et al., 1990), difficulty with or loss of employment (Lopez, Mathers, Ezzati, Jamison, & Murray, 2006), and interactions with the criminal justice system (Cooke et al., 1996). Although early recognition and treatment can reduce the risk of negative outcomes (e.g., Geller et al., 1998), the average gap between onset of symptoms and diagnosis of BD is more than ten years. Multiple studies indicate that improper assessment of BD is partially responsible for the increased interval between onset of symptoms and proper formal diagnosis (e.g. Lish, Dime-Meenan, Whybrow, Price, & Hirschfeld, 1994; Mantere, Suominen, Leppamaki, Arvilommi, & Isometsa, 2004).

Racial Differences in Bipolar Disorder Diagnoses

BD is difficult to properly assess and diagnose in individuals from all backgrounds (Johnson, Miller, & Eisner, 2008). However, some research suggests that clinicians have particular difficulty accurately diagnosing BD in Black individuals. Findings from several studies indicate that Black individuals are under-diagnosed with BD and over-diagnosed with schizophrenia or antisocial personality (e.g., Chrishon et al., 2012; Chen, Alan, Swann, & Johnson, 1998; Kilbourne et al., 2004; Neighbors et al., 2007; Strakowski et al., 2003). Yet, other researchers have identified comparable rates of bipolar disorder between Black and White individuals when using structured interviews or assessments (Moreno et al., 2013). Notably, when racial/ethnic cues are removed and clinicians review data alone, Black participants are less likely to receive more severe and stigmatizing diagnoses (e.g., schizophrenia; Chrishon et al., 2012; Strakowski et al., 2003). Moreover, racial discrepancies in diagnosis may be mitigated through the use of structured diagnostic

interviews (e.g., Perron et al., 2010; Strakowski et al., 1995) or evidence-based algorithms for interpreting risk factors and rating scales (Jenkins et al., 2011). Presumably, clinicians who do not use structured diagnostic interviews or similar instrumentation rely more heavily on clinical judgment and, therefore, may be more likely to be influenced by racial stereotypes (Smedley, Stith, & Nelson, 2002) and other cognitive biases (see Croskerry, 2003; Dawes, Faust, & Meehl, 1989, for reviews).

Assessment of Bipolar Disorder

Although structured diagnostic interviews are considered to be the gold standard for BD assessment (Rogers, Jackson, & Cashel, 2001), they are often seen as impractical for use in clinical settings due to unrealistic time requirements. Self-report measures, on the other hand, require less time commitment and, if psychometrically sound, may help detect BD. However, most existing self-report measures report only modest sensitivity and specificity (e.g., Johnson et al., 2008; Miller, Klugman, Berv, Rosenquist, & Ghaemi, 2004; Ghaemi et al., 2005). Notably, questionnaires may also lead to racial discrepancies in bipolar diagnosis (Smith et al., 2013) and, as such, should be closely scrutinized in regard to fairness, bias, and validity across racial groups. The GBI is considered one of the best available self-report measures for assessing symptoms of BD (Alloy et al., 2008; Miller et al., 2009). In particular, the GBI has been shown to have strong sensitivity and specificity (Miller et al., 2009) as well as convergent validity with many other measures (e.g., structured interviews, fMRI findings; e.g., Bebko et al., 2014). The GBI is best used as a diagnostic aid to identify individuals who may benefit from more in-depth structured interviews in both clinical and research settings, although it also has shown significant prognostic value predicting future conversion to bipolar disorder (e.g., Alloy et al., 2012) and also has shown good sensitivity to treatment effects (Findling et al., 2003; Youngstrom, Zhao et al. 2013). The GBI has been studied extensively in both clinical- and community-based samples, and the majority of research on the GBI has focused on young adult populations (e.g., Depue, Krauss, Spont, & Arbisi, 1989; Klein, Dickstein, Taylor, & Harding, 1989).

Unfortunately, racial minorities have been severely underrepresented in the psychometric literature on the GBI. Existing studies of the psychometric properties of the 73-item self-report version of the GBI among adults have used racially homogenous samples wherein 92–100% of participants identified as White (Depue, Krauss, Spont, & Arbisi, 1989; Klein, Dickstein, Taylor, & Harding, 1989) or did not report participant race (Depue et al., 1981). Although an abbreviated (10-item) parent report version of the GBI has been shown to be effective in assessing pediatric bipolar disorder among Black and White participants in academic and community mental health settings (Freeman et al., 2012), the validity of the self-report form of the GBI for assessing bipolar disorder among Black adults still needs examination.

Present Study

For many decades, research findings have consistently revealed that using an actuarial diagnostic approach that integrates quantitative data, such as scores from behavioral rating scales, increases diagnostic accuracy (e.g., Dawes, Faust, & Meehl, 1989). However, the

superiority of actuarial approaches in clinical diagnosis rests upon the premise that the scales produce reliable and valid scores for members of the population with which they are used. The purpose of this study is to examine the structural invariance of GBI scores among Black and White young adults. Strong evidence of measurement invariance would indicate that the GBI could help reduce sources of bias in the assessment process. Conversely, significant discrepancies in the performance of the GBI or portions of it for different racial groups could help identify potential sources of differential rates of diagnosis of bipolar disorder apart from cognitive heuristics or clinician biases.

Method

Participants and Procedures

GBI data included in this study were originally used to screen non-referred undergraduate students for participation in the Longitudinal Investigation of Bipolar Spectrum Disorders (LIBS) project (see Alloy et al., 2008; 2012 for details). Participants were 291 Black and 994 White undergraduate students, 74% female, at a Northeastern urban university. Participants ranged in age from 17–45 (*Mdn* age=18 years, *SD*=2.06). No statistically significant gender or age differences were evident on total scores from either the (hypo)manic/biphasic or depressive subscales.

Participants were recruited through undergraduate courses, visits to residence halls, and on-campus advertisements. No compensation was provided for completion of the GBI, and most participants completed the scale in approximately 15 minutes. However, students who qualified for and participated in the LIBS project received payment for an additional diagnostic interview. This study was approved by Institutional Review Boards at both Temple University and the University of Wisconsin–Madison.

Measures

Overview of the GBI—The GBI (Depue, 1987) is a commonly used instrument designed to tap depressive and (hypo)manic symptoms in adults. The GBI is comprised of 73 items with high scores representing increased pathology. For each item, respondents use a 4-point Likert-type scale to indicate the frequency with which they experience a particular phenomenon (e.g., “*Have others said you seem down or lonely?*”) with one representing *never or hardly ever* and four representing *almost always or constantly*.

GBI Validity Evidence—In predominantly (i.e., 92–100%) White samples, GBI scores have shown excellent internal consistency and convergent and discriminative validity in nonclinical samples (e.g., Depue et al., 1989), samples of adults at-risk for bipolar disorders (Depue et al., 1985), and adult outpatient samples (e.g., Klein et al., 1989). However, the psychometric properties of the GBI never have been tested with samples representative of other racial groups. Moreover, few studies have explicitly examined the construct validity of the GBI using techniques such as factor analysis among members of any racial group.

The GBI items were developed and selected based on rational, rather than empirical, criteria. Items were designed to be consistent with *Diagnostic and Statistical Manual for Mental Disorders – Third Edition (DSM-III)*; American Psychiatric Association, 1980) and *Research*

Diagnostic Criteria (RDC) for bipolar disorder both of which describe a two-dimensional structure comprised of (hypo)manic/biphasic and depressive symptoms (Depue et al., 1981). Although extensive analyses were conducted to examine the predictive validity of GBI scores (e.g., Depue et al., 1981; Danielson, Youngstrom, Findling, & Calabrese, 2003), few studies have examined the structural validity. In an initial study of the structure of the GBI, Depue et al. (1981) determined that the majority of GBI items loaded on a single major factor. However, later studies employed more contemporary factor analytic techniques, as well as item parcels, and identified a two factor structure (hypomanic/biphasic and depressive factors) for the standard (Danielson et al., 2003) and adapted parent report versions (Youngstrom, Findling, Danielson, Kmett, & Calabrese, 2001) of the GBI. Notably, the GBI is well written from a clinical construct perspective, with deep roots in clinical observation, and has strong internal consistency and criterion validity (e.g., imaging, longitudinal predictive validity; Bebko et al., 2014) as well as discriminative validity (e.g., the signal detection and ROC analyses; Danielson et al., 2003). However, the GBI includes many double- and triple-barreled items, which make factor analysis difficult. These issues are not limited to this sample; they have occurred in every GBI sample with which we have worked. Fortunately, the parceling techniques developed in prior research (e.g., Danielson et al., 2003) allow for the examination of structural validity via factor analysis – despite the presence of double- and triple-barreled items. Although the evidence for a two-factor structure of the GBI is compelling and consistent with expectations based on theory and diagnostic criteria, the extent to which the two-factor structure is applicable across racial groups remains unclear.

GBI Item Parcels—Most researchers examining the GBI have relied on item parcels developed by Danielson et al. (2003). The GBI item parcels were developed in a study wherein three independent raters assigned GBI items to parcels with 98% agreement, and each parcel was determined to be unidimensional based on confirmatory factor analyses (CFAs; Danielson et al., 2003). The parcels were confirmed to be unidimensional when parents completed the items to describe their child or adolescent’s mood (Youngstrom et al., 2001). Likewise, confirmatory factor analyses (CFAs) conducted as part of this study supported the unidimensionality of items within each parcel. (Findings from one-factor CFAs are available from the first author.) Table 1 reports parcel item content and internal consistency. Internal consistency estimates obtained in this study are comparable to those obtained in prior research (e.g., Youngstrom et al., 2001; Danielson et al., 2003). This method of item parceling is considered to be particularly useful for combining clinical and empirical standards in scale development and validation (Fabrigar, Wegener, MacCallum, & Strahan, 1999). Parceling also helps integrate items that ask directly about *changes* in mood or energy, aggregating items asking about changes versus having them cross-load or show considerable correlated disturbance terms with other items that assess the manic or depressive content in relative isolation. This is particularly important for what Depue et al. (1981) called the “biphasic” items, which ask about shifts from high to low mood or energy (the first and second parcels).

Data Analysis

To maintain a sufficient participant to variable ratio and improve the reliability of factor indicators, the 73 GBI items were grouped into the same 20 parcels used in prior research (Danielson et al, 2003; Youngstrom et al, 2001; Youngstrom, Murray, Johnson, & Findling, 2013). Each parcel was composed of three or four items with relatively homogenous content. CFAs on items from each parcel confirmed unidimensionality within each parcel. Additionally, differential item functioning (DIF) analyses were conducted using an item response theory (IRT) approach to examine racial differences among items within each parcel relative to other items in the same parcel. Findings from these analyses are available online as supplementary material. Overall, items showed minimal item-level DIF, and the few differences that were statistically significant tended to cancel out at the scale score level, consistent with findings with other data (Freeman et al., 2012).

Structural Invariance—Prior to beginning invariance analyses, two-factor GBI models were tested separately for each racial group. CFAs with robust maximum likelihood (MLM) estimation (Kline, 2005) were conducted using MPlus Version 6.1 (Muthen & Muthen, 2010) based on item parcel scores. Model fit was evaluated via standardized root mean residual (SRMR), the root mean error of approximation (RMSEA), and the comparative fit index (CFI; Kline, 2005; Tanaka, 1993). Criteria for evaluating model fit were established a priori: RMSEA values less than .075, SRMR values less than .08, and CFI values greater than .90 (Hu & Bentler, 1995; Markland, 2006). Although more stringent criteria for evaluating model fit have been proposed (e.g., CFI >.95; Hu & Bentler, 1999), scholars have warned against rigid adherence to these more stringent criteria as it could lead to inappropriate rejection of well-fitting models (e.g., Marsh, Hau, & Wen, 2009).

Multi-group CFAs were used to evaluate structural invariance across race. Invariance of the two-factor structure was assessed by applying increasingly restrictive constraints across groups to examine:(a) configural invariance (all parameters free to vary across groups), (b) weak invariance (also known as metric invariance; factor loadings constrained to be equal across groups), (c) strong invariance (scalar invariance; intercepts of item parcels also constrained to be equal), (d) strict invariance (residual invariance; residuals also constrained to be equal), (e) invariance of factor means, and (f) invariance of factor variances and covariances (e.g., Dimitrov, 2010; Meredith, 1993). The criteria for weak (metric) invariance are met if factor loadings are invariant across groups (i.e., the slopes of the regression lines between items and factors are equivalent across groups). In other words, weak invariance means that for all items, a one unit change in an item score results in an equal change in factor score across groups. For example, if endorsing an item (e.g., “cries often) was strongly associated with scores on a depression factor for males but had only a small relationship with depression factor scores for females, then the scale would *not* meet criteria for weak invariance across gender. The criteria for strong (scalar) invariance are met if the intercepts as well as factor loadings (intercepts and slopes of regression lines between items and factors) are equivalent across groups. Finally, the criteria for strict invariance are met if the slopes, intercepts, and residuals are equivalent across groups. If criteria for strong invariance are met, examination of invariance of factor means, variance, and covariance is warranted (Wu, Li, & Zumbo, 2007). According to Meredith and Teresi (2006), weak

invariance across groups is a requirement in order to compare scores or latent variable means. Strong and strict invariance are less important for research purposes but are desirable and allow for greater fairness and equity in assessments.

Change in Satorra-Bentler chi-square (χ^2) values and change in CFI (CFI) values were used to compare nested models. As the models grew more restrictive, non-significant χ^2 ($p > .05$) and CFI ($< .01$; Cheung & Rensvold, 2002) were considered to be evidence that the more restrictive model fit the data as well as the less restrictive one (Byrne, 2011; Meade, Johnson, & Braddy, 2008; Satorra & Bentler, 2001). The CFI value was weighted more heavily than the χ^2 in evaluating change in model fit because CFI values are less sensitive to sample size (Cheung & Rensvold, 2002).

Results

Preliminary Analyses

Cases were deleted list-wise if participants failed to respond to more than 30% of GBI items ($n=13$). Missing data were imputed for those responding to 70% of the items using the multiple imputation algorithms (based on Bayesian estimation) in MPlus version 6.1 ($n=70$; Roth & Switzer, 1999). Data from Black participants were missing at a higher rate ($n=6$; 2.1% exclusion rate) than data from White participants ($n=7$; 0.7% exclusion rate), but the difference was not statistically significant. Notably, most participants with missing data failed to respond to only one or two of the 73 GBI items and no systematic pattern of missing data could be identified. The final sample consisted of 285 Black and 987 White undergraduate students. Item-level descriptive statistics and correlation matrices are available by request through the first author, and descriptive statistics for subscale scores are provided in Table 2. As expected given the nature of the construct, scatterplots and univariate skewness and kurtosis statistics revealed mild departures from normality for some item parcels. As such, robust estimation methods were used in SEM analyses (Kline, 2005). Item parcel scores were used for all CFA and invariance analyses.

Invariance Analyses

As recommended by Meade, Johnson, and Braddy (2008), model fit was examined separately for each racial group before testing for invariance. The tested model was guided by the theoretical model and scale development (Depue et al., 1989) as well as prior investigations of structure using the parcels (Danielson et al., 2003; Youngstrom et al., 2001; 2013). One post-hoc re-specification was made for both groups based on theory and inspection of modification indices: the residuals from parcels 9 (feels sad) and 18 (dysthymic rumination) were allowed to co-vary. Items in both parcels 9 (feels sad) and 18 (dysthymic rumination) were designed to tap emotional symptoms of depression. For example, item 3 in parcel 9 (feels sad) is “Have you become sad, depressed, or irritable for several days or more without really understanding why?” and item 34 in parcel 18 (dysthymic rumination) is “Have there been long periods in your life when you felt sad, depressed, or irritable most of the time?” Given the similarity in wording and content between items in parcel 9 (feels sad) and parcel 18 (dysthymic rumination), it is reasonable to expect that the residuals may co-vary. Overall (in keeping with theory and prior research),

eight parcels loaded on the (hypo)manic/ biphasic subscale, and fourteen parcels loaded on the depression subscale. Parcels 1 (extremes of mood and energy) and 2 (mood never in the middle) loaded on both factors.

Table 3 reports findings from all CFA analyses of the two-factor GBI model. Among White participants, all fit indices of the two-factor model (Model 1) were within the limits specified a priori. For Black participants, most fit indices of the model (Model 2) met a priori criteria for adequate fit; however, the RMSEA value (.085) was somewhat high. Nonetheless, because empirical recommendations state that multiple, as opposed to single, fit indices should be considered when evaluating fit (e.g., Tanaka, 1993), the two-factor GBI model was considered to have adequate fit for members of both groups. For illustrative purposes, a diagram of the tested GBI model is provided in Figure 1. For both Black and White participants, all factor loadings were statistically significant, except the path between parcel 2 (mood never in the middle) and the Hypomanic/biphasic factor was non-significant for both groups.

In the next step, configural invariance across groups was established. Fit indices for the configural model (Model 3) fell within the specified ranges. Thus, the model was determined to have adequate fit across groups. The inter-factor correlation was fairly high: .85 among White participants and .90 among Black participants. All factor loadings (aside from parcel 2, mood never in the middle, on the Hypomanic/biphasic factor) were statistically significant ($p < .01$) and very similar across racial groups, ranging from .30 to .90. Factor loadings and inter-factor correlations remained relatively constant in the subsequent, more restrictive invariance analyses. Because parcel 2 (mood never in the middle) did not significantly load on factor 2 (Depression), the path was dropped. The analyses were re-run and there was no change to overall model fit. As such, the path between parcel 2 and factor 2 was omitted in subsequent analyses.

Model 4 also constrained factor loadings to be equal across racial groups to evaluate weak (metric) invariance. The model had an adequate fit to the data based on a priori criteria. Additionally, change-in-model fit statistics (i.e., χ^2 and CFI) suggested that the fit of the weak (metric) invariance model (model 4) did not significantly differ from that of the configural model, supporting weak (metric) invariance across racial groups.

Model 5 tested a scalar (strong) invariance model by constraining the factor loadings and intercepts of indicator variables across racial groups. The findings supported scalar invariance of GBI scores across racial groups.

Next, Model 6 constrained the residual (strict) variances and covariances to be equal across racial groups. Based on all fit indices and change-in-fit indices, the strict (residual) invariance model fit poorly (e.g., CFI=.85; RMSEA=.10) and had significantly worse fit than the scalar invariance model. Thus, the findings did not support strict (residual) invariance of GBI scores across racial groups.

Many scholars suggest that weak (metric) and strong (scalar) invariance are necessary and sufficient conditions for examining latent invariance (e.g., Dimitrov, 2010). Given that weak (metric) and strong (scalar) invariance were established with this sample, analyses

examining invariance of factor means and factor variances were conducted, and the factorial invariance models were compared with the strong (scalar) invariance model. Both factorial invariance models fit the data well, as evidenced by fit indices, and did not significantly differ from the strong (scalar) invariance model based on the change-in-model fit indices. Therefore, factorial invariance of GBI scores across racial groups was supported. In summary, GBI scores have similar relations with latent factors (weak invariance), intercepts (strong invariance), and means, variances, and co-variances of latent factors across racial groups. However, because strict (residual) invariance was not supported, it cannot be said that all group differences in scores are due *only* to group differences on the latent factors (Chen, 2007).

Discussion

The goal of this study was to examine whether there is evidence of structural invariance when comparing the self-report performance on the GBI in samples of Black and White participants drawn from a non-clinical setting. The GBI is an excellent choice for investigating invariance between these groups because (a) it measures depressive and mixed/biphasic symptoms, as well as symptoms of hypomania and mania, providing coverage of the range of emotional dysregulation associated with bipolar disorder; (b) it assesses sub-syndromal symptoms and associated features in addition to DSM symptoms, extending the content into areas where there might be additional cultural differences; (c) the GBI has been exceptionally well-validated for a rating scale in terms of demonstrating criterion validity with diagnoses of mood disorder (Danielson et al., 2003; Depue et al., 1981; Klein, Depue, & Slater, 1986; Mallon, Klein, Bornstein, & Slater, 1986), physiological correlates of mood disorder (Depue, Kleiman, Davis, Hutchinson, & Krauss, 1985; Depue, Luciana, Arbisi, Collins, & Leon, 1994), family history of mood disorder (Depue et al., 1981; Findling et al., 2005; Klein et al., 1986), prospective longitudinal prediction of mood disorder (Alloy et al., 2012; Klein & Depue, 1984) and response to treatment for mood disorder (Findling et al., 2003; Youngstrom, Zhao, et al., 2013).

The GBI's inclusion of mixed mood items – describing shifts in extreme energy as well as mood – anticipated the emphasis on energy as a characteristic of bipolar disorders in the *DSM-5* revision (APA, 2013). Similarly, the biphasic content and mixed items align with careful descriptions of bipolar phenomenology (Ghaemi et al., 2008) and with the *DSM-5* specifier “with mixed features” which now can be coded on hypomanic and depressive as well as manic episodes. The topic of measurement invariance versus bias in the assessment of bipolar disorder across racial groups is crucial in light of the large disparities between Black and White persons in the United States in terms of access to mental health services (Alegria, Chatterji, et al., 2008), discrepant rates of clinical diagnosis of bipolar disorder versus schizophrenia and antisocial personality (Chrishon et al., 2012; Strakowski, Keck, et al., 2003), and differences in medication prescription and service utilization (Barnes, 2004; DelBello et al., 2001; Strakowski, Keck, et al.).

Findings indicate that the GBI showed invariant factor structure between the Black and White groups in a non-clinical setting, measuring two highly correlated factors of depressive and hypomanic/mixed symptoms. Progressive models of invariance revealed acceptable fit

for tests of configural invariance, equal factor loadings, and equal intercepts (i.e., “strong/ scalar invariance”), but not invariance of the residuals and co-variances. Thus, the GBI did not demonstrate complete measurement invariance, but appears to be measuring the same factors of mood symptoms with configural and strong (scalar) invariance. The high correlation between factors replicates prior findings with the GBI (Depue et al., 1981) and is consistent with both the phenomenology of mood disorder, where mixed presentations are fairly common (Phelps, Angst, Katzow, & Sadler, 2008), as well as the complexity of the GBI items, which often combine different moods and energy levels in the same item. The complex item content reflects longstanding clinical observations of the mood spectrum (Kraepelin, 1921) as well as current data about the spectrum of presentations between “pure” depression and “pure” mania that are evident both in cross-sectional data as well as longitudinal follow-up studies (e.g., Angst et al., 2011; Fiedorowicz et al., 2011).

Overall, findings indicate that the GBI provides functionally invariant measurement of mood symptoms in both the Black and White groups drawn from a non-clinical setting. The invariance finding is important, not just as an addition to the growing nomothetic network of validation for the GBI, but also because of the larger literature on different rates of diagnoses and service utilization between Blacks and Whites in the USA (Alegria, Chatterji, et al., 2008; Neighbors et al., 2007). In this broader context, findings indicate that the structure of mood dimensions and the relationship of most symptoms to the mood factors are similar across the two groups. When asked the same questions about mood symptoms, both Black and White participants gave similar answers that showed mostly invariant relationships to their underlying level of mood problems. Notably, this study did not directly examine whether the GBI can reduce differential rates of diagnosis between Black and White individuals, and studies that directly address that question (e.g., using ROC/AUC analyses) will be an important area for future inquiry. However, the use of functionally invariant instruments such as the GBI has the potential to contribute consistent information to the clinical assessment process, hopefully reducing discrepancies in diagnosis and attendant differences in choice of medication and services (Alegria, Nakash, et al., 2008).

The present investigation relies on a large sample drawn from a university setting. The GBI has been studied extensively with both university and clinical or epidemiological samples, and consistently has demonstrated criterion validity in all these samples, but it would be valuable to test measurement invariance in a clinical sample. The combination of studies in clinical and non-clinical samples will help to understand the extent to which differences in treatment seeking for similar levels of mood problems may contribute to the larger patterns of discrepant diagnosis and intervention. Moreover, because a post-hoc modification was used in this study and the fit statistics were slightly below the generally used standards, replication is warranted to ensure generalizability of findings. It also would be valuable to study invariance in other countries and cultural groups, and at other age ranges.

Finally, although GBI scores were largely invariant across racial group, the overall fit for Black participants was marginal. As such, further research using a mix of qualitative and quantitative approaches to investigate how Black Americans normally think about mood and behavior problems, as well as attitudes towards treatment (Carpenter-Song, Whitley, Lawson, Quimby, & Drake, 2011) will be crucial to fully understand and identify possible

racial differences in symptom expression and reporting. There is evidence that many cultural groups focus more on somatic or behavioral aspects of mood disorders rather than conceptualizing them as problems of emotion (Angst et al., 2010). A medical anthropologist observing clinical interviews of families seeking mental health services for youths found that Black families were more likely to concentrate on externalizing behaviors in their descriptions of the presenting problem, whereas White families were more likely to describe “mood swings” for youths that all met criteria for bipolar spectrum disorders after completing comprehensive, semi-structured research interviews (Carpenter-Song, 2009). Clinicians doing unstructured interviews tend to focus on the initial description of the presenting problem and seek more confirmatory than disconfirmatory evidence (Croskerry, 2003; Garb, 1998). Thus, it may be that cultural differences in the initial framing of the presenting problem combine with cognitive heuristics to lead clinicians to assign different diagnoses to cases that would warrant identical diagnoses if the clinician engaged in more systematic assessment. The present findings suggest that the GBI may have a role as a measure of mood problems that can provide functionally invariant results to increase the chances of accurate diagnosis, case formulation, and treatment planning for Black and White individuals alike.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This research was supported by NIMH grants MH52617 and MH77908 to Lauren B. Alloy.

References

- Alegria M, Chatterji P, Wells K, Cao Z, Chen CN, Takeuchi D, Meng XL. Disparity in depression treatment among racial and ethnic minority populations in the United States. *Psychiatric Services*. 2008; 59:1264–1272.10.1176/appi.ps.59.11.1264 [PubMed: 18971402]
- Alegria M, Nakash O, Lapatin S, Oddo V, Gao S, Lin J, Normand SL. How missing information in diagnosis can lead to disparities in the clinical encounter. *Journal of Public Health Management and Practice*. 2008; 14(Suppl):S26–35.10.1097/01.PHH.0000338384.82436.0d [PubMed: 18843234]
- Alloy LB, Abramson LY, Walshaw PD, Cogswell A, Sylvia LG, Hughes ME, Hogan ME. Behavioral Approach System (BAS) and Behavioral Inhibition System (BIS) sensitivities and bipolar spectrum disorders: Prospective prediction of bipolar mood episodes. *Bipolar Disorders*. 2008; 10:310–322. [PubMed: 18271911]
- Alloy LB, Urosevic S, Abramson LY, Jager-Hyman S, Nusslock R, Whitehouse WG, Hogan ME. Progression along the bipolar spectrum: A longitudinal study of predictors of conversion from bipolar spectrum conditions to bipolar I and II disorders. *Journal of Abnormal Psychology*. 2012; 121:16–27.10.1037/a0023973 [PubMed: 21668080]
- American Psychiatric Association. *Diagnostic and statistical manual of mental health disorders: DSM-5*. 5. Washington, DC: American Psychiatric Publishing; 2013.
- American Psychiatric Association. *Diagnostic and statistical manual of mental disorders*. 3. Washington, D.C: Author; 1980.
- Angst J. The emerging epidemiology of hypomania and bipolar II disorder. *Journal of Affective Disorders*. 1998; 50:143–151. [PubMed: 9858074]
- Angst J, Azorin JM, Bowden CL, Perugi G, Vieta E, Gamma A, Young AH. Prevalence and characteristics of undiagnosed bipolar disorders in patients with a major depressive episode: the

- BRIDGE study. *Archives of General Psychiatry*. 2011; 68:791–798.10.1001/archgenpsychiatry.2011.87 [PubMed: 21810644]
- Angst J, Meyer TD, Adolfsson R, Skeppar P, Carta M, Benazzi F, Gamma A. Hypomania: A transcultural perspective. *World Psychiatry*. 2010; 9:41–49. [PubMed: 20148160]
- Barnes A. Race, schizophrenia, and admission to state psychiatric hospitals. *Administration and Policy in Mental Health*. 2004; 31:241–252. [PubMed: 15160786]
- Bebko G, Bertocci MA, Fournier JC, Hinze AK, Bonar L, Almeida JR, Phillips ML. Parsing Dimensional vs Diagnostic Category-Related Patterns of Reward Circuitry Function in Behaviorally and Emotionally Dysregulated Youth in the Longitudinal Assessment of Manic Symptoms Study. *JAMA Psychiatry*. 2014; 71(1):71–80.10.1001/jamapsychiatry.2013.2870 [PubMed: 24285346]
- Byrne, BM. *Structural equation modeling with Mplus: Basic concepts, applications, and programming*. Mahwah, NJ: Lawrence Erlbaum; 2011.
- Cai, L.; du Toit, SHC.; Thissen, D. *IRTPRO: Flexible, multidimensional, multiple categorical IRT modeling*. Chicago, IL: Scientific Software International; 2011.
- Carpenter-Song E. Caught in the psychiatric net: Meanings and experiences of ADHD, pediatric bipolar disorder and mental health treatment among a diverse group of families in the United States. *Culture, Medicine, and Psychiatry*. 2009; 33:61–85.10.1007/s11013-008-9120-4
- Carpenter-Song E, Whitley R, Lawson W, Quimby E, Drake RE. Reducing disparities in mental health care: Suggestions from the Dartmouth-Howard collaboration. *Community Mental Health Journal*. 2011; 47:1–13.10.1007/s10597-009-9233-4 [PubMed: 19662528]
- Chen FF. Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling*. 2007; 14:464–404.
- Chen R, Alan Y, Swann C, Johnson BA. Stability of diagnosis in bipolar disorder. *Journal of Nervous and Mental Disease*. 1998; 186:17–23. [PubMed: 9457143]
- Cheung GW, Rensvold RB. Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal*. 2002; 9:233–255.10.1207/S15328007SEM0902_5
- Chrishon K, Anderson D, Arora G, Bailey TK. Race and psychiatric diagnostic patterns: understanding the influence of hospital characteristics in the National Hospital Discharge Survey. *Journal of the National Medical Association*. 2011; 104(11–12):505–509. [PubMed: 23560352]
- Cohen AS, Kim SO, Wollack JA. An investigation o the likelihood ratio test for detection of differential item functioning. *Applied Psychological Measurement*. 1996; 20:15–26.10.1177/014662169602000102
- Cooke RG, Robb JC, Young LT, Joffe RT. Well-being and functioning in patients with bipolar disorder assessed using the MOS 20-ITEM short form (SF-20). *Journal of Affective Disorders*. 1996; 2:93–97. [PubMed: 8827417]
- Croskerry P. The importance of cognitive errors in diagnosis and strategies to minimize them. *Academic Medicine*. 2003; 78:775–780.10.1097/00001888-200308000-00003 [PubMed: 12915363]
- Danielson CK, Youngstrom EA, Findling RL, Calabrese JR. Discriminative validity of the General Behavior Inventory using youth report. *Journal of Abnormal Child Psychology*. 2003; 31:29–39. [PubMed: 12597697]
- Dawes RM, Faust D, Meehl PE. Clinical versus actuarial judgment. *Science*. 1989; 243:1668–1674. [PubMed: 2648573]
- Danielson CK, Youngstrom EA, Findling RL, Calabrese JR. Discriminative validity of the General Behavior Inventory using youth report. *Journal of Abnormal Child Psychology*. 2003; 31:29–39.10.1023/A:1021717231272 [PubMed: 12597697]
- DelBello MP, Lopez-Larson MP, Soutullo CA, Strakowski SM. Effects of race on psychiatric diagnosis of hospitalized adolescents: A retrospective chart review. *Journal of Child and Adolescent Psychopharmacology*. 2001; 11:95–103. [PubMed: 11322750]
- Depue RA, Kleiman RM, Davis P, Hutchinson M, Krauss SP. The behavioral high-risk paradigm and bipolar affective disorder, VIII: Serum free cortisol in nonpatient cyclothymic subjects selected by

- the General Behavior Inventory. *American Journal of Psychiatry*. 1985; 142:175–181. [PubMed: 3970242]
- Depue RA, Krauss S, Spoont MR, Arbisi P. General behavior inventory identification of unipolar and bipolar affective conditions in a nonclinical university population. *Journal of Abnormal Psychology*. 1989; 98:117–126. [PubMed: 2708652]
- Depue RA, Luciana M, Arbisi P, Collins P, Leon A. Dopamine and the structure of personality: Relation of agonist-induced dopamine activity to positive emotionality. *Journal of Personality and Social Psychology*. 1994; 67:485–498. [PubMed: 7965602]
- Depue RA, Slater JF, Wolfstetter-Kausch H, Klein DN, Goplerud E, Farr DA. A behavioral paradigm for identifying persons at risk for bipolar depressive disorder: A conceptual framework and five validation studies. *Journal of Abnormal Psychology*. 1981; 90:381–437. [PubMed: 7298991]
- Dimitrov DM. Testing for factorial invariance in the context of construct validation. *Measurement and Evaluation in Counseling and Development*. 2010; 43:121–149.10.1177/0748175610373459
- Embretson, SE.; Reise, SP. *Item response theory for psychologists*. Mahwah, NJ: Erlbaum; 2000.
- Fabrigar LR, Wegener DT, MacCallum RC, Strahan EJ. Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods*. 1999; 4:272–299.
- Fiedorowicz JG, Endicott J, Leon AC, Solomon DA, Keller MB, Coryell WH. Subthreshold hypomanic symptoms in progression from unipolar major depression to bipolar disorder. *American Journal of Psychiatry*. 2011; 168:40–48.10.1176/appi.ajp.2010.10030328 [PubMed: 21078709]
- Findling RL, McNamara NK, Gracious BL, Youngstrom EA, Stansbrey RJ, Reed MD, Calabrese JR. Combination lithium and divalproex in pediatric bipolarity. *Journal of the American Academy of Child & Adolescent Psychiatry*. 2003; 42:895–901. [PubMed: 12874490]
- Findling RL, Youngstrom EA, McNamara NK, Stansbrey RJ, Demeter CA, Bedoya D, Calabrese JR. Early symptoms of mania and the role of parental risk. *Bipolar Disorders*. 2005; 7:623–634. [PubMed: 16403188]
- Fountoulakis KN, Vieta E. Treatment of bipolar disorder: A systematic review of available data and clinical perspectives. *International Journal of Neuropsychopharmacology*. 2008; 11:999–1029.10.1017/S1461145708009231 [PubMed: 18752718]
- Freeman AJ, Youngstrom EA, Frazier TW, Youngstrom JK, Demeter C, Findling RL. Portability of a screener for pediatric bipolar disorder to a diverse setting. *Psychological assessment*. 2012; 24(2): 341. [PubMed: 21942229]
- Garb, HN. *Studying the clinician: Judgment research and psychological assessment*. Washington, DC: American Psychological Association; 1998.
- Geller B, Cooper TB, Sun K, Zimmerman B, Frazier J, Williams M, Heath J. Double-blind and placebo-controlled study of lithium for adolescent bipolar disorders with secondary substance dependency. *Journal of the American Academy of Child and Adolescent Psychiatry*. 1998; 37:171–178. [PubMed: 9473913]
- Ghaemi SN, Bauer M, Cassidy F, Malhi GS, Mitchell P, Phelps J, Youngstrom E. Diagnostic guidelines for bipolar disorder: a summary of the international society for bipolar disorders diagnostic guidelines task force report. *Bipolar Disorders*. 2008; 10(1p2):117–128. [PubMed: 18199230]
- Ghaemi SN, Miller CJ, Berv DA, Klugman J, Rosenquist KJ, Pies RW. Sensitivity and specificity of a new bipolar spectrum diagnostic scale. *Journal of Affective Disorders*. 2005; 84:273–277.10.1016/s0165-0327(03)00196-4 [PubMed: 15708426]
- Harmon-Jones E, Abramson LY, Nusslock R, Sigelman JD, Urosevic S, Turonie L, Fearn M. Effect of bipolar disorder on left frontal cortical responses to goals differing in valence and task difficulty. *Biological Psychiatry*. 2008; 63:693–698. [PubMed: 17919457]
- Hu, LT.; Bentler, PM. Evaluating model fit. In: Hoyle, RH., editor. *Structural equation modeling: Concepts, issues, and applications*. Thousand Oaks, CA: Sage Publications; 1995. p. 76-99.
- Jenkins MM, Youngstrom EA, Washburn JJ, Youngstrom JK. Evidence-based strategies improve assessment of pediatric bipolar disorder by community practitioners. *Professional Psychology: Research and Practice*. 2011; 42(2):121.

- Johnson, SL.; Miller, CJ.; Eisner, L. Bipolar Disorder. In: Hunsley, J.; Mash, EJ., editors. *A Guide to Assessments That Work*. New York: Oxford University Press; 2008. p. 121-137.
- Kilbourne AM, Haas GL, Mulsant BH, Bauer MS, Pincus HA. Concurrent psychiatric diagnoses by age and race among persons with bipolar disorder. *Psychiatric Services*. 2004; 55:931-933. [PubMed: 15292544]
- Klein DN, Depue RA. Continued impairment in persons at risk for bipolar affective disorder: results of a 19-month follow-up study. *Journal of Abnormal Psychology*. 1984; 93:345-347. [PubMed: 6470321]
- Klein DN, Depue RA, Slater JF. Inventory identification of cyclothymia. IX. Validation in offspring of bipolar I patients. *Archives of General Psychiatry*. 1986; 43:441-445. [PubMed: 3964023]
- Klein DN, Dickstein S, Taylor EB, Harding K. Identifying chronic affective disorders in outpatients: Validation of the general behavior inventory. *Journal of Consulting and Clinical Psychology*. 1989; 57:106-111. [PubMed: 2925959]
- Kline, RB. *Principles and practice of structural equation modeling*. New York, NY: Guilford Press; 2005.
- Kraepelin, E. *Manic-depressive insanity and paranoia*. Edinburgh: Livingstone; 1921.
- Lish JD, Dime-Meenan S, Whybrow PC, Price RA, Hirschfeld R. The National Depressive and Manic-depressive Association (DMDA) survey of bipolar members. *Journal of Affective Disorders*. 1994; 31:281-294.10.1016/0165-0327(94)90104-X [PubMed: 7989643]
- Lopez AD, Mathers CD, Ezzati M, Jamison DT, Murray CJ. Global and regional burden of disease and risk factors, 2001: Systematic analysis of population health data. *Lancet*. 2006; 367:1747-1757. [PubMed: 16731270]
- Mallon JC, Klein DN, Bornstein RF, Slater JF. Discriminant validity of the General Behavior Inventory: An outpatient study. *Journal of Personality Assessment*. 1986; 50:568-577. [PubMed: 3820050]
- Mantere O, Suiminen K, Leppamaki S, Arvilommi P, Isometsa E. The clinical characteristics of DSM-IV bipolar I and II disorders: Baseline findings from the Jorvi Bipolar Study (JoBS). *Bipolar Disorders*. 2004; 6:395-405. [PubMed: 15383132]
- Markland D. The golden rule is that there are no golden rules: A commentary on Paul Barrett's recommendations for reporting model fit in structural equation modeling. *Personality and Individual Differences*. 2006; 42:851-858.10.1016/j.paid.2006.09.023
- Marsh HW, Hau KT, Wen Z. In search of golden rules: Comment on hypothesis-testing approaches to setting cutoff values for fit indexes and dangers in overgeneralizing Hu and Bentler's (1999) findings. *Structural Equation Modeling: A Multidisciplinary Journal*. 2004; 11:320-341.10.1207/s15328007sem1103_2
- Meade AW, Johnson EC, Braddy PW. Power and sensitivity of alternative fit indices in tests of measurement invariance. *Journal of Applied Psychology*. 2008; 93:568-592.10.1037/0021-9010.93.3.568 [PubMed: 18457487]
- Meredith W. Measurement invariance, factor analysis, and factorial invariance. *Psychometrika*. 1993; 58:523-543.10.1007/BF02294825
- Meredith W, Teresi JA. An essay on measurement and factorial invariance. *Medical care*. 2006; 44(11):S69-S77. [PubMed: 17060838]
- Merikangas KR, Akiskal HS, Angst J, Greenberg PE, Hirschfeld RM, Petukhova M, Kessler RC. Lifetime and 12-Month Prevalence of Bipolar Spectrum Disorder in the National Comorbidity Survey Replication. *Archives of General Psychiatry*. 2007; 64:543-552. [PubMed: 17485606]
- Miller CJ, Johnson SL, Eisner L. Assessment tools for adult bipolar disorder. *Clinical Psychology: Science and Practice*. 2009; 16:188-201. [PubMed: 20360999]
- Miller CJ, Klugman J, Berv DA, Rosenquist KJ, Ghaemi SN. Sensitivity and specificity of the Mood Disorder Questionnaire for detecting bipolar disorder. *Journal of Affective Disorders*. 2004; 81:167-171. [PubMed: 15306144]
- Moreno C, Hasin DS, Arango C, Oquendo MA, Vieta E, Liu S, Blanco C. Depression in bipolar disorder versus major depressive disorder: results from the National Epidemiologic Survey on Alcohol and Related Conditions. *Bipolar Disorders*. 2013; 14(3):271-282. [PubMed: 22548900]
- Muthen, LK.; Muthen, BO. *MPlus 6.1*. Los Angeles, CA: Muthen & Muthen; 2010.

- Neighbors HW, Caldwell C, Williams DR, Nesse R, Taylor RJ, Bullard KM, et al. Race, Ethnicity, and the Use of Services for Mental Disorders: Results From the National Survey of American Life. *Archives of General Psychiatry*. 2007; 64:485–494.10.1001/archpsyc.64.4.485 [PubMed: 17404125]
- Perron BE, Fries LE, Kilbourne AM, Vaughn MG, Bauer MS. Racial/ethnic group differences in bipolar symptomatology in a community sample of persons with bipolar I disorder. *The Journal of Nervous and Mental Disease*. 2010; 198:16–21. [PubMed: 20061864]
- Phelps J, Angst J, Katzow J, Sadler J. Validity and utility of bipolar spectrum models. *Bipolar Disorders*. 2008; 10:179–193. BDI562 [pii]. 10.1111/j.1399-5618.2007.00562.x [PubMed: 18199236]
- Regier DA, Farmer ME, Rae DS, Locke BZ, Keith SJ, Judd LL, Goodwin FK. Comorbidity of mental disorders with alcohol and other drug abuse: Results from the epidemiological catchment area (ECA) study. *JAMA*. 1990; 264:2511–2518. [PubMed: 2232018]
- Robins, LN.; Regier, DA., editors. *Psychiatric Disorders in America*. Free Press; New York: 1991. The Epidemiologic Catchment Area Study
- Rogers, R.; Jackson, RL.; Cashel, M. The schedule for affective disorders and schizophrenia (SADS). In: Rogers, R., editor. *Handbook of diagnostic and structural interviewing*. New York: Guilford Press; 2001. p. 84-102.
- Roth, PL.; Switzer, FS. Missing data: Instrument-level heffalumps and item-level woozles. 1999. Retrieved from: <http://www.aom.pace.edu>
- Samejima, F. Graded response modesl. In: Hambleton, WJ., editor. *Handbook of modern item response theory*. Springer-Verlag; New York: 1997. p. 85-100.
- Satorra A, Bentler PM. A scaled difference chi-square test statistic for moment structure analysis. *Psychometrika*. 2001; 66:507–514.10.1007/BF02296192
- Smedley, BD.; Stith, AY.; Nelson, AR. *Unequal treatment: Confronting racial and ethnic disparities in health care (summary)*. Washington, DC: Institute of Medicine, National Academy Press; 2002.
- Smith DJ, Nicholl BI, Cullen B, Martin D, Ul-Haq Z, Evans J, Pell JP. Prevalence and characteristics of probable major depression and bipolar disorder within UK Biobank: cross-sectional study of 172,751 participants. *PLoS One*. 2013; 8(11):e75362. [PubMed: 24282498]
- Strakowski SM, Keck PE, Arnold LM, Collins J, Wilson RM, Fleck DE, Corey KB, Amicone J, Adebimpe VR. Ethnicity and diagnosis in patients with affective disorders. *Journal of Clinical Psychiatry*. 2003; 64:747–754. [PubMed: 12934973]
- Strakowski SM, Lonczak HS, Sax KW, West SA, Crist A, Mehta R, Thienhaus OJ. The effects of race on diagnosis and disposition from a psychiatric emergency service. *Journal of Clinical Psychiatry*. 1995; 56:101–107. [PubMed: 7883727]
- Tanaka, JS. Multifaceted conceptions of fit in structural equation models. In: Bollen, KA.; Long, JS., editors. *Testing structural equation models*. Newbury Park, CA: Sage; 1993. p. 10-39.
- Woods CM. Empirical selection of anchors for tests of differential item functioning. *Applied Psychological Measurement*. 2009; 33:42–57.
- Wu AD, Li Z, Zumbo BD. Decoding the meaning of factorial invariance and updating the practice of multi-group confirmatory factor analysis: A demonstration with TIMSS data. *Practical Assessment, Research and Evaluation*. 2007; 12(3):1–26.
- Youngstrom EA, Findling RL, Danielson CK, Calabrese JR. Discriminative validity of parent report of hypomanic and depressive symptoms on the General Behavior Inventory. *Psychological Assessment*. 2001; 13:276–286.10.1037/1040-3590.13.2.267
- Youngstrom EA, Murray G, Johnson SL, Findling RL. The 7 Up 7 Down Inventory: A 14-item measure of manic and depressive tendencies carved from the General Behavior Inventory. *Psychological Assessment*. 2013; 25:1377–1383.10.1037/a0033975 [PubMed: 23914960]
- Youngstrom EA, Zhao J, Mankoski R, Forbes RA, Marcus RM, Carson W, Findling RL. Clinical significance of treatment effects with aripiprazole versus placebo in a study of manic or mixed episodes associated with pediatric bipolar. I. disorder. *Journal of Child & Adolescent Psychopharmacology*. 2013; 23:72–79.10.1089/cap.2012.0024 [PubMed: 23480324]

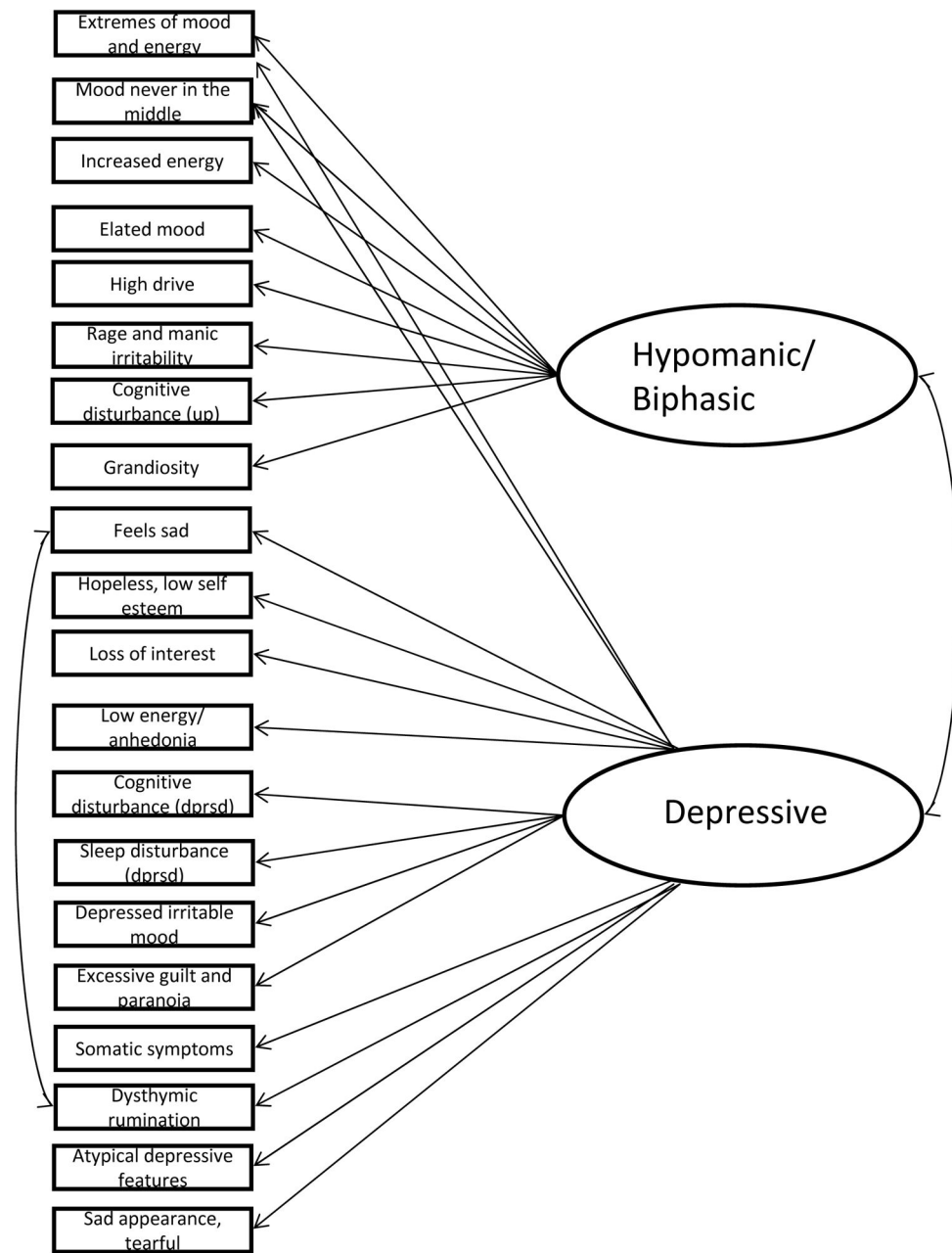


Figure 1. GBI structure tested, in separate analyses, for Black and White participants (models 1 and 2). This structure was also used for subsequent invariance testing.

Table 1

Reliability Estimates for GBI Item Parcels.

Parcel	Items	α
Biphasic (Cross-Loaded Items)		
1. Extremes of Mood and Energy	2, 24, 35, 48	0.72
2. Mood Never in the Middle	19, 40, 53	0.81
Hypomanic (Factor I)		
3. Increased Energy	4, 7, 15	0.67
4. Elevated Mood	22, 30, 31, 66	0.74
5. High Drive	11, 17, 42, 51	0.67
6. Rage and Manic Irritability	27, 44, 54	0.69
7. Cognitive Disturbance (up)	8, 57, 64	0.72
8. Grandiosity	38, 43, 46, 61	0.70
Depression (Factor II)		
9. Feels Sad	3, 23, 45, 63	0.88
10. Hopeless, Low Self-Esteem	47, 56, 62, 73	0.88
11. Loss of Interest	9, 10, 13, 70	0.76
12. Low Energy/ Anhedonia	21, 33, 49, 59	0.82
13. Cognitive Disturbance (down)	1, 12, 41	0.71
14. Sleep Disturbance	5, 25, 37, 52	0.75
15. Depressive Irritable Mood	14, 39, 55	0.79
16. Excessive Guilt and Paranoia	29, 36, 50, 71	0.81
17. Somatic Symptoms	16, 60, 65, 67	0.68
18. Dysthymic Rumination	20, 32, 34, 72	0.89
19. Atypical Depressive Features	18, 26, 58, 68	0.80
20. Sad Appearance, Tearful	6, 28, 69	0.67

Table 2

Descriptive Statistics for GBI Subscales by Race

	Hypomanic/Biphasic		Depressive	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Black participants (<i>n</i> =285)	27.77	26.12	16.33	14.18
White participants (<i>n</i> =985)	28.94	23.91	17.28	12.82
Total	28.68	24.42	17.07	13.04

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 3

Fit Statistics for Invariance of GBI across Racial Groups

Model	n	χ^2	Df	p	χ^2	df	p	CFI	CFI	RMSEA	RMSEA CI	SRMR
1 GBI White Participants	985	1011.58	167	.00	-	-	-	.938	-	.072	.067-.076	.041
2 GBI Black Participants	285	508.83	167	.00	-	-	-	.934	-	.085	.076-.093	.036
3 Configural Invariance	1270	1534.92	334	.00	-	-	-	.937	-	.075	.071-.079	.040
4 Weak Invariance	1270	1578.21	355	.00	35.98	21	.43	.936	-.001	.074	.070-.077	.068
5 Strong Invariance	1270	1674.99	373	.00	104.10	18	.09	.932	-.004	.074	.071-.078	.069
6 Strict Invariance	1270	2954.33	430	.00	1384.01	57	.00	.867	-.065	.096	.093-.099	.250
7* Invariant Factor Means	1270	1676.40	374	.00	1.41	1	.97	.932	0	.074	.070-.078	.069
8* Invariant Factor Variances	1270	1670.23	371	.00	4.76	2	.94	.932	0	.074	.071-.078	.045

Note.

* Change in chi-square was calculated using the Satorra-Bentler Chi-Square difference test formula. Models 1 and 2 refer to the CFAs with White and Black participants, respectively.

* Models 7 and 8 were evaluated relative to the strong (scalar) invariance model.