






## Assessing the Irritable and Defiant Dimensions of Youth Oppositional Behavior Using CBCL and YSR Items

Spencer C. Evans, F. Tony Bonadio, Sarah Kate Bearman, Ana M. Ugueto, Bruce F. Chorpita & John R. Weisz


To cite this article: Spencer C. Evans, F. Tony Bonadio, Sarah Kate Bearman, Ana M. Ugueto, Bruce F. Chorpita & John R. Weisz (2019): Assessing the Irritable and Defiant Dimensions of Youth Oppositional Behavior Using CBCL and YSR Items, Journal of Clinical Child & Adolescent Psychology, DOI: [10.1080/15374416.2019.1622119](https://doi.org/10.1080/15374416.2019.1622119)

To link to this article: <https://doi.org/10.1080/15374416.2019.1622119>

 View supplementary material 

 Published online: 05 Jul 2019.

 Submit your article to this journal 


 Article views: 178

 View related articles 

 View Crossmark data 



# Assessing the Irritable and Defiant Dimensions of Youth Oppositional Behavior Using CBCL and YSR Items

Spencer C. Evans 

*Department of Psychology, Harvard University*

F. Tony Bonadio

*Department of Psychology, Harvard University and Institute for Innovation and Implementation,  
School of Social Work, University of Maryland, Baltimore*

Sarah Kate Bearman

*Department of Educational Psychology, University of Texas at Austin*

Ana M. Ugueto

*Department of Psychiatry, McGovern Medical School, UT Health Science Center at Houston*

Bruce F. Chorpita

*Department of Psychology, University of California Los Angeles*

John R. Weisz

*Department of Psychology, Harvard University*

Research suggests that irritability and defiance are distinct dimensions of youth oppositionality that are differentially associated with internalizing and conduct problems, respectively. Because much of this evidence has emerged with limited psychometric evaluation, we conducted the first multi-informant examination of selected Child Behavior Checklist (CBCL) and Youth Self Report (YSR) items for measuring irritability and defiance in a large clinical sample. Clinically referred youths ( $N = 1,030$ ; ages 6–15; 43% female, 42% ethnic minority) were assessed prior to treatment using multi-informant rating scales and diagnostic interviews. Analyses examined factor structure, invariance, internal consistency, multi-informant patterns, and convergent, discriminant, and criterion validity with internalizing and externalizing problems/disorders. A correlated 2-factor model of irritability (stubborn/sullen/irritable, mood, temper) and defiance (argues, disobeys-home, disobeys-school) fit well for both informants. Adequate measurement invariance and scale consistency was consistently found for parent-report but not youth-report. With both informants, all hypothesized convergent and discriminant validity associations were supported: irritability and defiance with internalizing and conduct

---

Address correspondence to Spencer C. Evans, Department of Psychology, Harvard University, 1040 William James Hall, 33 Kirkland St, Cambridge, MA 02138. E-mail: [scevans@fas.harvard.edu](mailto:scevans@fas.harvard.edu)

Color versions of one or more of the figures in the article can be found online at [www.tandfonline.com/hcap](http://www.tandfonline.com/hcap).

This article has been republished with minor changes. These changes do not impact the academic content of the article.

scales, respectively. However, hypothesized criterion validity associations were largely found only by parent-report: irritability with anxiety and depressive disorders, defiance with conduct disorder, and both with oppositional defiant disorder. Results consistently supported the reliability and validity of the CBCL irritability and defiance scales, with somewhat less consistent support for the YSR scales. Thus, CBCL items may provide psychometrically sound assessment of irritability and defiance, whereas further research is needed to advance youth-report and multi-informant strategies. Results also provide further support for a two subdimension model of oppositional defiant disorder symptoms that includes irritability and defiance.

## INTRODUCTION

The heterogeneity of youth oppositional behavior has long been acknowledged but is not well understood. Recently, more than three dozen studies have helped disentangle this heterogeneity and suggested that the structure of oppositional defiant disorder (ODD) symptoms may be multidimensional in nature (see Evans et al., 2017, for a review). Across samples, methods, and analytic approaches, two dimensions of oppositionality are most consistently identified: *irritability* (e.g., touchy, angry, easily annoyed) and *defiance* (e.g., arguing, disobeying, refusing to comply; American Psychiatric Association [APA], 2013; Burke et al., 2014; Evans et al., 2017).<sup>1</sup> Cross-sectionally and longitudinally, irritability is uniquely linked to anxiety and depression, whereas defiance is more strongly linked to severe conduct problems (e.g., Althoff, Kuny-Slock, Verhulst, Hudziak, & van der Ende, 2014; Burke, 2012; Burke, Hipwell, & Loeber, 2010; Evans, Pederson, Fite, Blossom, & Cooley, 2016; Ezpeleta, Granero, de la Osa, Penelo, & Domènech, 2012; Rowe, Costello, Angold, Copeland, & Maughan, 2010; Stringaris & Goodman, 2009a, 2009b). These findings demonstrate the clinical and developmental significance of irritability and defiance as dimensions of oppositional behavior.

However, the dimensionality of oppositional behavior remains a matter of debate and open inquiry, and there are

some important methodological considerations warranting closer attention. First, most of this evidence comes from secondary analyses of items drawn from other measures to approximate irritability, defiance, or related constructs. As a result, the exact content and structure of these models varies depending on researchers' choice of methods, measures, and items. In many cases, ad hoc scales have been used to explore interesting hypotheses about irritability or dimensions of oppositionality but with limited attention to measurement reliability and validity. Thus, it remains important to examine the psychometric properties of these scales. Second, although many studies have investigated irritability and defiance together, others have focused predominately on severe irritability (see Brotman, Kircanski, & Leibenluft, 2017; Evans et al., 2017; Stringaris, Vidal-Ribas, Brotman, & Leibenluft, 2018, for recent reviews), including constructs such as chronic irritability (e.g., Stringaris, Cohen, Pine, & Leibenluft, 2009), severe mood dysregulation (Leibenluft, Charney, Towbin, Bhangoo, & Pine, 2003), and disruptive mood dysregulation disorder (APA, 2013; Roy, Lopes, & Klein, 2014). The measures, methods, dimensions, and categories used in these irritability, severe mood dysregulation, and disruptive mood dysregulation disorder studies are highly overlapping with those used in irritability-defiance studies, yielding similar results (Evans et al., 2017; Vidal-Ribas, Brotman, Valdivieso, Leibenluft, & Stringaris, 2016). Thus, this evidence should be considered collectively, and research would benefit from scales that can measure irritability and defiance both as *correlated constructs* (i.e., as dimensions of oppositionality) and as *independent constructs* (e.g., irritability alone, if defiance is not of interest). This study seeks to address these gaps by examining brief parent-report and self-report scales for measuring irritability and defiance among treatment-referred youths.

## Approaches to Measuring Youth Irritability and Defiance

One issue for any study of youth irritability and defiance is that only a few relevant methods and measures exist, each with strengths and limitations. The bulk of evidence and the most rigorous measurement studies have been conducted with items drawn from the eight diagnostic criteria for ODD defined in the *Diagnostic and Statistical Manual of Mental Disorders* (5th ed.; *DSM-5*; APA, 2013). Using this

<sup>1</sup> With minor variations in content and terminology, nearly all studies identify the dimensions of *irritability* and *defiance*; some identify a third area related to *spiteful/vindictive* behavior (Evans et al., 2017). However, this third factor is typically only identified using alternative measures and item sets (e.g., DAWBA). ODD criteria include only one indicator of spiteful/vindictive behavior, and one indicator is not sufficient for identifying a latent construct. The bulk of the evidence supports two ODD dimensions, irritability and defiance. For example, Burke et al.'s (2014) large multi-sample factor analysis found evidence for a two-factor structure, which was recently replicated and validated by Waldman et al. (2018). Regarding the CBCL/YSR, there are zero items in the oppositional/aggressive item set that correspond to a spiteful/vindictive area. We found only one study of CBCL/YSR oppositionality (Aebi et al., 2013) that included such a third factor; this was done by pulling an item from the DSM-oriented scale for conduct problems, widely regarded as distinct from oppositionality (e.g., Lahey et al., 2008). For all of these reasons, we focus on the two-factor model best supported by the available evidence; however, we also acknowledge that the number and nature of these dimensions remains an important question for further investigation.

*DSM-5*-based approach, Burke et al. (2014) tested competing models of ODD symptom dimensions across five samples (aggregate  $N = 16,280$ ) and found considerable support for a correlated bifactor model comprising irritability and defiance factors (see also Waldman, Rowe, Boylan, & Burke, 2018, for recent replication and validation). These results were extremely helpful in clarifying the multidimensional structure of youth oppositionality; however, it is important to note that *DSM-5* criteria were developed as categorical indicators for defining and diagnosing one pattern of psychopathology (Frances & Widiger, 2012). Diagnostic criteria should not be considered a gold standard for assessing psychological constructs, nor should *DSM-5* criteria, or any single measure, be the sole basis for diagnosing particular categories (Frick, Barry, & Kamphaus, 2010). Multiple methods and measures are needed for myriad purposes, such as research, screening, diagnosis, progress monitoring, and outcomes evaluation (Hunsley & Mash, 2007).

To this end, researchers have developed rating scales to assess irritability, defiance, and related constructs in youths. Notable examples include the Affective Reactivity Index (Stringaris et al., 2012a), for caregiver and youth self-report assessment of irritability; the Multidimensional Assessment of Preschool Disruptive Behavior (Wakschlag et al., 2014; Wiggins et al., 2018), which includes Temper Loss and Noncompliance subscales; and the Cranky Thermometers (Melvin et al., 2018), a developmentally sensitive, self-report, visual analog scale. Although each of these measures has some psychometric support and utility, each also has a certain practical limitation. Collectively, these limitations include being rather lengthy or time-consuming, focusing on a narrow age range, relying on a single informant, measuring only one dimension of oppositionality, or assessing constructs that are closely related but distinct from irritability and defiance. Such characteristics may restrict their potential for use outside of certain research or academic-clinical contexts.

### Brief Scales Derived from Widely Used Measures

One promising approach involves using items from broader, established measures—most often selected items from the Child Behavior Checklist or Youth Self Report (CBCL/YSR; Achenbach & Rescorla, 2001). Over the last decade, researchers have used subsets of these items to form brief scales measuring irritability, sometimes with a parallel scale for defiance. Stringaris, Zavos, Leibenluft, Maughan, & Eley (2012b) pioneered this approach (and see subsequent work by Aebi, Plattner, Metzke, Bessler, & Steinhausen, 2013; Althoff et al., 2014; Roberson-Nay et al., 2015; da Silva et al., 2014), using CBCL/YSR items to create brief irritability and defiance scales, conceptualized as dimensions of oppositionality. Later, Wiggins, Mitchell, Stringaris, and Leibenluft (2014), followed by others (Cornacchio, Crum, Coxe, Pincus, &

Comer, 2016; Salum et al., 2017; Savage et al., 2015; Tseng et al., 2017), similarly extracted a few items to create a brief scale measuring irritability without defiance. Taken together, these studies provide a body of evidence—based on exploratory, confirmatory, and conceptual approaches—suggesting that select CBCL/YSR items might be used to measure irritability and defiance.

Items from measures like the CBCL/YSR offer several potential advantages. These measures are already widely used in youth mental health research and service settings. Thus, if psychometrically well supported, irritability and defiance scales derived from CBCL and/or YSR items could contribute to research and practice in many cases without the need to administer any additional measures beyond those already being collected. Indeed, such brief scales could be especially useful for secondary analysis of data sets containing CBCL/YSR items (as in the present study). There is potential for further developmental, assessment, and intervention research; but first, psychometric examination is needed.

### The Need for Multi-Informant Evaluation in Clinical Samples

Two final aspects of the literature warrant consideration. First, most of the evidence has come from large community data sets (e.g., Althoff et al., 2014; Roberson-Nay et al., 2015; Wiggins et al., 2014), with little attention to clinically referred youths. Building on the progress made with community samples, it is important to now investigate the validity and reliability of measures among specific populations in which they are most likely to be used, whether for research or clinical purposes (American Psychological Association, 2017). Regardless of sample type, the results of research on clinical symptoms are often translated, generalized, or otherwise applied to clinical populations. In fact, findings from community samples have already influenced the creation of new diagnostic categories for youths, with important clinical implications (Evans et al., 2017; Roy et al., 2014). To enhance the relevance of the evidence base to its clinical applications, psychometric research in clinically referred samples is needed.

Second, studies have focused on either parent- or youth-report measures, or they have not compared the two approaches. Multi-informant research, involving both caregivers and youths, is needed, because parent-report properties cannot be assumed to generalize to youth-report, perhaps especially in clinical samples. For example, Hawley and Weisz (2003) found that, in a treatment-referred sample, approximately half of parent-child dyads could not agree on a single target problem, and virtually none (1.6%) showed total agreement on three out of three problems. More broadly, meta-analyses show low parent-child correlations on such measures, ranging from .22 to .28 (Achenbach, McConaughy, & Howell,

1987; De Los Reyes et al., 2015). In community samples, children report higher overall symptoms than parents; however, the opposite pattern has been reported for clinical samples (Salbach-Andrae, Klinkowski, Lenz, & Lehmkuhl, 2009). These discrepancies may be related to a number of factors, including parent mental health, family conflict, setting/context, and youth–caregiver differences in observing, encoding, and reporting accuracy (De Los Reyes & Kazdin, 2005; Dirks, De Los Reyes, Briggs-Gowan, Cella, & Wakschlag, 2012; Youngstrom, Loeber, & Stouthamer-Loeber, 2000). In addition, there are observed differences in the psychometric properties of some parallel youth and parent measures (Cole, Hoffman, Tram, & Maxwell, 2000). Some studies have found that model fit, item factor-loadings, and convergent and discriminant relationships vary across parallel parent and youth measures (e.g., Colburn, Bonadio, & Tompsett, 2018; Goodman, 2001; Storch et al., 2006). For example, the factor structure of the Strengths and Difficulties Questionnaire shows a stronger fit by parent report compared to youth report, and some items on the youth-report version do not load adequately on any one factor (Goodman, 2001). Given this evidence that parents and youths may report differently about similar constructs, multi-informant psychometric evaluation is critical.

## Study Goals and Hypotheses

The purpose of the present study was not to develop a new scale or to modify an existing one. Rather, this study provides a psychometric examination of CBCL/YSR item sets that have been used to measure irritability and defiance. Using a large aggregated sample of treatment-referred youths, we examined (a) internal consistency; (b) factor structure, and its invariance across gender and age; (c) convergent/discriminant and (d) criterion validity in relation to internalizing and externalizing problems/disorders; and (e) multi-informant patterns. Based on the literature, we hypothesized adequate fit for a correlated two-factor or bifactor model of oppositionality. Irritability was expected to be most strongly correlated with anxiety and depression, defiance with conduct, and both with ODD. Based on previous research, we expected parents to report higher levels of all problem behaviors than youths, potentially showing similar or better performance in terms of factor structure, invariance, and strength of hypothesized associations. Whether YSR items would show sufficient psychometric integrity to warrant recommended use was a question of special interest.

## METHODS

Baseline clinical assessment data for 1,030 treatment-referred youths (ages 6–15; 43.0% female) were aggregated from four outpatient mental health samples; see Table 1 for descriptive characteristics. In the aggregate sample, the availability of

CBCL and YSR data were 100.0% and 89.8%, respectively. Availability of other measures varied by subsample as indicated next. Within-sample missing data rates were less than 2%. Overall, participants were demographically and clinically diverse. Data were aggregated for analyses to increase sample diversity and size, and in consideration of the similarities across samples: All youths were identified based on very similar procedures and criteria, were seeking treatment for emotional or behavioral problems in community mental health centers, and were assessed by trained staff from the same research team (see Weisz et al., 2012). Institutional review board approval, parent consent, and youth assent were obtained.

## Measures

### *Broadband Emotional and Behavioral Problems*

The CBCL and YSR (Achenbach & Rescorla, 2001) are widely used measures of youth emotional and behavioral functioning. Items are roughly parallel in content between parent and youth versions and are rated on a 3-point scale of 0 (*not true*), 1 (*somewhat/sometimes true*), and 2 (*very/often true*). The original exploratory-empirical measurement structure of the CBCL/YSR consists of Total, Internalizing, and Externalizing Problems scales, which are composites derived from eight shorter narrowband syndrome scales such as Anxious/Depressed and Rule-Breaking. Alternative subscales have been developed through expert consensus and confirmatory-analytic methods to directly map onto *DSM* categories, such as Affective and Conduct Problems (Achenbach, Dumenci, & Rescorla, 2003). The Aggressive Behavior and *DSM* Oppositional Defiant Problems scales—from which irritability and defiance items (discussed next) were drawn—were omitted from consideration, leaving the remaining seven Achenbach System of Empirically Based Assessment syndrome and five *DSM* scales to be used in analyses (scale names are listed in the Results section). The CBCL and YSR composite, syndrome, and *DSM* scales all have ample evidence for their validity and reliability across diverse samples of youth ages 6–18 (e.g., Bellina et al., 2013; Ebesutani, Bernstein, Martinez, Chorpita, & Weisz, 2011; Ebesutani et al., 2010; Nakamura, Ebesutani, Bernstein, & Chorpita, 2009). Of interest here, these studies show that the Oppositional-Defiant Problems *DSM* has good specificity and sensitivity to ODD symptoms, criterion and predictive validity with ODD diagnoses, and longitudinal stability. In this analysis, CBCL/YSR data were available across all samples. All scales (*t* scores) were used to provide a complete picture of the results.

### *Irritability and Defiance*

The irritable and defiant dimensions of oppositionality were measured using CBCL/YSR items selected based on a review of previous studies and theoretical and practical

TABLE 1  
Sample Characteristics

	Clinically Referred Youth Samples				Combined Sample	Group Differences	
	Sample 1	Sample 2	Sample 3	Sample 4		$\chi^2$ or F (df = 3)	p
Total/Parent-Report <i>n</i>	333	307	127	263	1,030		
Youth-Report <i>n</i>	333	250	94	248	925		
Age in Years						8.93	< .001
<i>M</i> ( <i>SD</i> )	10.2 (1.7)	10.2 (2.7)	9.4 (2.6)	10.7 (2.6)	10.2 (2.4)		
Range	7–14	6–15	6–15	7–15	6–15		
Sex: Female, %	34.0	49.5	39.4	48.3	43.0	19.88	< .001
Ethnicity, %						198.90 <sup>a</sup>	< .001
White/Caucasian	46.5	81.4	85.8	32.3	58.2		
Black/African American	9.9	2.9	2.4	26.6	11.2		
Latino/Hispanic	7.5	1.6	2.4	25.5	9.7		
Asian	3.3	0.3	0.8	0.8	1.5		
Multiracial	29.1	12.4	8.7	13.7	17.7		
Other	3.0	1.3	0.0	1.1	1.7		
<i>T</i> Scores, <i>M</i> ( <i>SD</i> )							
CBCL Internalizing	64.6 (10.2)	64.0 (10.0)	67.3 (8.1)	64.2 (9.6)	64.7 (9.8)	3.60	.013
CBCL Externalizing	61.8 (10.7)	65.1 (10.4)	66.9 (8.4)	63.4 (9.7)	63.8 (10.2)	10.35	< .001
CBCL Total Problems	64.3 (9.6)	66.0 (9.3)	68.6 (6.4)	65.5 (8.0)	65.6 (8.8)	7.57	< .001
YSR Internalizing	55.7 (11.7)	52.4 (11.1)	59.0 (9.4)	56.6 (11.8)	55.4 (11.5)	9.96	< .001
YSR Externalizing	50.1 (11.6)	51.6 (11.0)	55.1 (9.7)	53.7 (10.5)	52.0 (11.1)	7.77	< .001
YSR Total Problems	54.0 (12.1)	53.6 (11.5)	59.0 (9.5)	56.9 (11.0)	55.2 (11.5)	8.30	< .001
Raw Scores, <i>M</i> ( <i>SD</i> )							
CBCL Defiance	0.97 (0.56)	1.00 (0.60)	1.10 (0.52)	0.96 (0.60)	0.99 (0.58)	1.79	.147
CBCL Irritability	1.07 (0.64)	1.18 (0.64)	1.36 (0.52)	1.12 (0.57)	1.15 (0.62)	7.57	< .001
YSR Defiance	0.57 (0.48)	0.62 (0.45)	0.74 (0.43)	0.70 (0.46)	0.64 (0.47)	5.70	.001
YSR Irritability	0.60 (0.52)	0.68 (0.58)	0.90 (0.59)	0.79 (0.58)	0.70 (0.57)	10.27	< .001

Note. Data were drawn from the initial baseline assessments (i.e., before any treatment took place) of four outpatient youth mental health samples in the United States: (a) Sample 1, a three-arm multisite randomized controlled trial (RCT) in two states in the Pacific West and Northeast; (b) Sample 2, a two-arm multisite RCT in a second northeastern state; (c) Sample 3, a multisite dissemination-implementation study in the same state as Sample 2; and (d) Sample 4, a two-arm multisite RCT in a third northeastern state. CBCL = Child Behavior Checklist; YSR = Youth Self Report.

<sup>a</sup>Chi-square test for ethnicity reflects proportions of White versus non-White participants across samples.

considerations. The literature revealed a preponderance of evidence but no clear consensus regarding which items should be assigned to which variables. The items used to measure irritability, defiance, and oppositionality have been drawn almost exclusively from the *DSM*-oriented scale, Oppositional Defiant Problems, or the overlapping and larger syndrome scale, Aggressive Behavior. In total, we identified five studies using CBCL and/or YSR items to measure both irritability and defiance and another five measuring irritability alone. Accordingly, we defined our irritability scale first based on this larger body of evidence, using the three items (temper, mood, stubborn) that have all been used to define irritability in nine of 10 studies identified (e.g., Stringaris, Zavos, Leibenluft, Maughan, & Eley, 2012b; Tseng et al., 2017). Defiance was then defined by the remaining three nonirritable items that were the most consistent indicators of defiance in five out of five studies (disobeys-home, disobeys-school, argues; e.g., Aebi et al., 2013; Althoff et al., 2014). All six items fall within the CBCL/YSR Aggressive Behavior and five of six (excluding temper) compose the Oppositional Defiant Problems,

thereby allowing comparisons with other CBCL/YSR scales. Analyses used raw item data and mean scale scores, thus retaining the original range from 0 (low) to 2 (high).

### Narrowband Anxiety and Depression Symptoms

The Revised Child and Adolescent Anxiety and Depression Scale (RCADS), available for Sample 1 only, was used to measure anxiety and depression symptoms by parent- and child-report. Complementing the CBCL/YSR Internalizing scales, the RCADS provides a more precise and thorough assessment of six specific internalizing syndromes—depression, generalized anxiety, obsessive-compulsive, panic disorder, separation anxiety, social phobia—that map onto *DSM* depressive and anxiety disorders. The anxiety scales also enhance interpretation of Sample 1's anxiety disorder results (see next). Evidence supports the reliability and validity of the RCADS in relation to converging measures, diagnoses, and in a variety of sample types (Chorpita, Moffitt, & Gray, 2005; Chorpita, Yim, Moffitt, Umemoto, & Francis, 2000; de Ross, Gullone, & Chorpita, 2002).

## Mental Health Diagnoses

The Children's Interview for Psychiatric Syndromes (ChIPS), available for Sample 1 only, is a highly structured interview designed for efficient and rigorous assessment of *DSM* diagnostic categories in youths ages 6–18. The sensitivity, specificity, and validity of the ChIPS has been well documented in community, outpatient, and inpatient samples (Weller, Weller, Fristad, Rooney, & Schecter, 2000). In this sample, parents and youth were interviewed separately by trained, masked examiners. Composite *DSM-IV* (American Psychiatric Association, 1994) diagnoses were then generated following Silverman and Nelles's (1988) integration procedures. This method used the following hierarchical algorithm: (a) If parent- and youth-report both met criteria for a condition, the diagnosis was assigned. (b) If a discrepancy arose, parent-report was emphasized for externalizing diagnoses and youth-report for internalizing. Composite diagnosis formation was performed by doctoral-level clinical psychologists with consensus on final diagnoses. Reliability was strong (overall kappas  $> .8$ ; all kappas  $> .7$ ) for parent- and youth-report interviews and for composite diagnoses (Weisz et al., 2012). Diagnosticians did not view the CBCL/YSR results prior to diagnosis, although these and other measures were often reviewed afterward to confirm the diagnosis. In the present analyses, ChIPS diagnoses were used in two ways. First, *total number of diagnoses* was calculated by summing the total number of ChIPS diagnoses assigned. This represented a transdiagnostic index accounting for any pattern or degree of comorbidity, consistent with (Weisz et al., 2012). Second, *individual diagnoses* were recoded to represent the presence of broad diagnostic clusters in accordance with current nosological frameworks. For example, "anxiety disorder" would be coded as 1 to reflect the presence of one or more diagnoses in this cluster (e.g., social, generalized, separation). Of the 333 youths assessed, 91.9% were assigned at least one diagnosis, with a median of 2 ( $M = 2.6$ ,  $SD = 1.6$ ). Diagnostic frequencies are as follows: attention deficit/hyperactivity disorder (ADHD; 55.9%), any anxiety disorder (53.2%), ODD (46.2%), any depressive disorder (34.2%), conduct disorder (16.5%), post-traumatic stress disorder (PTSD; 5.1%), and obsessive-compulsive disorder (OCD; 4.2%).

## Analytic Plan

Confirmatory factor analysis (CFA) models were estimated in Mplus Version 8 (Muthén & Muthén, 2017) using robust weighted least squares estimation (WLSMV).<sup>2</sup> Models were

specified to examine the factor structure, fit, and invariance of the irritability and defiance scales. Theta parameterization was needed for invariance models. We considered a correlated bifactor model (general oppositional factor on one side, irritability and defiance correlated factors on the other; e.g., Burke et al., 2014), and a correlated two-factor model, (correlated irritability and defiance factors; e.g., Evans et al., 2016). Ultimately, the bifactor model did not converge—possibly due to overfitting, one of several problems with bifactor models (Bonifay, Lane, & Reise, 2017)—so analyses proceeded with the correlated two-factor model. Following recommendations (Kline, 2016), CFA models were evaluated collectively based on incremental and absolute fit indices, item loadings and residuals, and theoretical considerations, with model fit evaluated at the following approximate thresholds for adequate/good: root mean square error of approximation (RMSEA)  $\leq .08/.05$ , comparative fit index (CFI)  $\geq .90/.95$ , and Tucker-Lewis index (TLI)  $\geq .90/.95$ .

Measurement invariance was tested across gender (boys vs. girls) and developmental stages (children [6–10 years] vs. adolescents [11–15 years]). We defined adolescents as 11+ because this corresponds with the American Academy of Pediatrics' demarcation of early adolescence, and it is the typical age of transition from primary to secondary schooling. Invariance was tested by imposing equality constraints and looking for decrements in model fit (i.e., chi-square difference,  $p < .05$ ; Kline, 2016). Specifically, we tested the hierarchical assumptions of (a) *configural invariance* (similarity of item-factor membership across groups; essential), (b) *metric invariance* (similarity of factor loadings across groups; essential), and (c) *scalar invariance* (similarity of intercepts across groups; nonessential but useful; Kline, 2016). Internal consistency was evaluated using alpha, omega, and mean inter-item and item-total correlations. Alpha and omega are interpreted similarly, but omega does not violate assumptions of normality and tau-equivalence, yielding a more reliable index of internal consistency (Dunn, Baguley, & Brunsten, 2014).

Next, linear regression models were used to examine convergent/discriminant validity with symptom scales (CBCL/YSR, RCADS), and logistic regression models were used to examine criterion validity with ChIPS composite diagnoses. In linear models, zero-order correlations were used to show the basic association between one dimension and, whereas semi-partial (adjusted) correlations examined this association after controlling for the other dimension. Similarly, in logistic regression models, unadjusted odds ratios (ORs) were used to describe the basic association between dimension and diagnostic outcome, with adjusted ORs showing this association after controlling for the other dimension. In all linear and logistic models, adjusted and unadjusted estimates can be interpreted similarly as standardized effect sizes for the magnitude of the association, testing two null hypotheses: (a) that the basic association is null ( $H_0: r = 0$ ,  $OR = 1$ ) and (b) that one dimension's association is not different from the other's ( $H_0:$

<sup>2</sup>WLSMV is appropriate for ordered-categorical data such as the CBCL/YSR three-point response scale. Because CBCL/YSR data are usually scored and interpreted as continuous measures, we also estimated all models using robust maximum likelihood estimator (MLR) appropriate for interval data. Results from MLR and WLSMV models were quite similar in terms of overall model structure, fit, and invariance.

$r_{\text{def}} = r_{\text{irr}}$ ,  $OR_{\text{def}} = OR_{\text{irr}}$ ). Both null hypotheses must be rejected in adjusted and unadjusted models to establish a robust differential correlate. For example, a robust differential irritability–depression link would be greater than zero *and* greater than the defiance–depression correlation, whether controlling for defiance or not. Correlations and *ORs* were compared using Fisher’s *z* transformation and 95% confidence intervals (CIs), respectively. To provide a full picture of the results, the main analyses of aggregated sample data were also performed for each of the four constituent samples (see the Supplement).

## RESULTS

Descriptive statistics and correlations are reported in Table 2. Consistent with the general pattern of CBCL/YSR results in this sample (Table 1), irritability and defiance scores were higher and showed greater variability by parent-report compared with youth-report. At the scale level, these scores were approximately normally distributed, with skewness from  $-0.22$  to  $0.46$  and kurtosis from  $-0.97$  to  $-0.44$ , without a preponderance of zero scores (7.9%–23.4%) on any scale. At the item level, responses were distributed approximately evenly

TABLE 2  
Bivariate, Univariate, and Scale Characteristics

	Parent Report		Youth Report	
	Defiance	Irritability	Defiance	Irritability
<b>Bivariate Correlations</b>				
Parent-Reported Defiance	1			
Parent-Reported Irritability	.58**	1		
Youth-Reported Defiance	.41**	.31**	1	
Youth-Reported Irritability	.24**	.27**	.54**	1
Age (Years)	-.10**	-.04	.11**	.18**
Gender (Female)	-.14**	-.02	-.06	.08*
Black/African American	.01	-.03	.05	.08*
Latino/Hispanic	-.01	-.06*	.03	.05
White/Caucasian	-.02	.07*	-.04	-.05
Other/Multiracial	.02	-.02	-.02	-.03
<b>Univariate and Scale Properties</b>				
Range	0–2	0–2	0–2	0–2
<i>M</i>	0.99	1.15	0.64	0.70
<i>SD</i>	0.58	0.62	0.47	0.57
Skewness	-0.06	-0.22	0.45	0.46
Kurtosis	-0.82	-0.97	-0.44	-0.62
Zero Inflation (%)	10.9	7.9	17.3	23.4
Cronbach’s $\alpha$	.70	.73	.61	.64
Omega Total	.74	.73	.64	.64
<i>M</i> Inter-Item Correlation	.44	.48	.34	.37
<i>M</i> Item-Total Correlation	.79	.81	.75	.76

Note. Gender and ethnicities are binary variables with point-biserial correlation coefficients. \* $p < .05$ . \*\* $p < .01$ .

or normally, except for YSR Item 22 (disobeys-home) and CBCL/YSR items 23 (disobey(s)-school) where zero responses exceeded 50%. Parent–child agreement ranged from .27 for irritability to .41 for defiant behavior. Within-rater correlations for different constructs were greater than between-rater correlations of same constructs. Age was negatively associated with parent-reported defiance and positively associated with youth-reported defiance and irritability. Girls had lower scores than boys on parent-reported defiance but higher scores than boys on youth-reported irritability. With only a few small and inconsistent exceptions, scores were uncorrelated with ethnicity (see Table 2, and Tables S1–S4).

### Internal Consistency, Measurement Structure, and Invariance

Internal consistency results are presented in Table 2 (see Tables S1–S4 for results broken down by sample). Alpha and omega coefficients were consistently adequate by parent-report but were somewhat weaker by youth-report. The mean inter-item and item-total correlations followed a similar pattern. Thus, the following youth-reported results should be interpreted with caution. Within the same informant, results were relatively similar for irritability and defiance. In the CBCL two-factor model, the model indices were mixed, suggesting marginal-to-adequate fit to the data,  $\chi^2(8) = 106.59$ ,  $p < .001$ , RMSEA = .109, 90% CI [.091, 0.128], CFI = .978, TLI = .960, weighted root mean square residual (WRMR) = 1.435. In contrast, the one-factor model,  $\chi^2(9) = 188.24$ ,  $p < .001$ , RMSEA = .139, 90% CI [.122, .157], CFI = .961, TLI = .935, WRMR = 1.987, showed significantly poorer fit,  $\chi^2_{\text{diff}}(1) = 61.36$ ,  $p < .001$ . Modification indices suggested a correlated residual between CBCL Items 22 (disobeys-home) and 23 (disobeys-school), which is understandable given their content similarity. When these item residuals were free to covary, model fit improved,  $\chi^2_{\text{diff}}(1) = 41.66$ ,  $p < .001$ , showing good fit to the data (see Figure 1, left panel). For YSR data, the initial two-factor model fit well and was ultimately retained as the final model with no modifications (see Figure 1, right panel). Although the one-factor YSR model fit the data adequately,  $\chi^2(9) = 53.90$ ,  $p < .001$ , RMSEA = .073, 90% CI (.055, .093), CFI = .976, TLI = .960, WRMR = 1.123, better fit was obtained from a correlated two-factor model,  $\chi^2_{\text{diff}}(1) = 22.18$ ,  $p < .001$ .

Table 3 presents results of tests of invariance across gender (boys vs. girls) and age (child vs. adolescent) groups. Per both informants, results showed configural invariance across gender (item-factor *membership* equivalent between groups). Metric invariance across gender (item-factor *loadings* equivalent between groups) was supported for parent-report but not for youth-report. Last, neither informant passed the test of scalar invariance (item *thresholds* being equivalent between groups; a nonessential criterion) across



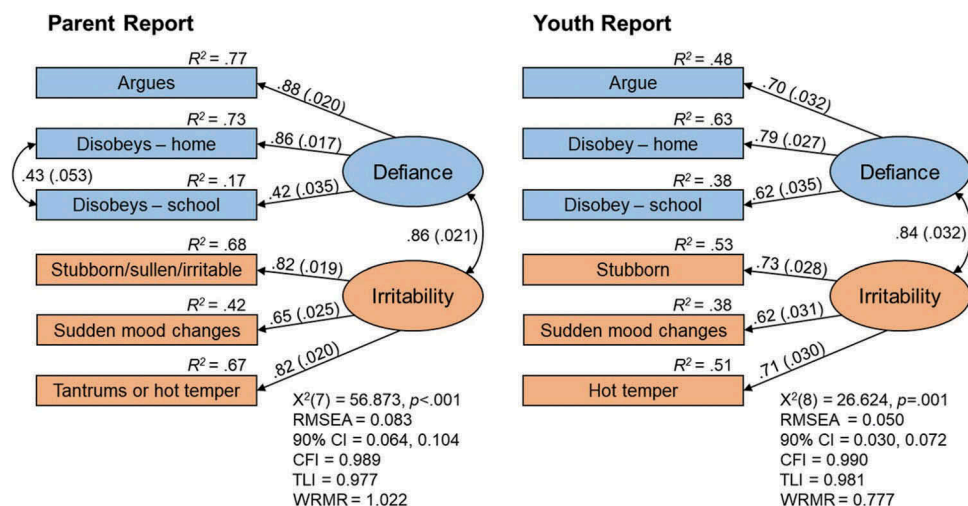


FIGURE 1 Final models with standardized estimates (Standard Errors) and fit statistics. All parameter estimates are significant at  $p < .001$ .

TABLE 3  
Results of Measurement Invariance Testing

	Model $\chi^2$			$\Delta\chi^2$			RMSEA (90% CI)			CFI	TLI	Passed?
	Est	df	p	Est	df	p	Est	LB	UB			
Parent Report												
Gender												
1. Configural	59.45	14	< .001	–	–	–	.079	.059	.101	.990	.979	Yes
2. Metric	44.91	20	.001	5.50	6	.482	.049	.030	.069	.995	.992	Yes
3. Scalar	133.24	32	< .001	77.62	12	< .001	.078	.065	.092	.978	.980	No
Age												
1. Configural	63.16	14	< .001	–	–	–	.083	.063	.104	.989	.977	Yes
2. Metric	42.77	20	.002	2.70	6	.846	.047	.027	.067	.995	.993	Yes
3. Scalar	57.81	32	.003	18.65	12	.097	.040	.023	.056	.994	.995	Yes
Youth Report												
Gender												
1. Configural	38.26	16	.001	–	–	–	.055	.033	.077	.989	.979	Yes
2. Metric	55.71	22	< .001	18.07	6	.006	.058	.039	.077	.983	.977	No
3. Scalar	116.19	34	< .001	57.16	12	< .001	.072	.058	.087	.959	.964	No
Age												
1. Configural	41.61	16	< .001	–	–	–	.059	.037	.081	.986	.973	Yes
2. Metric	50.81	22	< .001	13.09	6	.042	.053	.034	.073	.984	.978	No
3. Scalar	152.30	34	< .001	9.62	12	< .001	.087	.073	.101	.934	.942	No

Note. All models are structurally equivalent to those presented in Figure 1. The test of measurement invariance is considered “passed” if the model is not significantly worse (based on  $\Delta\chi^2$ ) when estimates are constrained to be equal across gender or age. Models were estimated and compared successively, with (a) configural invariance imposing no equality constraints, (b) metric invariance constraining loadings to be equal across groups, and (c) strong invariance constraining loadings and thresholds to be equal across groups. RMSEA = root mean square error of approximation; CI = confidence interval; LB = lower bound; UB = upper bound; CFI = comparative fit index; TLI = Tucker–Lewis index.

gender. Based on these results, we next examined gender-related differences for model estimates on the YSR and item-response patterns for both the CBCL and YSR. By self-report, models fit well across gender but slightly better for boys,  $\chi^2(8) = 11.81, p = .160, RMSEA = .030, 90\% CI [0.00, .064], CFI = .996, TLI = .992, WRMR = .513$ , as compared

to girls,  $\chi^2(8) = 25.95, p = .001, RMSEA = .075, 90\% CI [.044, .108], CFI = .983, TLI = .968, WRMR = .781$ . Girls had higher item loadings for YSR03 (argue;  $\beta = .76, SE = .047$ ) and YSR95 (hot temper;  $\beta = .80, SE = .038$ ) as compared to boys ( $\beta = .64, SE = .044; \beta = .63, SE = .044$ , respectively). Exploring item response patterns, the YSR and

CBCL both had one pronounced gender discrepancy: Item 23 (disobeys-school) was endorsed more often for boys (CBCL = 59.3%; YSR = 35.6%) than girls (CBCL = 32.3%; YSR = 17.2%); all other items showed similar ( $\pm 10\%$ ) frequencies.

Regarding age differences, CBCL models passed all three tests, demonstrating configural, metric, and scalar invariance of the CFA model between children and adolescents. Youth-reported models showed evidence of configural invariance across age groups but not metric or scalar invariance. Although the overall models fit similarly well for both YSR child ( $\chi^2(8) = 20.85, p = .008, RMSEA = .058, 90\% CI [.028, .090], CFI = .985, TLI = .973, WRMR = .684$ ), and YSR adolescent ( $\chi^2(8) = 20.77, p = .008, RMSEA = .059, 90\% CI [.028, .091], CFI = .986, TLI = .974, WRMR = .702$ ), there were age-related differences in parameter estimates. YSR23 (disobey-school) was a stronger indicator of defiance among children ( $\beta = .70, SE = .046$ ) than among adolescents ( $\beta = .52, SE = .053$ ). Regarding response pattern, YSR86 (stubborn) that showed the most prominent group differences, endorsed by 40.9% of children versus

63.9% of adolescents. Slightly smaller age-related differences were found for YSR22 (disobey-home; child = 40.3%, adolescent = 54.6%) and YSR95 (hot temper; child = 45.7%, adolescent = 59.7%); other items showed similar ( $\pm 8\%$ ) frequencies.

To summarize, the final CBCL and YSR models (Figure 1) both showed adequate fit ( $RMSEAs \leq .08, CFIs/TLIs \geq .98$ ), consistency in standardized factor loadings (all 0.6–0.9 except CBCL disobeys-school at 0.42), and similarly high factor/scale correlations per both informants (latent factor  $r_s = .86$  and  $.84$ , respectively; scale score  $r_s = .58$  and  $.54$ ). The CBCL items showed adequate internal consistency overall and measurement invariance across age and gender, but the YSR items were more questionable.

**Convergent and Discriminant Validity with Symptom Scales**

Table 4 presents the zero-order and semipartial correlations of irritability and defiance with all CBCL/YSR and RCADS-P/C scales. As shown, 10 coefficients

TABLE 4  
Common and Differential Correlations of Irritability and Defiance with Symptom Scales (Convergent/Discriminant Validity)

Scale	Domain	Parent Report				Youth Report			
		Zero-Order		Semipartial		Zero-Order		Semipartial	
		Def	Irr	Def	Irr	Def	Irr	Def	Irr
CBCL/YSR (Samples 1–4, $N = 1,030$ )									
Internalizing Problems	I	.16*	.44* <sup>†</sup>	-.11*	.43* <sup>†</sup>	.37*	.48* <sup>†</sup>	.13*	.33* <sup>†</sup>
Externalizing Problems	E	.82* <sup>†</sup>	.77*	.46* <sup>†</sup>	.36*	.78* <sup>†</sup>	.73*	.45* <sup>†</sup>	.38*
Total Problems	–	.61*	.69* <sup>†</sup>	.25*	.42* <sup>†</sup>	.61*	.66* <sup>†</sup>	.30*	.39* <sup>†</sup>
Anxious/Depressed	I	.15*	.41* <sup>†</sup>	-.10*	.39* <sup>†</sup>	.34*	.41* <sup>†</sup>	.13*	.28* <sup>†</sup>
Withdrawn/Depressed	I	.15*	.33* <sup>†</sup>	-.06	.31* <sup>†</sup>	.29*	.41* <sup>†</sup>	.08*	.30* <sup>†</sup>
Somatic Complaints	–	.07*	.22* <sup>†</sup>	-.07*	.22* <sup>†</sup>	.22*	.29* <sup>†</sup>	.08*	.20* <sup>†</sup>
Social Problems	–	.41*	.46* <sup>†</sup>	.17*	.28* <sup>†</sup>	.35*	.41* <sup>†</sup>	.15*	.26* <sup>†</sup>
Thought Problems	–	.29*	.42* <sup>†</sup>	.06*	.31* <sup>†</sup>	.35*	.47* <sup>†</sup>	.12*	.33* <sup>†</sup>
Attention Problems	–	.44* <sup>†</sup>	.38*	.27* <sup>†</sup>	.15*	.51*	.52*	.27*	.29*
Rule-Breaking	E	.68* <sup>†</sup>	.51*	.47* <sup>†</sup>	.14*	.56* <sup>†</sup>	.47*	.36* <sup>†</sup>	.20*
DSM Affective Problems	I	.26*	.44* <sup>†</sup>	.01	.36* <sup>†</sup>	.34*	.42* <sup>†</sup>	.14*	.28* <sup>†</sup>
DSM Anxiety Problems	I	.08*	.34* <sup>†</sup>	-.14*	.36* <sup>†</sup>	.27*	.34* <sup>†</sup>	.11*	.23* <sup>†</sup>
DSM Somatic Problems	–	.06	.19* <sup>†</sup>	-.06*	.19* <sup>†</sup>	.17*	.23* <sup>†</sup>	.05	.17* <sup>†</sup>
DSM ADHD Problems	–	.55* <sup>†</sup>	.46*	.35* <sup>†</sup>	.17*	.55*	.55*	.30*	.30*
DSM Conduct Problems	E	.73* <sup>†</sup>	.58*	.48* <sup>†</sup>	.19*	.63* <sup>†</sup>	.50*	.43* <sup>†</sup>	.19*
RCADS-P/C (Sample 1, $n = 333$ )									
Depression	I	.27*	.44* <sup>†</sup>	.00	.36* <sup>†</sup>	.18*	.36* <sup>†</sup>	-.04	.31* <sup>†</sup>
Generalized Anxiety	I	-.02	.24* <sup>†</sup>	-.20*	.32* <sup>†</sup>	.12*	.24* <sup>†</sup>	-.03	.21* <sup>†</sup>
Obsessive-Compulsive	–	-.04	.14* <sup>†</sup>	-.15*	.21* <sup>†</sup>	.06	.22* <sup>†</sup>	-.09	.24* <sup>†</sup>
Panic Disorder	I	.04	.23* <sup>†</sup>	-.12*	.26* <sup>†</sup>	.10	.31* <sup>†</sup>	-.10	.31* <sup>†</sup>
Separation Anxiety	I	-.06	.19* <sup>†</sup>	-.22*	.28* <sup>†</sup>	-.01	.11* <sup>†</sup>	-.09	.14* <sup>†</sup>
Social Phobia	I	-.02	.25* <sup>†</sup>	-.21*	.33* <sup>†</sup>	.20*	.38* <sup>†</sup>	-.03	.32* <sup>†</sup>

Note. Underlined cells denote hypothesized convergent/discriminant validity coefficients; all others are exploratory. Domain was assigned based on current DSM/ICD conceptualizations to help guide and display hypotheses. Def = defiance; Irr = irritability; CBCL/YSR = Child Behavior Checklist/Youth Self Report; I = internalizing; E = externalizing; DSM = Diagnostic and Statistical Manual of Mental Disorders; RCADS-P/C = Revised Children's Anxiety and Depression Scale – Parent/Child Forms.

\* $p < .05$  for correlation significance ( $H_0: r = 0$ ). <sup>†</sup> $p < .05$  for differential association significance ( $H_0: r_{def} = r_{irr}$ ).

(internalizing) were hypothesized to be more strongly associated with irritability, three coefficients (externalizing) to be more strongly linked to defiance, and the remainder were considered exploratory. Across informants and models, irritability was differentially associated with 10 of 10, and defiance with three of three, hypothesized variables. Generally, the magnitude of the convergent validity coefficients was stronger by parent- than by youth-report; the magnitude of the convergent-discriminant discrepancies was nearly twice as large for parent- than youth-report. This pattern suggests parent-report showed stronger evidence of convergent and discriminant validity than youth-report. This is consistent with the relatively lower internal consistency of youth-report scales just noted. Exploratory associations revealed a strikingly consistent pattern. Per both informants, all scales of total, social, thought, somatic, and obsessive-compulsive problems were more strongly linked to irritability than to defiance. In contrast, there was less interrater consistency on the attention/ADHD problems scales; by parent-report, both inattention measures were more strongly linked to defiance than to

irritability, but by youth-report, inattention was equivalently correlated with irritability and defiance.

### Criterion Validity with ChIPS Diagnoses

Logistic regression *ORs* from Sample 1 data are reported in Table 5. By both informants, irritability and defiance were significantly and equivalently associated with increased likelihood of ODD, consistent with hypotheses. All other results differed by informant. As expected, parent-reported defiance was a significant overall and differential predictor of conduct disorder and ADHD. Similarly, parent-reported irritability was a significant differential predictor of anxiety and depressive disorders, and a greater total number of diagnoses. Parent-reported irritability and defiance were not reliably associated with OCD or PTSD. Results showed less support for hypothesized predictions by youth-report; only one of three expected differential associations was supported. Youth-reported defiance predicted conduct disorder, both greater than chance and greater than irritability. Of interest, although youth-reported irritability did not increase risk for anxiety disorders as hypothesized, youth-reported defiance decreased

TABLE 5  
Irritability and Defiance as Common and Differential Predictors of Mental Health Diagnoses (Criterion Validity)

ChIPS Composite Diagnoses	Domain	Predictor Variables			
		Unadjusted OR [95% CI]		Adjusted OR [95% CI]	
		Defiance	Irritability	Defiance	Irritability
<b>Parent-Report Models</b>					
ADHD	–	4.28 [2.72, 6.72]* <sup>†</sup>	2.01 [1.42, 2.87]*	4.25 [2.45, 7.38]* <sup>†</sup>	0.98 [0.62, 1.54]
Anxiety	I	0.71 [0.48, 1.04]	1.47 [1.04, 2.06]* <sup>†</sup>	0.36 [0.21, 0.62]*	2.54 [1.61, 4.02]* <sup>†</sup>
Conduct	E	5.49 [2.91, 10.37]* <sup>†</sup>	1.89 [1.17, 3.03]*	6.19 [3.00, 12.78]* <sup>†</sup>	0.86 [0.48, 1.53]
Depressive	I	1.43 [0.95, 2.15]	1.81 [1.26, 2.62]*	0.92 [0.55, 1.54]	1.84 [1.17, 2.90]* <sup>†</sup>
OCD	–	0.54 [0.21, 1.42]	1.30 [0.56, 3.02] <sup>†</sup>	0.27 [0.08, 0.97]*	2.56 [0.87, 7.50] <sup>†</sup>
ODD	E	3.42 [2.22, 5.28]*	3.38 [2.30, 4.97]*	1.94 [1.16, 3.24]*	2.44 [1.56, 3.82]*
PTSD	–	1.18 [0.49, 2.82]	1.14 [0.53, 2.46]	1.13 [0.38, 3.32]	1.08 [0.42, 2.76]
<i>N</i> Diagnoses <sup>a</sup>	–	0.28*	0.38* <sup>†</sup>	0.07	0.27* <sup>†</sup>
<b>Youth-Report Models</b>					
ADHD	–	1.49 [0.83, 2.65]	0.91 [0.54, 1.54]	1.44 [0.91, 2.29]	1.17 [0.77, 1.77]
Anxiety	I	0.55 [0.31, 0.98]* <sup>†</sup>	1.08 [0.63, 1.82]	0.60 [0.38, 0.96]*	0.79 [0.52, 1.20]
Conduct	E	2.96 [1.43, 6.13]* <sup>†</sup>	0.89 [0.45, 1.76]	2.62 [1.46, 4.71]*	1.57 [0.91, 2.70]
Depressive	I	1.24 [0.69, 2.24]	1.30 [0.76, 2.23]	1.54 [0.96, 2.46]	1.51 [0.98, 2.34]
OCD	–	2.37 [0.65, 8.59] <sup>†</sup>	0.41 [0.11, 1.60]	1.38 [0.47, 4.05]	0.67 [0.22, 2.02]
ODD	E	2.34 [1.30, 4.22]* <sup>†</sup>	0.96 [0.56, 1.62]	2.34 [1.46, 3.74]*	1.56 [1.02, 2.37]*
PTSD	–	2.18 [0.66, 7.17]	1.96 [0.66, 5.81]	3.11 [1.21, 7.98]*	2.89 [1.20, 6.98]*
<i>N</i> Diagnoses <sup>a</sup>	–	0.28* <sup>†</sup>	0.15	0.26* <sup>†</sup>	0.09

Note. Data are from Sample 1 only ( $N = 333$ ). Underlined cells denote hypothesized criterion validity coefficients; all others are exploratory. Defiance and irritability were entered simultaneously predicting one outcome at a time within informant. Domain was assigned based on current DSM/ICD conceptualizations to help guide and display hypotheses. OR = odds ratio; CI = confidence interval; ChIPS = Children's Interview for Psychiatric Syndromes; ADHD = attention deficit/hyperactivity disorder; I = internalizing; E = externalizing; OCD = obsessive-compulsive disorder; ODD = oppositional defiant disorder; PTSD = posttraumatic stress disorder.

<sup>a</sup>Values are zero-order and semipartial correlation coefficients from linear models.

\* $p < .05$  for OR significance ( $H_0: OR = 1, r = 0$ ). <sup>†</sup> $p < .05$  for differential association significance ( $H_0: OR_{def} = OR_{irr}, r_{def} = r_{irr}$ ).

risk for anxiety disorders. In contrast to parent-report, for youths it was defiance, not irritability, that was associated with a greater number of total diagnoses. Youth-reported irritability and defiance showed no distinct reliable associations with ADHD, depressive disorders, OCD, or PTSD.

## DISCUSSION

This study examined the internal consistency, factor structure, invariance, and validity of the irritable and defiant dimensions of oppositionality using the CBCL/YSR items that had been used most often in previous studies of these constructs (e.g., Althoff et al., 2014; Stringaris et al., 2012b; Wiggins et al., 2018). Hypotheses regarding the measurement and associations of irritability and defiance were uniformly supported by parent-report and partially supported by youth-report. Overall, results suggest that brief scales from CBCL items can be used to measure irritability and defiance in clinically referred youths; however, the corresponding YSR scales should not be used for this purpose, at least not without further psychometric examination or refinement. As reliability is often considered a precondition for validity, results should be interpreted successively, beginning with measurement characteristics and then moving on to validity associations and multi-informant variations.

By all indices, internal consistency was supported for parent-report, but results were mixed for youth-report. Further, correlated two-factor models showed good fit by both informants, but these results were qualified in two respects. First, age and gender invariance were found for parent- but not youth-report. This does not merely reflect different levels of irritability and defiance between groups but rather indicates that items perform differently as indicators of an underlying construct. Among girls (compared to boys), the YSR model showed slightly poorer fit and suggested that self-reported temper problems and arguing may be stronger indicators of irritability and defiance, respectively. Among children (compared to adolescents), YSR disobedience at school was a stronger indicator of defiance. These results should be interpreted cautiously, not as general findings about age and gender differences but rather only about problems in the invariance of the YSR irritability and defiance items. Second, the CBCL model showed good fit only after accounting for the residual covariance between the two “disobey” items. Despite its lower factor loading, disobedience at school is highly relevant to youth adjustment—reflecting the severity, pervasiveness, and impairment of psychopathology (APA, 2013)—so this is important to assess in a measure of defiance. However, parents are not direct observers of their children’s school behavior, consistent with our results. School-related items may also perform differently at different times of year. Improved assessment of irritability and defiance across settings might be achieved by incorporating a third

perspective via corresponding Teacher Report Form (Achenbach & Rescorla, 2001).

With parent- but not youth-report having passed these initial tests of internal reliability and invariance, our subsequent evaluations of convergent, discriminant, and criterion validity should be interpreted accordingly. By both informants, patterns of common and differential associations were consistent with hypotheses and previous literature. Irritability exhibited stronger links with mood, anxiety, and internalizing problems, whereas defiance exhibited stronger associations with conduct and externalizing problems. Generally, these effects were at least moderate in magnitude (i.e.,  $r > .3$ ) and remained robust after controlling for the other subtype. Although the patterns of significance were similar across parent- and child-report, the magnitude of the effects tended to be stronger for parent-report.

In the criterion validity analyses, a similar multi-informant pattern was found, but it showed different patterns of significance for parent- versus child-report. Specifically, significant hypothesized associations were found only by parent-report, with defiance robustly and differentially increasing risk of conduct disorder, and irritability conferring similar risk for anxiety and depressive disorders. In contrast, only one of these hypothesized results was found by youth-report: Defiance increased risk for conduct disorder, but with a smaller effect size, and it did not remain significantly greater than irritability in the adjusted model. Of interest, *lower* levels of youth-reported anxiety significantly predicted conduct disorder. This pattern is consistent with the relative direction of the association of irritability and defiance with anxiety disorders; however, it was unexpected that conduct disorders would be predicted by *lower* levels of anything. This finding resonates with early typologies of antisocial behavior in which youths with conduct problems were distinguished based on the presence versus absence co-occurring anxiety, the latter being severe (Quay, 1987). More recent research has similarly shown that, among youths with ADHD, co-occurring anxiety is linked to lower levels of aggression (Falk, Lee, & Chorpita, 2017).

Important to note, the one area in which youth and parent criterion results did converge was that, per both informants, irritability and defiance were each associated with increased likelihood of ODD, and when combined in the adjusted model, neither dimension was significantly stronger than the other. Moreover, irritability and defiance were highly correlated with each other. Consistent with prior research (e.g., Burke et al., 2014; Evans et al., 2017), these findings support the conclusion that irritability and defiance are best conceptualized as subdimensions of oppositional behaviors rather than as standalone constructs.

Patterns of parent-child ratings were consistent with meta-analytic estimates of parent-child agreement (Achenbach et al., 1987; De Los Reyes et al., 2015). For example, there was greater agreement in reports of defiance than irritability, which parallels evidence for greater interrater agreement on externalizing versus internalizing problems (De Los Reyes

et al., 2015). This pattern may be due to the parent's ability to observe these behaviors directly, whereas emotional symptoms can be less observable and require more inference (De Los Reyes & Kazdin, 2005). Additionally, contextual variations may contribute to discrepancies in reporting (Dirks et al., 2012; Kraemer et al., 2003). For example, if a child is defiant only when with parents, the parent may report that the child is highly defiant but the child may report lower levels of defiance given their obedience in other contexts. Alternatively, clinically referred children may have different views on their emotions from their parents, and perhaps have difficulty identifying and rating specific emotional states (Trentacosta & Fine, 2010). Latent transdiagnostic clinical profiles have been differentiated partly based on whether irritability was reported by parent versus youth (Kircanski et al., 2017), and irritability is more likely to be identified and viewed as distressing per caregiver report (Freeman, Youngstrom, Freeman, Youngstrom, & Findling, 2011). Finally, these discrepancies may be partly explained by measurement error (De Los Reyes et al., 2015; De Los Reyes, Thomas, Goodman, & Kunday, 2013). Consistent with this hypothesis, the evidence for youth-report was consistently less robust than for parent-report across all phases of analysis, from consistency to validity. On the other hand, favorable psychometric results have been found for self-report externalizing measures among samples as young as 5 years old (Arseneault, Kim-Cohen, Taylor, Caspi, & Moffitt, 2005). Thus, youths' performance in reporting their own defiance and irritability may be influenced by a complex array of developmental, clinical, and methodological factors. Disentangling these factors is an important task for future research. For now, multi-informant assessment should be accompanied by multi-informant hypotheses to guide the interpretation of results (De Los Reyes et al., 2015).

### Limitations

Some limitations should be noted. First, like most of the extant research on irritability and oppositionality, this study used ad hoc scales and secondary analysis of existing data. However, we aimed to help address this issue by providing the first rigorous multi-informant investigation in a large clinical sample. Second, although the CBCL and YSR offer roughly parallel forms across different informants, there was some degree of incongruity in item content, which may be meaningful for construct definition (e.g., "stubborn, sullen, and irritable" on CBCL vs. "stubborn" on YSR). Developmentally sensitive content is critical (children might not know what "irritable" and "sullen" mean), but these differences in content could help explain some of the lower robustness of youth-reported results. Other age-appropriate terms (e.g., grumpy, cranky) or formats, such as visual analogue scales, might be used in future self-report assessments of irritability in youths and particularly with younger children (Melvin et al., 2018).

Third, a multi-informant approach would be advanced by including additional raters, particularly teachers. School disobedience is referenced on the CBCL but not typically observed by parents, which may contribute to parent-child discrepancies. Including teacher ratings via the parallel Teacher Report Form may provide a more complete picture of child behavior across settings. Because interrater discrepancies could introduce as much confusion as clarity, behavioral observations would help clarify the reliability and validity of these different approaches. Last, despite the large, diverse, aggregated clinical sample, RCADS and ChIPS data were available only for Sample 1; thus, results are not equally powered and robust across all variables. The diagnostic criterion validity results should be interpreted with further caution because such diagnoses may be biased toward parent report (De Los Reyes et al., 2015). Future research should include other indices of criterion validity that are less subject to criterion contamination via informant bias; here again, behavioral observations may prove especially useful. Relatedly, these findings might not generalize to other samples, such as older adolescents or nonclinical samples.

### Implications

This study helps establish support for a set of items that has been used ad hoc with limited psychometric evidence. Findings synthesize and aid in the interpretation of extant evidence while guiding future applications. Irritability and defiance are meaningful, correlated constructs that can be measured extremely efficiently using items from widely used measures, offering advantages for clinical and research utility. Findings such as these could set the stage for an array of longitudinal analyses tracking developmental trajectories associated with irritability and defiance, and for intervention studies tracking responses to prevention and treatment. Clinically, in contexts where the entire CBCL is already being given, these items may be useful as brief screeners for irritability and defiance, supplementing the interpretation of Achenbach System of Empirically Based Assessment standard scoring. Elevated scores could flag the clinician to conduct a more thorough assessment of irritability or defiance while guiding in treatment planning decisions. For example, when only defiance is elevated, clinicians may choose to administer traditional parent management interventions (Barkley, 2013), but when irritability is elevated, this may lead a clinician to consider complementing parent-focused behavioral strategies along with child-focused cognitive-behavioral approaches for anger and aggression (Lochman, Wells, & Murray, 2007; Sukhodolsky & Scahill, 2012) or for co-occurring depression, anxiety, or traumatic stress (Chorpita & Weisz, 2009). For measurement-based care, an important future direction will be to develop measures that can be administered very quickly, at no cost, repeatedly throughout treatment, to

monitor progress and guide clinical decision-making. Such efforts would benefit from item response theory and computerized testing approaches to help optimize precision, efficiency, and utility. Initiatives such as Patient-Reported Outcomes Measurement Information System are already pursuing these goals. Of note, the Patient-Reported Outcomes Measurement Information System Pediatric Anger Scale includes a six-item short form and has been developed and supported through CFA and item response theory methods (Irwin et al., 2012).

At the same time, shorter is not always better, and irritability and defiance warrant further investigation apart from the CBCL/YSR item sets. The present scales are an artifact of available data. A more rigorous assessment of irritability and defiance would require new measure development. Similarly, a better understanding of the multidimensional structure of oppositionality would need to consider a broader array of item content, such as a third dimension of spiteful/vindictive behavior measured with multiple items. For many research purposes, it may be more appropriate to consider other instruments such as the Affective Reactivity Index, Cranky Thermometers, Multidimensional Assessment of Preschool Disruptive Behavior (Melvin et al., 2018; Stringaris et al., 2012a; Wakschlag et al., 2014), and alternative methodologies, such as the Disruptive Behavior Disorders Diagnostic Observation Scale (Wakschlag et al., 2008) and the affective Posner paradigm (Tseng et al., 2017). Similarly, alternative broadband measures (e.g., Conners, 2008; Reynolds & Kamphaus, 2015) can also be useful for measuring irritability, defiance, oppositionality, and related constructs.

Findings also offer implications regarding multi-informant assessment. Although our results for parents versus youths can be seen as concerning, they are useful in addressing a perennial issue in youth mental health and developmental psychopathology: How to interpret and integrate information when parent and youth reports differ—as they often do. Our results suggest that, for irritability and defiance, we have reason for greater confidence in the reliability and validity of parent reports. These findings may have practical utility in that parent reports are frequently accessible in clinical settings, where measures such as the CBCL may be a routine part of intake assessment. This is not to suggest that one informant is “better” than the other, but rather that different informants offer different perspectives, and this should be taken into consideration in clinical and research assessments (De Los Reyes et al., 2015).

Last, this study contributes much-needed multi-informant and clinical data to a growing body of evidence showing that irritability and defiance are differentially associated with internalizing and externalizing problems. Consistent with recent conceptualizations of ODD in *DSM-5* and the International Classification of Diseases, 11th Revision, these findings suggest that irritability and defiance are clinically

important dimensions of oppositional behavior in youth. It is important to note that our analyses were designed in such a way that conclusions can be drawn about the measurement of irritability and defiance together (as a correlated two-factor scale) or of either construct on its own (as a single three-item rating scale, as has become common in irritability research). Overall, these CBCL scales strike a balance between psychometric robustness and efficiency that makes them potentially useful for measuring the irritable and defiant dimensions of youth oppositional behavior.

## ACKNOWLEDGMENTS

We thank the youths, caregivers, staff, and administrators who participated in this research.

## DISCLOSURE STATEMENT

The authors report that they have no potential conflicts of interest.

## FUNDING

We gratefully acknowledge support from the John D. and Catherine T. MacArthur Foundation (grants 83,423 and 90,231), Casey Family Programs (grant FY08-02200), the Annie E. Casey Foundation (grants 209.0037 and 211.0004), the Norlien Foundation, the Child Health and Development Institute of Connecticut (contract 14DCF6673AA), and AIM for Mental Health (AIM Clinical Science Fellowship to SCE).

## SUPPLEMENTAL MATERIAL

Supplemental data for this article can be accessed on the [publisher's website](#).

## ORCID

Spencer C. Evans  <http://orcid.org/0000-0001-8644-817X>

## REFERENCES

- Achenbach, T. M., Dumenci, L., & Rescorla, L. A. (2003). DSM-oriented and empirically based approaches to constructing scales from the same item pools. *Journal of Clinical Child and Adolescent Psychology, 32*, 328–340. doi:10.1207/S15374424JCCP3203\_02
- Achenbach, T. M., McConaughy, S. H., & Howell, C. T. (1987). Child/adolescent behavioral and emotional problems: Implications of

- cross-informant correlations for situational specificity. *Psychological Bulletin*, *101*, 213–232.
- Achenbach, T. M., & Rescorla, L. (2001). *ASEBA school-age forms & profiles*. Burlington, VT: ASEBA.
- Aebi, M., Plattner, B., Metzke, C. W., Bessler, C., & Steinhausen, H.-C. (2013). Parent- and self-reported dimensions of oppositionality in youth: Construct validity, concurrent validity, and the prediction of criminal outcomes in adulthood. *Journal of Child Psychology and Psychiatry*, *54*, 941–949. doi:10.1111/jcpp.12039
- Althoff, R. R., Kuny-Slock, A. V., Verhulst, F. C., Hudziak, J. J., & van der Ende, J. (2014). Classes of oppositional-defiant behavior: Concurrent and predictive validity. *Journal of Child Psychology and Psychiatry*, *55*, 1162–1171. doi:10.1111/jcpp.12233
- American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th ed.). Arlington, VA: Author.
- American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Arlington, VA: American Psychiatric Publishing.
- American Psychological Association. (2017, January 1). *American psychological association ethical principles of psychologists and code of conduct*. Retrieved from <http://www.apa.org/ethics/code/>
- Arseneault, L., Kim-Cohen, J., Taylor, A., Caspi, A., & Moffitt, T. E. (2005). Psychometric evaluation of 5- and 7-year-old children's self-reports of conduct problems. *Journal of Abnormal Child Psychology*, *33* (5), 537–550.
- Barkley, R. A. (2013). *Defiant children: A clinician's manual for assessment and parent training*. New York, NY: Guilford Press.
- Bellina, M., Brambilla, P., Garzitto, M., Negri, G. A., Molteni, M., & Nobile, M. (2013). The ability of CBCL DSM-oriented scales to predict DSM-IV diagnoses in a referred sample of children and adolescents. *European Child & Adolescent Psychiatry*, *22*, 235–246. doi:10.1007/s00787-012-0343-0
- Bonifay, W., Lane, S. P., & Reise, S. P. (2017). Three concerns with applying a bifactor model as a structure of psychopathology. *Clinical Psychological Science*, *5*, 184–186. doi:10.1177/2167702616657069
- Brotman, M. A., Kircanski, K., & Leibenluft, E. (2017). Irritability in children and adolescents. *Annual Review of Clinical Psychology*, *13*, 317–341. doi:10.1146/annurev-clinpsy-032816-044941
- Burke, J. D. (2012). An affective dimension within oppositional defiant disorder symptoms among boys: Personality and psychopathology outcomes into early adulthood. *Journal of Child Psychology and Psychiatry*, *53*, 1176–1183. doi:10.1111/j.1469-7610.2012.02598.x
- Burke, J. D., Boylan, K., Rowe, R., Duku, E., Stepp, S. D., Hipwell, A. E., & Waldman, I. D. (2014). Identifying the irritability dimension of ODD: Application of a modified bifactor model across five large community samples of children. *Journal of Abnormal Psychology*, *123*, 841–851. doi:10.1037/a0037898
- Burke, J. D., Hipwell, A. E., & Loeber, R. (2010). Dimensions of oppositional defiant disorder as predictors of depression and conduct disorder in preadolescent girls. *Journal of the American Academy of Child and Adolescent Psychiatry*, *49*, 484–492.
- Chorpita, B. F., Moffitt, C. E., & Gray, J. (2005). Psychometric properties of the Revised Child Anxiety and Depression Scale in a clinical sample. *Behaviour Research and Therapy*, *43*, 309–322. doi:10.1016/j.brat.2004.02.004
- Chorpita, B. F., & Weisz, J. R. (2009). *MATCH-ADTC: Modular approach to therapy for children with anxiety, depression, trauma, or conduct problems*. Satellite Beach, FL: PracticeWise.
- Chorpita, B. F., Yim, L., Moffitt, C., Umemoto, L. A., & Francis, S. E. (2000). Assessment of symptoms of DSM-IV anxiety and depression in children: A revised child anxiety and depression scale. *Behaviour Research and Therapy*, *38*, 835–855.
- Colburn, S., Bonadio, F. T., & Tompsett, C. J. (2018). Factor structure of parent- and adolescent-perceived functional impairment. *Journal of Child and Family Studies*, *27*, 2811–2823. doi:10.1007/s10826-018-1145-x
- Cole, D. A., Hoffman, K., Tram, J. M., & Maxwell, S. E. (2000). Structural differences in parent and child reports of children's symptoms of depression and anxiety. *Psychological Assessment*, *12*, 174–185.
- Conners, C. K. (2008). *Conners 3rd Edition Manual*. Toronto, ON: Multi-Health Systems.
- Cornacchio, D., Crum, K. I., Cox, S., Pincus, D. B., & Comer, J. S. (2016). Irritability and anxiety severity among youth with anxiety. *Journal of the American Academy of Child and Adolescent Psychiatry*, *55*, 54–61. doi:10.1016/j.jaac.2015.10.007
- da Silva, N. T., Schestatsky, P., Winckler, P. B., Salum, G. A., Petrocchi, A. W., & Da Heldt, E. P. S. (2014). Oppositionality and sympathetic skin response in adolescents: Specific associations with the headstrong/hurtful dimension. *Biological Psychology*, *103*, 242–247. doi:10.1016/j.biopsycho.2014.09.009
- De Los Reyes, A., Augenstein, T. M., Wang, M., Thomas, S. A., Drabick, D. A., Burgers, D. E., & Rabinowitz, J. (2015). The validity of the multi-informant approach to assessing child and adolescent mental health. *Psychological Bulletin*, *141*, 858–900. doi:10.1037/a0038498
- De Los Reyes, A., & Kazdin, A. E. (2005). Informant discrepancies in the assessment of childhood psychopathology: A critical review, theoretical framework, and recommendations for further study. *Psychological Bulletin*, *131*, 483–509. doi:10.1037/0033-2909.131.4.483
- De Los Reyes, A., Thomas, S. A., Goodman, K. L., & Kundery, S. M. A. (2013). Principles underlying the use of multiple informants' reports. *Annual Review of Clinical Psychology*, *9*, 123–149. doi:10.1146/annurev-clinpsy-050212-185617
- de Ross, R. L., Gullone, E., & Chorpita, B. F. (2002). The revised child anxiety and depression scale: A psychometric investigation with Australian youth. *Behaviour Change*, *19*, 90–101. doi:10.1375/bech.19.2.90
- Dirks, M. A., De Los Reyes, A., Briggs-Gowan, M., Cella, D., & Wakschlag, L. S. (2012). Embracing not erasing contextual variability in children's behavior: Theory and utility in the selection and use of methods and informants in developmental psychopathology. *Journal of Child Psychology and Psychiatry*, *53*, 558–574. doi:10.1111/j.1469-7610.2012.02537.x
- Dunn, T. J., Baguley, T., & Brunson, V. (2014). From alpha to omega: A practical solution to the pervasive problem of internal consistency estimation. *British Journal of Psychology*, *105*, 399–412. doi:10.1111/bjop.12046
- Ebesutani, C., Bernstein, A., Martinez, J. I., Chorpita, B. F., & Weisz, J. R. (2011). The youth Self Report: Applicability and validity across younger and older youths. *Journal of Clinical Child and Adolescent Psychology*, *40*, 338–346. doi:10.1080/15374416.2011.546041
- Ebesutani, C., Bernstein, A., Nakamura, B. J., Chorpita, B. F., Higa-McMillan, C. K., & Weisz, J. R.; Research Network on Youth Mental Health. (2010). Concurrent validity of the Child Behavior Checklist DSM-oriented scales: Correspondence with DSM diagnoses and comparison to syndrome scales. *Journal of Psychopathology and Behavioral Assessment*, *32*, 373–384. doi:10.1007/s10862-009-9174-9
- Evans, S. C., Burke, J. D., Roberts, M. C., Fite, P. J., Lochman, J. E., de la Peña, F. R., & Reed, G. M. (2017). Irritability in child and adolescent psychopathology: An integrative review for ICD-11. *Clinical Psychology Review*, *53*, 29–45. doi:10.1016/j.cpr.2017.01.004
- Evans, S. C., Pederson, C. A., Fite, P. J., Blossom, J. B., & Cooley, J. L. (2016). Teacher-reported irritable and defiant dimensions of Oppositional Defiant Disorder: Social, behavioral, and academic correlates. *School Mental Health*, *8*, 292–304. doi:10.1007/s12310-015-9163-y
- Ezpeleta, L., Granero, R., de la Osa, N., Penelo, E., & Domènech, J. M. (2012). Dimensions of oppositional defiant disorder in 3-year-old preschoolers. *Journal of Child Psychology and Psychiatry*, *53*, 1128–1138. doi:10.1111/j.1469-7610.2012.02545.x
- Falk, A., Lee, S., & Chorpita, B. F. (2017). Differential association of youth attention-deficit/hyperactivity disorder (ADHD) and anxiety with

- delinquency and aggression. *Journal of Clinical Child and Adolescent Psychology*, 46, 653–660. doi:10.1080/15374416.2015.1055858
- Frances, A. J., & Widiger, T. (2012). Psychiatric diagnosis: Lessons from the DSM-IV past and cautions for the DSM-5 future. *Annual Review of Clinical Psychology*, 8, 109–130. doi:10.1146/annurev-clinpsy-032511-143102
- Freeman, A. J., Youngstrom, E. A., Freeman, M. J., Youngstrom, J. K., & Findling, R. L. (2011). Is caregiver-adolescent disagreement due to differences in thresholds for reporting manic symptoms? *Journal of Child and Adolescent Psychopharmacology*, 21, 425–432. doi:10.1089/cap.2011.0033
- Frick, P. J., Barry, C. T., & Kamphaus, R. W. (2010). *Clinical assessment of child and adolescent personality and behavior* (3rd ed.). New York, NY: Springer.
- Goodman, R. (2001). Psychometric properties of the strengths and difficulties questionnaire. *Journal of the American Academy of Child and Adolescent Psychiatry*, 40, 1337–1345. doi:10.1097/00004583-200111000-00015
- Hawley, K. M., & Weisz, J. R. (2003). Child, parent and therapist (dis) agreement on target problems in outpatient therapy: The therapist's dilemma and its implications. *Journal of Consulting and Clinical Psychology*, 71(1), 62.
- Hunsley, J., & Mash, E. J. (2007). Evidence-based assessment. *Annual Review of Clinical Psychology*, 3, 29–51. doi:10.1146/annurev-clinpsy.3.022806.091419
- Irwin, D. E., Stucky, B. D., Langer, M. M., Thissen, D., DeWitt, E. M., Lai, J. S., ... DeWalt, D. A. (2012). PROMIS pediatric anger scale: An item response theory analysis. *Quality of Life Research*, 21, 697–706. doi:10.1007/s11136-011-9969-5
- Kircanski, K., Zhang, S., Stringaris, A., Wiggins, J. L., Towbin, K. E., Pine, D. S., ... Brotman, M. A. (2017). Empirically derived patterns of psychiatric symptoms in youth: A latent profile analysis. *Journal of Affective Disorders*, 216, 109–116. doi:10.1016/j.jad.2016.09.016
- Kline, R. B. (2016). *Principles and practice of structural equation modeling* (4th ed.). New York, NY: Guilford.
- Kraemer, H. C., Measelle, J. R., Ablow, J. C., Essex, M. J., Boyce, W. T., & Kupfer, D. J. (2003). A new approach to integrating data from multiple informants in psychiatric assessment and research: Mixing and matching contexts and perspectives. *American Journal of Psychiatry*, 160, 1566–1577. doi:10.1176/appi.ajp.160.9.1566
- Lahey, B. B., Rathouz, P. J., Van Hulle, C., Urbano, R. C., Krueger, R. F., Applegate, B., ... Waldman, I. D. (2008). Testing structural models of DSM-IV symptoms of common forms of child and adolescent psychopathology. *Journal of Abnormal Child Psychology*, 36, 187–206. doi:10.1007/s10802-007-9169-5
- Leibenluft, E., Charney, D. S., Towbin, K. E., Bhangoo, R. K., & Pine, D. S. (2003). Defining clinical phenotypes of juvenile mania. *American Journal of Psychiatry*, 160, 430–437. doi:10.1176/appi.ajp.160.3.430
- Lochman, J. E., Wells, K. C., & Murray, M. (2007). The coping power program: preventive intervention at the middle school transition. In P. Tolan, J. Szapocznik, & S. Sambrano (Eds.), *Preventing youth substance abuse: Science-based programs for children and adolescents* (pp. 185–210). Washington, DC: American Psychological Association.
- Melvin, G. A., Tonge, B. J., Mulraney, M., Gordon, M., Taffe, J., & Klimkeit, E. (2018). The Cranky Thermometers: Visual analogue scales measuring irritability in youth. *Journal of Adolescence*, 64, 146–154. doi:10.1016/j.adolescence.2018.02.008
- Muthén, L. K., & Muthén, B. O. (2017). *Mplus user's guide* (8th ed.). Los Angeles, CA: Muthén & Muthén.
- Nakamura, B. J., Ebesutani, C., Bernstein, A., & Chorpita, B. F. (2009). A psychometric analysis of the child behavior checklist DSM-oriented scales. *Journal of Psychopathology and Behavioral Assessment*, 31, 178–189. doi:10.1007/s10862-008-9119-8
- Quay, H. C. (1987). Patterns of delinquent behavior. In H. C. Quay (Ed.), *Handbook of juvenile delinquency* (pp. 118–138). New York, NY: Wiley.
- Reynolds, C. R., & Kamphaus, R. W. (2015). *Behavior assessment system for children—Third edition (BASC-3)*. Bloomington, MN: Pearson.
- Roberson-Nay, R., Leibenluft, E., Brotman, M. A., Myers, J., Larsson, H., Lichtenstein, P., & Kendler, K. S. (2015). Longitudinal stability of genetic and environmental influences on irritability: From childhood to young adulthood. *American Journal of Psychiatry*, 172, 657–664. doi:10.1176/appi.ajp.2015.14040509
- Rowe, R., Costello, E. J., Angold, A., Copeland, W. E., & Maughan, B. (2010). Developmental pathways in oppositional defiant disorder and conduct disorder. *Journal of Abnormal Psychology*, 119, 726–738. doi:10.1037/a0020798
- Roy, A. K., Lopes, V., & Klein, R. G. (2014). Disruptive mood dysregulation disorder: A new diagnostic approach to chronic irritability in youth. *American Journal of Psychiatry*, 179, 918–924. doi:10.1176/appi.ajp.2014.13101301
- Salbach-Andrae, H., Klinkowski, N., Lenz, K., & Lehmkuhl, U. (2009). Agreement between youth-reported and parent-reported psychopathology in a referred sample. *European Child & Adolescent Psychiatry*, 18, 136–143. doi:10.1007/s00787-008-0710-z
- Salum, G. A., Mogg, K., Bradley, B. P., Stringaris, A., Gadelha, A., Pan, P. M., ... Leibenluft, E. (2017). Association between irritability and bias in attention orienting to threat in children and adolescents. *Journal of Child Psychology and Psychiatry*, 58, 595–602. doi:10.1111/jcpp.12659
- Savage, J., Verhulst, B., Copeland, W., Althoff, R. R., Lichtenstein, P., & Roberson-Nay, R. (2015). A genetically informed study of the longitudinal relation between irritability and anxious/depressed symptoms. *Journal of the American Academy of Child and Adolescent Psychiatry*, 54, 377–384. doi:10.1016/j.jaac.2015.02.010
- Silverman, W. K., & Nelles, W. B. (1988). The anxiety disorders interview schedule for children. *Journal of the American Academy of Child and Adolescent Psychiatry*, 27, 772–778. doi:10.1097/00004583-198811000-00019
- Storch, E. A., Murphy, T. K., Adkins, J. W., Lewin, A. B., Geffken, G. R., Johns, N. B., ... Goodman, W. K. (2006). The children's yale-brown obsessive-compulsive scale: Psychometric properties of child-and parent-report formats. *Journal of Anxiety Disorders*, 20, 1055–1070. doi:10.1016/j.janxdis.2006.01.006
- Stringaris, A., Cohen, P., Pine, D. S., & Leibenluft, E. (2009). Adult outcomes of youth irritability: A 20-year prospective community-based study. *American Journal of Psychiatry*, 166, 1048–1054. doi:10.1176/appi.ajp.2009.08121849
- Stringaris, A., & Goodman, R. (2009a). Longitudinal outcome of youth oppositionality: Irritable, headstrong, and hurtful behaviors have distinctive predictions. *Journal of the American Academy of Child and Adolescent Psychiatry*, 48, 404–412. doi:10.1097/CHI.0b013e3181984f30
- Stringaris, A., & Goodman, R. (2009b). Three dimensions of oppositionality in youth. *Journal of Child Psychology and Psychiatry*, 50, 216–223. doi:10.1111/jcpp.2009.50.issue-3
- Stringaris, A., Goodman, R., Ferdinando, S., Razdan, V., Muhrer, E., Leibenluft, E., & Brotman, M. A. (2012a). The Affective Reactivity Index: A concise irritability scale for clinical and research settings. *Journal of Child Psychology and Psychiatry*, 53, 1109–1117. doi:10.1111/jcpp.2012.53.issue-11
- Stringaris, A., Vidal-Ribas, P., Brotman, M. A., & Leibenluft, E. (2018). Practitioner Review: Definition, recognition, and treatment challenges of irritability in young people. *Journal of Child Psychology and Psychiatry*, 59, 721–739. doi:10.1111/jcpp.12823
- Stringaris, A., Zavos, H., Leibenluft, E., Maughan, B., & Eley, T. C. (2012b). Adolescent irritability: Phenotypic associations and genetic



- links with depressed mood. *American Journal of Psychiatry*, 169, 47–54. doi:10.1176/appi.ajp.2011.10101549
- Sukhodolsky, D. G., & Scahill, L. (2012). *Cognitive-behavioral therapy for anger and aggression in children*. New York, NY: Guilford Press.
- Trentacosta, C. J., & Fine, S. E. (2010). Emotion knowledge, social competence, and behavior problems in childhood and adolescence: A meta-analytic review. *Social Development*, 19(1), 1–29.
- Tseng, W.-L., Moroney, E., Machlin, L., Roberson-Nay, R., Hettema, J. M., Carney, D., ... Brotman, M. A. (2017). Test-retest reliability and validity of a frustration paradigm and irritability measures. *Journal of Affective Disorders*, 212, 38–45. doi:10.1016/j.jad.2017.01.024
- Vidal-Ribas, P., Brotman, M. A., Valdivieso, I., Leibenluft, E., & Stringaris, A. (2016). The status of irritability in psychiatry: A conceptual and quantitative review. *Journal of the American Academy of Child and Adolescent Psychiatry*, 55, 556–570. doi:10.1016/j.jaac.2016.04.014
- Wakschlag, L. S., Briggs-Gowan, M. J., Choi, S. W., Nichols, S. R., Kestler, J., Burns, J. L., ... Henry, D. (2014). Advancing a multidimensional, developmental spectrum approach to preschool disruptive behavior. *Journal of the American Academy of Child and Adolescent Psychiatry*, 53, 82–96.e3. doi:10.1016/j.jaac.2013.10.011
- Wakschlag, L. S., Briggs-Gowan, M. J., Hill, C., Danis, B., Leventhal, B. L., Keenan, K., ... Carter, A. S. (2008). Observational assessment of preschool disruptive behavior, Part II: Validity of the Disruptive Behavior Diagnostic Observation Schedule (DB-DOS). *Journal of the American Academy of Child and Adolescent Psychiatry*, 47, 632–641. doi:10.1097/CHI.0b013e31816c5c10
- Waldman, I. D., Rowe, R., Boylan, K., & Burke, J. D. (2018). External validation of a bifactor model of oppositional defiant disorder. *Molecular Psychiatry*. Advance online publication. doi: 10.1038/s41380-018-0294-z
- Weisz, J. R., Chorpita, B. F., Palinkas, L. A., Schoenwald, S. K., Miranda, J., Bearman, S. K., ... Gray, J. (2012). Testing standard and modular designs for psychotherapy treating depression, anxiety, and conduct problems in youth: A randomized effectiveness trial. *Archives of General Psychiatry*, 69(3), 274–282. doi:10.1001/archgenpsychiatry.2011.147
- Weller, E. B., Weller, R. A., Fristad, M. A., Rooney, M. T., & Schechter, J. (2000). Children's interview for psychiatric syndromes (ChIPS). *Journal of the American Academy of Child and Adolescent Psychiatry*, 39, 76–84. doi:10.1097/00004583-200001000-00019
- Wiggins, J. L., Briggs-Gowan, M. J., Estabrook, R., Brotman, M. A., Pine, D. S., Leibenluft, E., & Wakschlag, L. S. (2018). Identifying clinically significant irritability in early childhood. *Journal of the American Academy of Child and Adolescent Psychiatry*, 57, 191–199. doi:10.1016/j.jaac.2017.12.008
- Wiggins, J. L., Mitchell, C., Stringaris, A., & Leibenluft, E. (2014). Developmental trajectories of irritability and bidirectional associations with maternal depression. *Journal of the American Academy of Child and Adolescent Psychiatry*, 53, 1191–1205. doi:10.1016/j.jaac.2014.08.005
- Youngstrom, E., Loeber, R., & Stouthamer-Loeber, M. (2000). Patterns and correlates of agreement between parent, teacher, and male adolescent ratings of externalizing and internalizing problems. *Journal of Consulting and Clinical Psychology*, 68, 1038–1050.