Measurement and correlates of irritability in clinically referred youth: Further examination of the Affective Reactivity Index

Spencer C. Evans [a,b,*], Madelaine R. Abel [c], Rachel L. Doyle [c], Hilary Skov [a,d], Sherelle L. Harmon [a,b]

[a] *Department of Psychology, Harvard University, Cambridge, MA, USA*
[b] *Department of Psychology, University of Miami, Coral Gables, FL, USA*
[c] *Clinical Child Psychology Program, University of Kansas, Lawrence, KS, USA*
[d] *Department of Psychology, Tulane University, New Orleans, LA, USA*

A B S T R A C T

Background: Research on youth irritability has proliferated in recent years, largely facilitated by items from existing measures and by key new instruments like the Affective Reactivity Index (ARI). The present study extends this literature by investigating the psychometric properties of the parent- and youth-report ARI and the correlates of irritability in an independent, clinically referred sample.

Method: Baseline assessment data were collected from 237 youths (ages 3-18; 36% female) and their parents, seen for outpatient therapy and/or assessment. We examined the ARI in terms of (1) its item, scale, and factor properties; (2) convergent/discriminant validity with internalizing, externalizing, and emotion regulation problems; (3) specificity of associations with reactive aggression, anger, dysregulation, and coping; and (4) robustness of associations after controlling for demographic variables (e.g., age, gender).

Results: The ARI's internal consistency and unidimensional factor structure were acceptable or better, with some variation across items and informants. Irritability, as measured by parent- and youth-report, was associated with variables in the externalizing (inattention, hyperactivity, executive dysfunction, aggression), internalizing (anxiety, depression, suicidality), and emotion regulation domains. Associations with reactive aggression, anger, dysregulation, and coping problems were especially pronounced. Irritability's links with internalizing and externalizing problems remained robust after controlling for demographic covariates.

Limitations: The sample was limited in diversity and moderate in size.

Conclusions: Findings support the reliability and validity of the ARI for assessing parent- and youth-rated irritability among clinically referred youth. Future research is needed to understand variations in irritability's manifestations, measurement, and correlates across demographic groups.

1. Introduction

Research has demonstrated the clinical and developmental significance of irritability in children and adolescents[1]. Several dozen empirical reports published over the last 15 years indicate that severe, chronic, and impairing irritability in youth shows unique associations with concurrent and future problems including disruptive behavior, depression, anxiety, suicidality, and other difficulties in functioning (Brotman et al., 2017; Evans et al., 2017; Vidal-Ribas et al., 2016).

However, a large portion of this research has advanced in the absence of instruments specifically designed and validated for measuring irritability. For example, studies have created variables of irritability formed from psychiatric diagnostic interview items (Brotman et al., 2006; Dougherty et al., 2015; Stringaris et al., 2009). Others have utilized items from existing scales such as the Child Behavior Checklist (Aebi et al., 2013; Althoff et al., 2014; Evans, Weisz, et al., 2020). Additionally, some have focused specifically on the subdimension of Oppositional Defiant Disorder (ODD) characterized by the touchy, angry, and

---

We have no known conflicts of interest to disclose.

* Corresponding author at: Department of Psychology, University of Miami, P.O. Box 248185, Coral Gables, FL 33124, USA.
E-mail address: sevans@miami.edu (S.C. Evans).

[1] Herein "youths," except where intended to mean children or adolescents specifically.

temper symptoms (Burke, 2012; Rowe et al., 2010; Stringaris & Goodman, 2009a, 2009b). While these ad hoc scales have some psychometric support (e.g., Burke et al., 2014; Evans et al., 2019), it remains clear that there is a need for the validation of new instruments specifically measuring irritability.

Substantial progress has been made in the development of such instruments. Measures like the Affective Reactivity Index (ARI; Stringaris et al., 2012), the Multidimensional Assessment of Preschool Disruptive Behavior (Wakschlag et al., 2014), and the Cranky Thermometers (Melvin et al., 2018) offer utility for measuring youth irritability. Of interest here is the ARI, a concise, 7-item parent- and youth-report measure. The ARI has been tested in clinical and community samples in the US, UK, and Australia (Mulraney et al., 2014; Stringaris et al., 2012), as well as translated and evaluated cross-culturally in a Brazilian community sample (DeSousa et al., 2013) and a Chinese outpatient sample (Pan & Yeh, 2018). The ARI has been translated into at least 15 languages (https://bit.ly/nimh-ari). Recently, a clinician interview version (Haller et al., 2020) and a Spanish teacher-report version (Ezpeleta et al., 2020) have been developed. Thus, the ARI stands uniquely poised as a useful measure of youth irritability with advantages including its brevity, accessibility, multiple informants, clinical cutoffs, translations, and possible convergence with clinician and teacher instruments.

Despite these advantages, the original English-language ARI has not yet been psychometrically evaluated in an independent clinical sample. Stringaris et al.'s (2012) original validation study included two samples: a U.S. sample (*N* = 218, ages 6-17), comprised of diagnostic groups with severe mood dysregulation, bipolar disorder, bipolar familial risk, and healthy volunteers; and a UK sample (*N* = 88, ages 5-18), including 34 referred for mental health services and 54 recruited from the community. Results showed strong support for the structure, validity, and utility of the ARI across groups. However, an independent investigation of the original measure's psychometric properties is still needed. Further, the associations between irritability and psychopathology (e.g., internalizing, externalizing, and emotion regulation problems) underscores the need for research investigating irritability's characteristics in general youth mental health settings (Evans, Blossom, et al., 2020; Humphreys et al., 2019; Stringaris et al., 2018). Indeed, previous ARI studies have mostly been conducted with community populations, in a language other than English, and/or with only one informant (e.g., youth-report only). Thus, there is a need for research evaluating the reliability and validity of the multi-informant ARI among clinically referred youth. As with any new assessment tool, generating and evaluating evidence for the instrument's use for its intended purposes and settings is a matter of ethical, professional, and scientific importance (American Psychological Association, 2017; National Council on Measurement, 2014).

The present study seeks to fill this gap, offering an independent psychometric examination of the ARI while also extending knowledge concerning youth irritability. Using parent- and self-report clinical assessment data, specific aims were as follows:

1 To evaluate the ARI's **item and scale characteristics**, including its internal consistency, unidimensional factor structure, item-response properties, distributional characteristics, demographic relations, and cross-informant patterns;
2 To test **convergent-discriminant validity hypotheses** that irritability would be correlated with *internalizing problems* (anxiety, depression, and suicidality), *externalizing problems* (inattention, hyperactivity-impulsivity, aggression, executive dysfunction, and peer problems), and *emotion dysregulation* (problems regulating anger, sadness, and worry).
3 To probe **specificity hypotheses** that irritability would be more strongly and uniquely associated with *reactive aggression* (relative to proactive aggression), with emotion *dysregulation* and *coping*

(relative to emotion inhibition), and with *anger* (relative to sadness and worry).
4 To examine the **robustness of irritability's associations** with internalizing and externalizing problems (Aim 2) after controlling for age, gender, race/ethnicity, and family income.

We expected the ARI to show adequate psychometric properties, including factor, item-response, and scale characteristics. The validity and specificity associations noted above were hypothesized for both youth- and parent-report, although cross-informant and multi-informant results were also examined. Finally, we anticipated that these results would be robust after adding covariates.

2. Method

2.1. Participants

Participants were 237 youths (M_{age} =9.14 years, *SD* =3.51, range 3-18; 64% male) referred for mental health services in the U.S. Midwest. Demographic and parent-report information were provided by the primary caregiver. This was typically the biological mother (79%) or father (10%), but also included adoptive parents (5%), grandparents (2%), and other relatives/caregivers (e.g., foster parent, aunt; 4%). In cases where multiple caregivers completed measures, we followed an algorithm—giving preference to complete data, biological parents, and female caregivers—to identify a single primary caregiver for analysis. Youths' racial/ethnic backgrounds were 80% White/Caucasian, 4% Black/African American, 2% Hispanic/Latinx, 11% multiracial, 2% other, and 2% no response. Some 62% of caregivers reported being married, 19% single, 19% divorced/separated, 1% living with partner, and 1% no response. Median annual household income was $50,000 (interquartile range: $27,863-$90,000); 9% did not respond. These sociodemographic data were similar to census estimates for the surrounding communities.

The sample was from a university-based training clinic, which provides psychological assessment and treatment services to community-referred youths on a low-cost sliding fee scale (Abel et al., 2020; Fite et al., 2017). Clinicians were graduate student therapists in an accredited clinical child psychology PhD program. Participants had sought therapy (48%), assessment (41%), or both (11%) between February 2013 and March 2019. Common referral concerns included disruptive behavior, anxiety, depression, inattention/hyperactivity, learning challenges, and trauma, stress, and adjustment problems. Of those for whom diagnostic data were available (*n* =134; 57%), the most frequent diagnoses were ADHD (54%), Anxiety Disorders (31%), ODD (25%), Learning Disorders (21%), Autism Spectrum Disorder (13%), and Major Depressive Disorder (12%). Most (64%) had two or more diagnoses (Table S1).

2.2. Procedures

Data were collected as part of a standard assessment battery given to all new patients. Parents provided written consent and youths gave verbal assent. Records indicate that over 90% of families who initiated services provided research consent. All families were informed that inclusion in research was voluntary and would not impact clinical care. Parent-reported measures were administered to parents of children 3-18 years of age at intake; youth-reported measures were given to those 6-18 years of age. Developmental considerations were given close attention in procedures and analyses. It can be hard to obtain reliable self-report data from younger children (e.g., < 11 years of age; Achenbach & Rescorla, 2001; American Psychiatric Association, 2013). Thus, when self-report measures were given to preadolescent children, clinicians typically read them aloud interview style. Clinicians could also omit self-report measures in cases where they were not appropriate (e.g., due to intellectual disability). As noted below, age was coded into developmentally

Table 1
Scale and Item Properties of the Parent- and Youth-Report ARI.

	ARI-Parent (N = 225)		ARI-Youth (N = 161)		Multi-Informant (Ns = 149-237)
Scale Properties					
M (SD)	4.38 (3.56)		4.22 (3.04)		Cross-informant observed r = .32**
Mdn (Range)	4.00 (0-12)		4.00 (0-12)		Cross-informant latent r = .38**
Mode	0.00		1.00		
Skewness [95% CI]	0.39 [0.07, 0.70]		0.51 [0.13, 0.88]		
Kurtosis [95% CI]	-1.06 [-1.70, -0.43]		-0.66 [-1.40, 0.09]		
Cronbach's α [95% CI]	0.88 [0.86, 0.91]		0.79 [0.74, 0.84]		
McDonald's ω [95% CI]	0.90 [0.88, 0.92]		0.80 [0.76, 0.85]		
Bivariate Correlations					
Age in Years	-0.10		-0.17*		
Gender: Male	0.08		0.17*		
Ethnicity: White	-0.00		-0.02		
Family Income	0.01		-0.13		
ARI-Parent Impairment [a]	0.76**		0.32**		
ARI-Youth Impairment [a]	0.26**		0.52**		
CFA Model Results					
WLSMV χ^2	42.12** (df = 9)		19.42** (df = 9)		109.49** (df = 53)
RMSEA [90% CI]	0.128 [0.09, 0.17]		0.085 [0.03, 0.14]		0.067 (0.05, 0.09)
CFI	0.992		0.981		0.985
TLI	0.987		0.969		0.982
WRMR	1.001		0.626		0.988
Item Properties: Observed Data	M (SD)	% 0-1-2 [b]	M (SD)	% 0-1-2 [b]	Cross-informant observed r
1. Easily annoyed	0.97 (0.78)	32-40-29	1.08 (0.71)	21-50-29	.23**
2. Often lose temper	1.00 (0.80)	32-36-32	0.81 (0.76)	40-39-21	.19*
3. Angry for long time	0.50 (0.70)	62-27-12	0.68 (0.73)	48-37-16	.33*
4. Angry most of time [c]	0.26 (0.53)	79-17-04	0.42 (0.64)	67-25-08	.11
5. Angry frequently	0.72 (0.78)	48-32-20	0.48 (0.67)	62-28-10	.32**
6. Lose temper easily	0.94 (0.81)	36-35-30	0.76 (0.80)	47-30-23	.21*
7. Causes problems [a]	0.85 (0.82)	42-31-27	0.74 (0.72)	42-42-16	.18*
Item Properties: CFA Model Results	Loading (SE)	R^2	Loading (SE)	R^2	Cross-informant residual r
1. Easily annoyed	0.75 (.041)	.560	0.44 (.090)	.190	.27
2. Often lose temper	0.95 (.017)	.903	0.81 (.046)	.661	-.25
3. Angry for long time	0.81 (.037)	.657	0.69 (.064)	.482	.57*
4. Angry most of time [c]	0.81 (.044)	.651	0.76 (.063)	.583	.11
5. Angry frequently	0.94 (.018)	.873	0.81 (.062)	.651	.72
6. Lose temper easily	0.94 (.022)	.880	0.79 (.063)	.631	-.05

Note. Loadings are standardized. Multi-informant factor models merged parent and youth models together while adding covariances connecting corresponding factors (latent r) and leftover item-level variance (cross-informant residual r). [a] ARI item 7 (irritability impairment) is not part of the total sum score, and therefore is included in the correlation and observed item-level results but not in the CFA models. [b] Represents percentage of sample selecting 0 (*not true*), 1 (*somewhat true*), or 2 (*certainly true*); may not sum to 100% due to rounding. [c] Given item 4's distribution, CFA models were re-estimated with alternative specifications of item 4 (fixing to zero, dropping, and re-binning; see Table S5). Overall, results did not show any improvement in model fit over the primary CFA models reported here.

* $p < .05$
** $p < .01$.

meaningful covariates for analysis.

2.3. Measures

Sum scores were used for all scales, such that 0 always represents the minimum possible score. Univariate characteristics are reported in Table 1 for the ARI and Table 2 for other measures.

2.3.1. Irritability

Irritability was assessed using the ARI (Stringaris et al., 2012). This parent- and youth-report scale assesses irritability in terms of threshold, frequency, and duration for angry feelings and behaviors. Participants responded on a three-point scale from 0 (*not true*) to 2 (*certainly true*). The ARI has seven items with parallel content across informants (items 1-6 toward the total score; item 7 for impairment). Evidence supports the reliability and validity of the ARI a concise measure of youth irritability (Mulraney et al., 2014; Stringaris et al., 2012). Dougherty et al. (2020) recently compared several parent- and youth-report irritability measures, with results showing strongest support for the ARI across ages. The self-report ARI has been administered to children as young as 6-8 years (Evans et al., 2020; Kalvin et al., 2020; Pan & Yeh, 2019; Stringaris et al., 2012), with Stringaris et al.'s US sample showing no differences by age. In the present study, the youth-report ARI sample (n =161) had a median age of 10.36 years (M =10.00, SD =2.97), with

about half (n =89; 55%) being younger than 11.

2.3.2. Disruptive Behavior

Parents completed the Conners 3rd Edition, Parent Short Form (Conners, 2008) to assess the severity of externalizing-related behavior problems. The Conners includes scales for inattention, hyperactivity-impulsivity, learning problems, executive dysfunction, aggression, and peer problems. Items are rated on from 0 (*never*) to 2 (*always*). The Conners has a reliable measurement structure with strong evidence for internal reliability and convergent validity (Conners, 2008).

2.3.3. Anxiety

The Revised Children's Manifest Anxiety Scale 2nd Edition, Short Form (RCMAS; Reynolds & Richmond, 2008) was used to assess total anxiety. Youths responded to 10 items regarding social anxiety, worry, and physiological symptoms, indicating whether each item was true (*yes* =1, *no* =0). The RCMAS has shown evidence of adequate to good internal consistency and construct validity among culturally diverse youth samples (Ahmad & Mansoor, 2011; Ang et al., 2011; Lowe, 2015).

2.3.4. Depression

Depressive symptoms were assessed using the Children's Depression Inventory 2nd Edition: Self-Report, Short Form (CDI; Kovacs, 2010).

Table 2
Study Variables and their Zero-Order Correlations with ARI Irritability.

	Univariate Characteristics			Correlations with ARI	
	M (*SD*)	*N*	Range	ARI-Parent (*N*=225)	ARI-Youth (*N*=161)
Externalizing Problems (Conners)					
Parent Report					
Inattention	7.48 (4.70)	180	0-15	**.26****	.23**
Hyperactivity	7.03 (5.80)	180	0-18	**.38****	.34**
Learning Problems	5.35 (4.36)	179	0-15	**.12**	.15
Executive Function	7.25 (4.14)	178	0-15	**.24****	.22**
Aggression	2.45 (3.05)	180	0-13	**.66****	.33**
Peer Problems	4.36 (4.08)	179	0-15	**.34****	.06
Internalizing Problems					
Youth Report					
Anxiety (RCMAS)	3.64 (2.78)	137	0-10	-.07	**.32****
Depression (CDI)	5.90 (4.25)	154	0-21	.17*	**.42****
Suicidality (SBQ)	1.39 (2.61)	142	0-12	.12	**.23****
Functions of Aggression (PRA)					
Parent Report					
Reactive Aggression	5.49 (3.26)	217	0-12	**.66****	.28**
Proactive Aggression	1.39 (2.02)	218	0-10	**.51****	.27**
Youth Report					
Reactive Aggression	4.08 (2.94)	154	0-12	.20*	**.53****
Proactive Aggression	0.79 (1.44)	154	0-8	.01	**.33****
Regulation of Emotions (CEMS)					
Parent Report					
Anger Coping	3.14 (2.24)	215	0-8	**-.62****	-.24**
Anger Inhibition	1.65 (2.00)	215	0-8	**-.33****	-.19*
Anger Dysregulation	2.78 (1.74)	215	0-6	**.63****	.27**
Sadness Coping	4.33 (2.38)	202	0-10	**-.32****	-.19*
Sadness Inhibition	2.23 (2.27)	202	0-8	**.03**	-.05
Sadness Dysregulation	2.81 (1.71)	202	0-6	**.31****	.06
Worry Coping	3.05 (1.49)	219	0-6	**-.30****	-.02
Worry Inhibition	2.58 (2.17)	219	0-8	**-.01**	-.17*
Worry Dysregulation	1.95 (1.74)	219	0-6	**.19****	.08
Youth Report					
Anger Coping	4.19 (2.11)	160	0-8	-.30**	**-.41****
Anger Inhibition	3.46 (2.15)	159	0-8	-.12	**-.23****
Anger Dysregulation	1.99 (1.56)	160	0-6	.23**	**.49****
Sadness Coping	5.23 (2.15)	161	0-10	-.29**	**-.20***
Sadness Inhibition	3.80 (2.36)	161	0-8	-.12	**.04**
Sadness Dysregulation	2.26 (1.64)	161	0-6	.08	**.21****
Worry Coping	3.33 (1.50)	160	0-6	-.17*	**-.19***
Worry Inhibition	4.02 (2.32)	160	0-8	-.10	**.06**
Worry Dysregulation	1.81 (1.54)	161	0-6	.03	**.28****

Note. Boldface indicates correlations where variables were rated by the same informant. RCMAS = Reynolds Child Manifest Anxiety Scale; CDI = Children's Depression Inventory; SBQ = Suicide Behavior Questionnaire; PRA = Proactive Reactive Aggression; CEMS = Children's Emotion Management Scales.

* *p*<.05
** *p*<.01.

Each item of this 12-item self-report measure has three distinct response options ranging from low (0) to high (2) severity. Youths were instructed to choose which of three statements best described them in the past 2 weeks. The CDI has shown good internal consistency and test-retest stability (Abel et al., 2020; Bae, 2012).

2.3.5. Suicidality

Suicidal thoughts and behaviors were assessed using the 4-item self-report Suicide Behavior Questionnaire (SBQ; Cotton & Range, 1993). Each item is rated on a distinct Likert scale: "Have you ever thought about or tried to kill yourself?" (0 =*Never* to 5 =*All the Time*); "How many times have you thought about killing yourself?" (0 =*Never* to 5 =*All the Time*); "Have you ever told someone that you were going to kill yourself?" (0 =*Never* to 2 =*Often*); and "Do you think you might kill yourself someday?" (0 =*Never* to 4 =*Definitely Will*). The SBQ has been found to accurately measure suicide risk in youth, with good test-retest stability and adequate internal consistency (Campos & Holden, 2019; Cotton et al., 1995; Cotton & Range, 1993; Osman et al., 2001).

2.3.6. Proactive and Reactive Aggression

The 6-item, parent- and youth-report versions of the Proactive/Reactive Aggression Rating Scale (PRA; Dodge & Coie, 1987) were used to measure proactive aggression (three items; e.g., "threatens or bullies others to get own way") and reactive aggression (3 items; e.g., "when teased or threatened, gets angry easily and strikes back"). Participants responded on a 5-point Likert scale from 1 (*never*) to 5 (*almost always*). The PRA has ample evidence for its reliability and validity as a measure of youth aggression (Evans et al., 2018; Fite et al., 2009).

2.3.7. Emotion Regulation

The Children's Emotion Management Scales (CEMS) (Zeman et al., 2001, 2002, 2010) was used to examine how youth managed their *sadness* (12 items), *anger* (11 items), and *worry* (13 items). Within each emotion, CEMS subscales assess different patterns of regulation: *inhibition, dysregulation*, and *coping* (3-5 items each). Inhibition measures the suppression of emotional expression (e.g., "I get sad inside but don't show it"), dysregulation measures inappropriate emotional expression (e.g., "I do things like slam doors when I'm mad"), and coping measures adaptive emotion regulation (e.g., "I keep myself from losing control of my worried feelings"). Items are rated from 1 (*hardly ever*) to 3 (*often*). Evidence has supported the reliability and validity of the CEMS suite in several samples (Evans, Blossom, et al., 2020; Folk et al., 2014; Zeman et al., 2001, 2002, 2010).

2.4. Analytic Approach

First, univariate descriptive, distributional, and scale characteristics of the ARI were examined. Internal consistency was assessed via alpha and omega coefficients, the latter being recommended over alpha because it does not assume tau equivalence (McNeish, 2018). Next, confirmatory factor analysis (CFA) models were estimated to evaluate the ARI's unidimensional structure. These analyses used robust weighted least squares estimation (WLSMV) in Mplus Version 8, appropriate for ordered-categorical data such as 0-1-2 response scales (Muthén & Muthén, 2019). Models were evaluated according to fit statistics (with approximate thresholds for good/acceptable drawn at CFI and TLI =0.95/0.90, RMSEA =0.05/0.10) and consistency among factor loadings (Kline, 2016; Little, 2013). For a converging perspective on scale characteristics, we also analyzed ARI data in an item response theory (IRT) framework. These analyses were conducted in Mplus using robust maximum likelihood estimation (MLR) and a graded response model appropriate for ordered-categorical data (Linden & Hambleton, 1997). Although similar to CFA with categorical items, IRT offers additional information about items and scales at different levels of severity (Kline, 2016; Muthen, 2012). Here, we treat CFA as our primary latent variable framework due to its flexibility for evaluating overall fit, for testing item and factor correlations and for modifying models. As a secondary method, IRT is used particularly to visualize item and scale properties. Full CFA and IRT results are reported in the Supplement.

Hypotheses concerning convergent and discriminant validity associations with emotion regulation and aggression variables were tested through correlation/regression approaches. These included assessing the magnitude and significance of associations in *absolute* terms (via zero-order correlations) and in *relative* terms (after controlling for other variables, and testing for the equality of regression coefficients using model Wald tests). Further models tested whether associations with internalizing and externalizing problems remained significant and robust after adding demographic covariates, defined as follows: early childhood ($\leq$7 years of age), adolescence ($\geq$11 years), non-White ethnic minority (1 =White), and annual family income (1 =$\geq$\$50K), else =0. Inspection of variance inflation factors (VIFs) and zero-order correlations (see Table S2) indicated multicollinearity was acceptably low for all multiple regression models. Missing data rates were approximately 1-20%, in line with practical and statistical allowances (Mair, 2018). Nearly all of the 225 parent ARI cases also had data on the PRA and CEMS scales (data availability: 90-97%), and most had Conners scales (79-80%). Of the 161 youth-reported ARI cases, nearly all had PRA, CEMS, and CDI scale data (96-100%), and most had RCMAS and SBQ (85-88%). Missingness was addressed through full information maximum likelihood estimation.

3. Results

3.1. Item, Scale, and Univariate Characteristics

As shown in Table 1, ARI irritability scores per both informants were semi-normally distributed (skewness and kurtosis < |1.5|), appropriate for path models with robust estimation (Kline, 2016; Muthén & Muthén, 2019). Distributions for both informants covered the full range (0-12) with means and medians around 4. There was a right skew, with the mode being 0 and 1 for parent- and youth-report, respectively (see histograms in Figures S1 and S2). Internal consistency was good (alpha/omega range: 79-90, with 95% CIs above .70), but higher for parents than for youths (.88-.90 and 79-80, respectively).

In informant-specific CFA models, the unidimensional structure fit the data well according to CFI/TLI estimates (all > 0.95). RMSEA was questionable for parent-report, suggesting possibly poor fit; however, because the 90% CIs reached down to 0.09, RMSEA conclusions about fit are indeterminate. Further, RMSEA is biased to indicate poor fit when *N* and *df* are low (Kenny et al., 2015). After merging parent and youth data

into a multi-informant CFA model with correlated factors, the one-factor model fit the data well per all indicators (CFI/TLI > 0.98, RMSEA < 0.07).

Some item-level results are noteworthy. All response categories were frequently endorsed, with items tapping *annoyance* and *temper* tending to be endorsed at slightly higher rates/levels than items tapping *anger*. On item 4 ("angry most of the time"), majorities of parents and youths responded with "0." But this item's distribution did not seem to have a detrimental effect on model fit (Table S5). It is also notable that item 1 ("easily annoyed") showed low factor loadings, especially by youth-report. Overall, however, loadings were relatively consistent and high by parent-report (.75-.94) and youth-report (.44-.81). Total ARI scores were correlated with irritability-related impairment (item 7) ratings from the same informant (rs =.52-.76) and across informants (rs =.26-32).

Cross-informant ARI correlations were medium (r =.32) for observed scores[2] and slightly higher (r =.38) for CFA latent variables with no measurement error. Multiple regression models showed that the parent-youth correlation in ARI scores was not moderated by age or gender (ps > .21 for interaction terms), and it remained robust (ps < .001 for direct effects) after controlling for these covariates. At the item level, interrater correlations were more modest and variable (rs =.11-.33). These item-level cross-informant correlations generally attenuated to non-significance[3] in CFA models where cross-informant item residuals were allowed to correlate. See Tables S3-S5 for complete CFA results.

Results of IRT models (Table S6) offered further confirmation of the CFA results. At the scale level, total information curves (Fig. 1) showed that parent- and youth-rated ARI scores offered at least adequate reliability across most of the spectrum of irritability seen in youth outpatient care. Still, parent-rated irritability seemed to offer better reliability ("excellent") than youth-report ("good") at most levels of severity. Results suggest that youth-report provides somewhat less reliable information, particularly at the extremes—i.e., very high and very low levels of irritability. At the item level, there was generally clean separation of the ordered response categories, with probabilities clearly varying as a function of the latent irritability variable (Figures S3-S4). Item discrimination was stronger by parent-report and lower by youth-report. Item 1, "easily annoyed," showed somewhat lower discrimination, especially by youth-report.

3.2. Zero-Order Correlations

Table 2 presents the zero-order correlations of ARI scores with scales measuring externalizing, internalizing, aggression, and emotion regulation problems, plus the descriptive statistics for those scales (see Table S2 for other bivariate correlations). As anticipated, parent-reported irritability was significantly and positively correlated with all scales related to disruptive behavior: inattention, hyperactivity-impulsivity, executive dysfunction, aggression, and peer problems, with moderate to large effect sizes (rs =.26-.66). The discriminant validity correlation with learning problems was nonsignificant. Similar but more modest convergent validity associations emerged for youth-reported irritability, which was linked to inattention, hyperactivity-impulsivity, executive dysfunction, and aggression (rs =.22-.34). Youth-reported irritability was moderately associated with anxiety, depression, and suicidality (rs =.23-.42), but these variables showed little to no association with

---

[2] Given that ARI data showed some zero-inflation (Table 1; Figures S1-S2), we estimated the cross-informant correlation using linear (Pearson's r=.32, p<.001) and nonparametric (Spearman's ρ=.29, p<.001) approaches. The difference was minimal.

[3] With one exception: a significant residual correlation (r= .57) on item 3, "angry for a long time." This suggests there was significant cross-informant variance related to sustained anger that was not accounted for by total irritability.
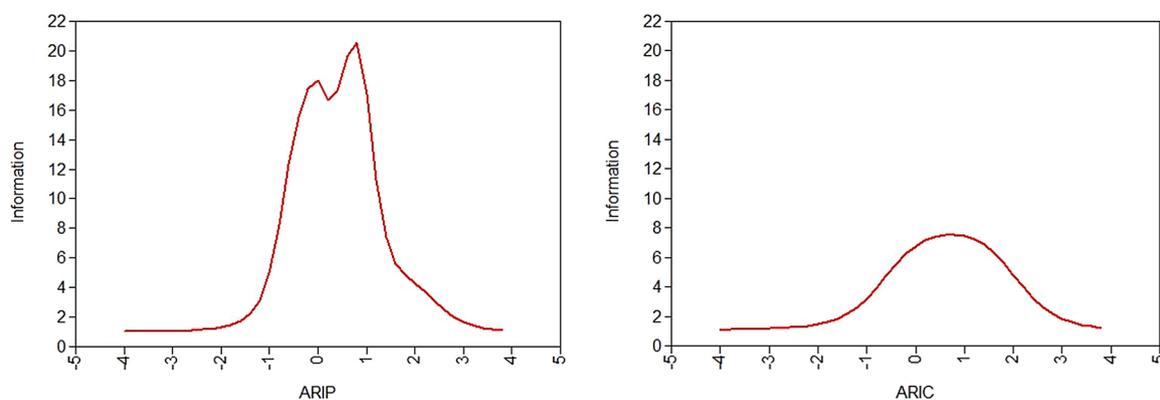
Fig. 1. Test Total Information Curve for Parent-Report (Left) and Youth-Report (Right) ARI Data.
Note. Charts show the amount of reliable information (Y-axis) the ARI obtained about irritability at varying levels of severity (X-axis; $M = 0$, $SD = 1$). Information values of 4 and 9 roughly correspond to internal consistency values commonly considered thresholds for good ($\alpha \approx .80$) and excellent ($\alpha \approx .90$), respectively. Extrapolating from the observed means and standard deviations, the relevant range on these charts—corresponding to the measure's possible total score range of 0 to 12—is approximately -1.2 to 2.1 for parent-report and -1.4 to 2.6 for youth-report. That is, values on the X-axes falling outside of these ranges are hypothetical according to model assumptions but are not actually obtainable on the ARI. Thus, the ARI generally offers at least good reliable information about irritability latent variables per both informants, with the possible exception of the most extreme high and low values possible for the scale, particularly by youth-report.

parent-rated irritability ($rs =−.07$-.17).

In considering irritability's correlations with aggression and emotion regulation variables (bottom two-thirds of Table 2), we focus here on the overall pattern of results emerging from same-informant correlations.[4] Irritability was strongly associated with reactive aggression ($rs =.55$-.66) and moderately with proactive ($rs =.33$-.51). Correlations between the ARI and CEMS scales allow us to consider associations with *specific regulation patterns* (coping, inhibition, dysregulation) pooled across different emotions, and with *specific emotions* (anger, sadness, worry) pooled across different regulation patterns. Regarding emotion regulation patterns, irritability was strongly associated with emotion dysregulation ($rs =.19$-.63) and coping ($rs =−.62$-−.19), but was less consistently linked to inhibition ($rs =−.33$-.06). Regarding specific emotions, irritability showed pronounced and consistent correlations with anger problems ($|rs| =.23$-.63)—less so for sadness and worry ($|rs| =.01$-.32).

3.3. Specificity of Associations with Aggression and Emotion Regulation Variables

To probe the specificity of irritability's correlates, we estimated a series of multiple regression models testing different sets of aggression and emotion-regulation variables predicting same-informant irritability. In this section, we first outline the analytic strategy and then present results by informant. The logic of our approach can be seen in Table 3, with Parent (top) and Youth (bottom) Models 1-11 estimated in sequence. First, Model 1 included PRA aggression subtypes as predictors to test whether irritability would be differentially correlated with reactive aggression, to a greater extent than, and after controlling for, proactive aggression. Next, CEMS scales were used to parse irritability's unique associations with different *emotions* (anger, sadness, worry) in relation to different emotion *regulation patterns* (coping, inhibition, dysregulation). Models 2-4 tested one regulation pattern at a time across all three emotions. For example, Model 2 examines irritability's unique

and total associations with *emotion coping* by including terms for anger coping, sadness coping, and worry coping. Conversely, Models 5-7 focused on one emotion at a time while pulling in all three regulation patterns. For example, Model 5 tests irritability's unique and total associations with *anger regulation* by including terms for anger coping, anger inhibition, and anger dysregulation. We concluded by examining all nine CEMS scales together (Model 8) and alongside aggression subtypes (Model 9), to parse their unique and total associations with irritability. These composite models were then trimmed (Models 10-11) by removing non-significant terms to determine the key predictor variables accounting for a sizeable portion of the variance in irritability.

Results of parent-report Model 1 (Table 3, top-left) showed that irritability was associated with reactive aggression ($\beta =.55$) after controlling for proactive aggression ($\beta =.20$). Both coefficients were statistically significant and equivalent. In Models 2-4 testing regulation patterns across emotions, anger showed the most pronounced associations with irritability scores ($|\beta s| =.44$-.62), all significantly greater than corresponding coefficients for sadness and worry ($|\beta s| \leq .18$). Models 5-7 showed that difficulties *coping* with emotions ($|\beta s| =.26$-.35) were uniquely and consistently associated with irritability within each of the three emotions. Further, irritability was linked to dysregulation of anger and sadness ($\beta s =.23$-.39) but not worry; and irritability was linked to inhibition only for sadness ($\beta =.21$). Ultimately, all aggression and emotion variables accounted for up to two-thirds (66%) of the variance in irritability (Models 8-9). Yet, of the 11 predictors, three stood out: reactive aggression, anger coping, and anger dysregulation. These three variables were stronger predictors of irritability than all others (Model 9); and each showed unique but equivalent associations with irritability (Models 9-10). When isolated in Model 10, they accounted for 59% of the variance in parent-rated irritability.

Similarly, results of youth-report Model 1 revealed a strong association of reactive aggression with irritability ($\beta =.48$; Model 1), controlling for the nonsignificant effect of proactive aggression ($\beta =.09$). In line with parent results, Models 2-4 showed that problems with *anger* ($|\beta s| =.37$-.46|) were uniquely and more strongly related to irritability, relative to problems with sadness and worry ($|\beta s| \leq .20$). In emotion-specific models (Models 5-7), irritability was uniquely associated with poor coping ($|\beta s| =.19$-.25) and dysregulation ($|\beta s| =.19$-.40), and these links were significant and equivalent within each model. Emotion inhibition showed no unique association with irritability ($|\beta s| \leq. 12$). When emotion regulation and aggression variables were combined (Models 8-9), this accounted for up to 45% of irritability's variance. Five variables stood out as the strongest significant and unique predictors of

---

[4] The corresponding cross-informant correlations were predictably smaller. Although we interpret results collectively across both informants, it is also informative to interpret these results separately for parent- and youth-reported data (see boldface coefficients in Table 2). Generally, correlations among parent-rated variables tended to be slightly higher than their corresponding correlations among youth-rated variables. Otherwise, the two informants showed virtually identical results.

Table 3

Multiple Regression Models Examining Subtypes of Aggression and Emotion Regulation as Predictors of Irritability.

Predictor Variables (β)	Aggression (PRA) Model 1: Rea-Pro Aggression	Regulation Patterns Across Different Emotions (CEMS) Model 2: Emotion Coping	Model 3: Emotion Inhibition	Model 4: Emotion Dysreg	Regulation Patterns Within the Same Emotion (CEMS) Model 5: Anger Regulation	Model 6: Sadness Regulation	Model 7: Worry Regulation	Composite Models Model 8: All CEMS	Model 9: CEMS + PRA	Trimmed Models Model 10: Trim 1	Model 11: Trim 2
Parent Models											
Rea Aggression	.55*[a]	-	-	-	-	-	-	-	.36*[a]	.35*[a]	
Pro Aggression	.20*[a]	-	-	-	-	-	-	-	.05[bc]	-	-
Anger Coping	-	−.62*[a]	-	-	−.35*[a]	-	-	−.28*[ab]	−.17*[ab]	−.25*[a]	-
" Inhibition	-	-	−.44*[a]	-	−.05[b]	-	-	−.12*[bc]	−.07[bc]	-	-
" Dysreg	-	-	-	.61*[a]	.39*[a]	-	-	.39*[a]	.28*[ab]	.30*[a]	-
Sadness Coping	-	.04[b]	-	-	-	−.26*[a]	-	.02[c]	.00[c]	-	-
" Inhibition	-	-	.18*[b]	-	-	.21*[a]	-	.11[bc]	.04[c]	-	-
" Dysreg	-	-	-	.14*[b]	-	.23*[a]	-	.13[bc]	.07[bc]	-	-
Worry Coping	-	−.09[b]	-	-	-	-	−.27*[a]	−.12[bc]	−.06[bc]	-	-
" Inhibition	-	-	.09[b]	-	-	-	−.01[b]	.05[bc]	.08[bc]	-	-
" Dysreg	-	-	-	.05[b]	-	-	.07[ab]	.01[c]	.05[bc]	-	-
Model VIFs ≤	1.47	1.41	1.74	1.26	1.82	1.57	1.56	2.34	2.43	1.57	-
Model R²	.467	.412	.137	.446	.477	.159	.092	.541	.656	.587	-
Youth Models											
Rea Aggression	.48*[a]	-	-	-	-	-	-	-	.34*[a]	.33*[a]	.33*[a]
Pro Aggression	.09[a]	-	-	-	-	-	-	-	−.01[b]	-	-
Anger Coping	-	−.40*[a]	-	-	−.25*[ab]	-	-	−.21*[ab]	−.19*[ab]	−.20*[a]	−.23*[a]
" Inhibition	-	-	−.37*[a]	-	−.08[a]	-	-	−.21*[ab]	−.18*[ab]	−.07[a]	-
" Dysreg	-	-	-	.46*[a]	.40[b]	-	-	.33*[a]	.20*[ab]	.24*[a]	.24*[a]
Sadness Coping	-	.02[b]	-	-	-	−.21*[a]	-	−.02[b]	−.03[b]	-	-
" Inhibition	-	-	.18*[b]	-	-	.12[a]	-	.13[b]	.14[ab]	-	-
" Dysreg	-	-	-	.05[b]	-	.19*[a]	-	.00[b]	.01[bc]	-	-
Worry Coping	-	−.03[b]	-	-	-	-	−.19*[ab]	−.01[b]	.01[b]	-	-
" Inhibition	-	-	.10[b]	-	-	-	.09[a]	.09[b]	.08[b]	-	-
" Dysreg	-	-	-	.20*[b]	-	-	.29*[b]	.25*[a]	.18*[ab]	.14*[a]	.13*[a]
Model VIFs ≤	1.31	1.48	1.78	1.36	1.28	1.12	1.03	2.13	2.09	1.40	1.40
Model R²	.288	.165	.097	.294	.316	.085	.119	.384	.452	.430	.428

Note. In all models, sets of aggression/emotion-regulation predictor variables were rationally selected from those listed on the left and entered as predictors of ARI scores. Parent and youth models (top and bottom, respectively) report associations among same-informant variables. Within each model, standardized beta coefficients with the same letter were equivalent at $p > .05$ according to model Wald tests for the equality of their absolute values. PRA = Proactive Reactive Aggression, CEMS = Children's Emotion Management Scales, rea = reactive, pro = proactive, dysreg = dysregulation. VIFs = variance inflation factor estimates.

* $p < .05$.

Table 4

Irritability Predicting Internalizing and Externalizing Problems, Controlling for Demographic Variables.

Predictor Variables	Externalizing (Conners Parent Report)						Internalizing (Youth Report)		
	Inattention	Hyperactivity-Impulsivity	Learning Problems	Executive Dysfunction	Aggression	Peer Problems	Anxiety (RCMAS)	Depression (CDI)	Suicidality (SBQ)
Parent Models, *B*									
ARI Irritability	0.31**	0.57**	0.14	0.23*	0.57**	0.35**	0.04	0.25+	0.13+
Gender: Male	2.17**	3.15**	1.53*	1.46*	0.68+	1.50*	-0.88	-1.00	-0.42
Ethnicity: White	0.37	-0.49	-0.77	0.33	0.29	-0.08	-1.78*	-0.94	-1.15
Income ≥ $50K	-2.43**	-2.82**	-0.74	-1.81**	-0.55	0.67	-0.31	-1.00	-0.91+
Age: Young Child (≤7)	-0.54	1.07	-0.48	-1.24	0.28	-0.98	-0.65	0.20	0.55
Age: Adolescent (≥11)	0.93	-1.60+	0.46	1.41+	0.03	0.31	-0.90	0.20	0.37
Youth Models, *B*									
ARI Irritability	0.36*	0.55**	0.27+	0.33*	0.42**	0.24+	0.31**	0.59**	0.16+
Gender: Male	2.77**	4.52**	2.15**	2.01**	1.28*	1.94**	-0.72	-0.49	-0.03
Ethnicity: White	1.26	0.10	0.10	0.52	0.70	0.06	-1.71**	-0.98	-1.14
Income ≥ $50K	-1.48+	-1.32	0.72	-1.34+	-0.11	2.24**	0.08	-0.05	-0.64
Age: Young Child (≤7)	-0.32	0.94	-0.33	-1.10	0.19	-1.14	-0.73	-0.14	0.55
Age: Adolescent (≥11)	2.14*	0.64	1.40	2.11*	1.32*	1.05	-0.42	1.12	0.84+

Note. RCMAS = Reynolds Child Manifest Anxiety Scale; CDI = Children's Depression Inventory; SBQ = Suicide Behavior Questionnaire.

+ *p* < .10
* *p* < .05
** *p* < .01.

irritability: reactive aggression, worry dysregulation, anger coping, anger inhibition, and anger dysregulation. When isolated in Model 10, they accounted for 43% of the variance. Anger inhibition attenuated to nonsignificance when entered mainly alongside other anger variables (Models 10 and 5). Dropping anger inhibition simplified the model, yielding four key variables: reactive aggression, anger coping, anger dysregulation, and worry dysregulation (Model 11). All were significant and equivalent (βs| =.13-33), still accounting for 43% of the variance in youth-rated irritability.

To summarize, these results were consistent with hypotheses: Per both informants, irritability was uniquely, consistently, and robustly associated with aggression—especially with *reactive aggression*—and with emotion regulation problems—especially with *anger, poor coping,* and *dysregulation*.

3.4. Robustness of Symptom Associations After Controlling for Covariates

Lastly, Table 4 presents the associations of ARI irritability with internalizing and externalizing problems, controlling for age, gender, ethnicity, and family income. All the correlations that were significant at the zero-order level (Table 2) remained significant or marginal after accounting for these demographic covariates (Table 4). Specifically, parent- and youth-rated irritability were robustly associated with inattention, hyperactivity-impulsivity, executive dysfunction, and aggression (*p*s < .02). Further, parent-rated irritability remained linked to peer problems (*p* < .001) and youth-rated irritability remained linked to anxiety and depression (*p*s < .001). The significant zero-order correlation between youth-rated irritability and suicidality was attenuated (*p* =.08) in the presence of covariates.

4. Discussion

This study examined the psychometric properties and correlates of the ARI as a measure of irritability in a clinically referred youth sample. Results revealed good internal consistency and a unidimensional factor structure for the ARI per both parent- and youth-report. These findings

were supported by CFA and IRT models showing the scale's reliability across the range of severity seen in this outpatient sample. Providing support for validity, ARI scores were correlated with internalizing, externalizing, and emotion regulation problems, and with irritability-related impairment; but they showed little to no associations with age, gender, race/ethnicity, family income, and learning problems. These findings support the use of the ARI with clinically referred youth.

At the same time, these major findings should be interpreted in light of some other notable results. First, the psychometric properties for youth-report ARI were slightly weaker than for parent-report. Similar findings have been obtained in previous studies of different irritability rating scales (e.g., Evans et al., 2019). This may warrant caution in interpreting youth-report results and extra effort to obtain parent-report. Second, parent and youth ARI scores were correlated at *r* =.32, suggesting informants might be rating somewhat different constructs or offering somewhat different perspectives. This correlation was not moderated by age or gender. It is striking that this coefficient was *identical* to meta-analytic estimates of parent-child correspondence for externalizing problems (*r* =.32), higher than for internalizing problems (*r* =.26; De Los Reyes et al., 2015). Third, there was some variability in the item-level results. "Angry most of time" (item 4) was rarely endorsed and "easily annoyed" (item 1) performed questionably, especially by youth-report, in CFA/IRT models. It seems plausible that the former occurs somewhat infrequently, whereas the latter occurs so frequently that it loses specificity for measuring irritability in clinically referred youth. These are intriguing questions for future research.

Tests of convergent and discriminant validity were largely consistent with hypotheses. Parent- and youth-rated irritability were significantly linked to inattention, hyperactivity-impulsivity, executive dysfunction, and aggression, as well as to general emotion regulation difficulties. The absence of an association with learning problems offers evidence for discriminant validity. By youth-report only, irritability was linked to / anxiety, depressive symptoms, and suicidality. These patterns of emotional and behavioral correlates between parent and youth reports are consistent with evidence that youths seem to be more accurate reporters of their own internalizing problems, whereas parents and

teachers may be more reliable for externalizing problems (De Los Reyes et al., 2015). Importantly, these validity correlations generally remained robust even after controlling for demographic covariates.

Finally, this is the first study to our knowledge to document hypothesized associations of irritability with specific facets of emotion regulation and aggressive behavior within a multi-informant framework. As expected, reactive aggression, anger problems, dysregulation, and poor coping showed the most pronounced associations with irritability. Notably, these results were obtained through a sequence of models that iteratively included (controlled for) and omitted (trimmed) other theoretically related variables in varying combinations. This process yielded a consistent and interesting pattern of results. Findings contribute uniquely to the literature by clarifying irritability's correlates while also identifying a few key facets of dysregulation and aggression that might warrant special attention in the treatment of irritable youths. Further research is needed to understand how these facets pertain to the etiology, phenomenology, and sequelae of youth irritability—and how they could be addressed clinically.

4.1. Limitations and Implications

This study has some limitations. First, this was a predominantly White, school-aged, and male sample. Although sociodemographic data were consistent with the clinic's location, further research is needed to advance the measurement and understanding of irritability among youth of different ages and with greater racial/ethnic diversity. Second, because data were collected on one occasion, results do not support directionality or causality. Third, although multi-informant data were present for many scales, we relied on youths for the CDI, RCMAS, and SBQ and on parents for the Conners. Comprehensive multi-informant assessment in each domain would be ideal in future research, especially in contexts where both parties are involved in treatment. A related limitation was the differences in sample constitution for each informant, meaning parent results and youth results are not perfectly comparable (e.g., parent-report for ages 3-18; youth-report for ages 6-18). Finally, additional health service data (e.g., clinical outcomes, service/billing codes, diagnoses) were largely unavailable and should be included in future research.

Despite these limitations, the present findings offer important implications for the measurement and understanding of youth irritability. This study fills a gap in the ARI's emerging evidence base, providing new evidence for its reliability and validity from an independent, multi-informant, clinically referred youth sample. Results support the use of the ARI in outpatient youth mental health care. The present sample's composition—including community-referred youths, representing a variety of ages and problems—enhances the generalizability of these results to other mental health settings. Insofar as irritability has been shown to be a nonspecific marker of and precursor to a diverse array of mental health concerns (Evans et al., 2017; Stringaris et al., 2018), it is important to have psychometrically sound measures for assessing irritability among clinically referred individuals across development. Our findings suggest that the ARI is useful for serving this purpose.

Author contributions

SCE: conceptualization, data curation, formal analysis, methodology, supervision, project administration, writing – original draft, reviewing, and editing; MRA: conceptualization, data curation, writing – original draft, reviewing, and editing; RLD: data curation, writing – original draft, reviewing, and editing; HS: writing – original draft, reviewing, and editing; SLH: conceptualization, writing – original draft, reviewing, and editing.

Acknowledgments

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.jad.2020.11.002.

References

Abel, M.R., Poquiz, J.L., Fite, P.J., Doyle, R.L., 2020. Reactive aggression and suicidal behaviors in children receiving outpatient psychological services: The moderating role of hyperactivity and inattention. Child Psychiatry & Human Development 51 (1), 2–12. https://doi.org/10.1007/s10578-019-00905-5.

Achenbach, T.M., Rescorla, L., 2001. Manual for the ASEBA school-age forms & profiles: An integrated system of multi-informant assessment. ASEBA, Burlington, VT.

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 (9), 941–949. https://doi.org/10.1111/jcpp.12039.

Ahmad, R., Mansoor, I., 2011. What I think and feel: Translation and adaptation of Revised Children's Manifest Anxiety Scale, (RCMAS-2) and its reliability assessment. The International Journal of Educational and Psychological Assessment.

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 (10), 1162–1171. https://doi.org/10.1111/jcpp.12233.

American Psychiatric Association. (2013). DSM-5 Online Assessment Measures. https://www.psychiatry.org/psychiatrists/practice/dsm/educational-resources/assessment-measures.

American Psychological Association, 2017. Ethical principles of psychologists and code of conduct https://www.apa.org. https://www.apa.org/ethics/code/index.

Ang, R.P., Lowe, P.A., Yusof, N., 2011. An examination of the RCMAS-2 scores across gender, ethnic background, and age in a large Asian school sample. Psychological Assessment 23 (4), 899–910. https://doi.org/10.1037/a0023891.

Bae, Y., 2012. Test review: Children's depression inventory 2 (CDI 2). Sage Publications Sage, CA: Los Angeles. CA.

Brotman, M.A., Kircanski, K., Leibenluft, E., 2017. Irritability in children and adolescents. Annual Review of Clinical Psychology 13 (1), 317–341. https://doi.org/10.1146/annurev-clinpsy-032816-044941.

Brotman, M.A., Schmajuk, M., Rich, B.A., Dickstein, D.P., Guyer, A.E., Costello, E.J., Egger, H.L., Angold, A., Pine, D.S., Leibenluft, E., 2006. Prevalence, clinical correlates, and longitudinal course of severe mood dysregulation in children. Biological Psychiatry 60 (9), 991–997. https://doi.org/10.1016/j.biopsych.2006.08.042.

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 (11), 1176–1183. https://doi.org/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 (4), 841–851. https://doi.org/10.1037/a0037898.

Campos, R.C., Holden, R.R., 2019. Portuguese version of the Suicidal Behaviors Questionnaire-Revised: Validation data and the establishment of a cut-score for screening purposes. European Journal of Psychological Assessment 35 (2), 190–195. https://doi.org/10.1027/1015-5759/a000385.

Conners, C.K., 2008. Conners 3rd edition: Manual. Multi-Health Systems Toronto, Ontario, Canada.

Cotton, C.R., Range, L.M., 1993. Suicidality, hopelessness, and attitudes toward life and death in children. Death Studies 17 (2), 185–191. https://doi.org/10.1080/07481189308252615.

Cotton, C.R., Peters, D.K., Range, L.M., 1995. Psychometric properties of the suicidal behaviors questionnaire. Death Studies 19 (4), 391–397. https://doi.org/10.1080/07481189508252740.

De Los Reyes, A., Augenstein, T.M., Wang, M., Thomas, S.A., Drabick, D.A.G., Burgers, D.E., Rabinowitz, J., 2015. The validity of the multi-informant approach to assessing child and adolescent mental health. Psychological Bulletin 141 (4), 858–900. https://doi.org/10.1037/a0038498.

DeSousa, D.A., Stringaris, A., Leibenluft, E., Koller, S.H., Manfro, G.G., Salum, G.A., 2013. Cross-cultural adaptation and preliminary psychometric properties of the Affective Reactivity Index in Brazilian Youth: Implications for DSM-5 measured irritability. Trends in Psychiatry and Psychotherapy 35 (3), 171–180.

Dodge, K.A., Coie, J.D., 1987. Social-information-processing factors in reactive and proactive aggression in children's peer groups. Journal of Personality and Social Psychology 53 (6), 1146.

Dougherty, L.R., Galano, M.M., Chad-Friedman, E., Olino, T.M., Bufferd, S.J., Klein, D.N., 2020. Using item response theory to compare irritability measures in early adolescent and childhood samples. Assessment, 1073191120936363.

Dougherty, L.R., Smith, V.C., Bufferd, S.J., Kessel, E., Carlson, G.A., Klein, D.N., 2015. Preschool irritability predicts child psychopathology, functional impairment, and service use at age nine. Journal of Child Psychology and Psychiatry 56 (9), 999–1007.

Evans, S.C., Blossom, J.B., Fite, P.J., 2020. Exploring longitudinal mechanisms of irritability in children: Implications for cognitive-behavioral intervention. Behavior Therapy 51 (2), 238–252. https://doi.org/10.1016/j.beth.2019.05.006.

Evans, S.C., Bonadio, F.T., Bearman, S.K., Ugueto, A.M., Chorpita, B.F., Weisz, J.R., 2019. Assessing the irritable and defiant dimensions of youth oppositional behavior using CBCL and YSR Items. Journal of Clinical Child & Adolescent Psychology 1–16.

Evans, S.C., Burke, J.D., Roberts, M.C., Fite, P.J., Lochman, J.E., Francisco, R., Reed, G. M., 2017. Irritability in child and adolescent psychopathology: An integrative review for ICD-11. Clinical Psychology Review.

Evans, S.C., Frazer, A.L., Blossom, J.B., Fite, P.J., 2018. Forms and functions of aggression in early childhood. Journal of Clinical Child & Adolescent Psychology 1–9.

Evans, S.C., Weisz, J.R., Carvalho, A.C., Garibaldi, P.M., Bearman, S.K., Chorpita, B.F., 2020. Effects of standard and modular psychotherapies in the treatment of youth with severe irritability. Journal of Consulting and Clinical Psychology 88 (3), 255–268. https://doi.org/10.1037/ccp0000456.

Ezpeleta, L., Penelo, E., de la Osa, N., Navarro, J.B., Trepat, E., 2020. How the Affective Reactivity Index (ARI) works for teachers as informants. Journal of Affective Disorders 261, 40–48. https://doi.org/10.1016/j.jad.2019.09.080.

Fite, P.J., Poquiz, J., Frazer, A.L., Reiter, N., 2017. Further evaluation of associations between reactive and proactive aggression and suicidal behavior in a treatment seeking sample of youth. Child Psychiatry & Human Development 48 (6), 903–910. https://doi.org/10.1007/s10578-017-0713-4.

Fite, P.J., Stoppelbein, L., Greening, L., 2009. Proactive and reactive aggression in a child psychiatric inpatient population. Journal of Clinical Child & Adolescent Psychology 38 (2), 199–205.

Folk, J.B., Zeman, J.L., Poon, J.A., Dallaire, D.H., 2014. A longitudinal examination of emotion regulation: Pathways to anxiety and depressive symptoms in urban minority youth. Child and Adolescent Mental Health 19 (4), 243–250. https://doi.org/10.1111/camh.12058.

Haller, S.P., Kircanski, K., Stringaris, A., Clayton, M., Bui, H., Agorsor, C., Cardenas, S.I., Towbin, K.E., Pine, D.S., Leibenluft, E., Brotman, M.A., 2020. The clinician affective reactivity index: Validity and reliability of a clinician-rated assessment of irritability. Behavior Therapy 51 (2), 283–293. https://doi.org/10.1016/j.beth.2019.10.005.

Humphreys, K.L., Schouboe, S.N.F., Kircanski, K., Leibenluft, E., Stringaris, A., Gotlib, I. H., 2019. Irritability, externalizing, and internalizing psychopathology in adolescence: Cross-sectional and longitudinal associations and moderation by sex. Journal of Clinical Child & Adolescent Psychology 48 (5), 781–789. https://doi.org/10.1080/15374416.2018.1460847.

Kalvin, C.B., Gladstone, T.R., Jordan, R., Rowley, S., Marsh, C.L., Ibrahim, K., Sukhodolsky, D.G., 2020. Assessing irritability in children with autism spectrum disorder using the affective reactivity index. Journal of Autism and Developmental Disorders 1–12.

Kenny, D.A., Kaniskan, B., McCoach, D.B., 2015. The performance of RMSEA in models with small degrees of freedom. Sociological Methods & Research 44 (3), 486–507.

Kline, R.B., 2016. Principles and practice of structural equation modeling, 4th ed. Guilford.

Kovacs, M.M.S., 2010. CDI 2: Children's Depression Inventory 2nd Edition (CDI 2), Technical Manual. MHS, Toronto, Canada.

Linden, W.J.van der, Hambleton, R.K. (Eds.), 1997. Handbook of Modern Item Response Theory. Springer-Verlag. https://doi.org/10.1007/978-1-4757-2691-6.

Little, T.D., 2013. Longitudinal structural equation modeling. Guilford.

Lowe, P.A., 2015. The revised Children's manifest anxiety scale–second edition short form: Examination of the psychometric properties of a brief measure of general anxiety in a sample of children and adolescents. Journal of Psychoeducational Assessment 33 (8), 719–730.

Mair, P., 2018. Modern psychometrics with R. Springer.

McNeish, D., 2018. Thanks coefficient alpha, we'll take it from here. Psychological Methods 23 (3), 412–433. https://doi.org/10.1037/met0000144.

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. https://doi.org/10.1016/j.adolescence.2018.02.008.

Mulraney, M.A., Melvin, G.A., Tonge, B.J., 2014. Psychometric properties of the Affective Reactivity Index in Australian adults and adolescents. Psychological Assessment 26 (1), 148.

Muthen, B.O., 2012. Mplus Discussion >>Binary CFA vs IRT. http://www.statmodel.com/discussion/messages/9/10401.html?1427253381.

Muthén, L.K., Muthén, B., 2019. Mplus. The Comprehensive Modelling Program for Applied Researchers: User's Guide 5.

National Council on Measurement, 2014. National Council on Measurement in Education Joint Committee on Standards for Educational and Psychological Testing. Standards for Educational and Psychological Testing.

Osman, A., Bagge, C.L., Gutierrez, P.M., Konick, L.C., Kopper, B.A., Barrios, F.X., 2001. The Suicidal Behaviors Questionnaire-Revised (SBQ-R): Validation with clinical and nonclinical samples. Assessment 8 (4), 443–454.

Pan, P.-Y., Yeh, C.-B., 2018. Irritability and maladaptation among children: the utility of Chinese versions of the affective reactivity index and aberrant behavior checklist-irritability subscale. Journal of Child and Adolescent Psychopharmacology 29 (3), 213–219. https://doi.org/10.1089/cap.2018.0070.

Reynolds, C.R., Richmond, B.O., 2008. Revised Children's Manifest Anxiety Scale: (RCMAS-2) Torrance. CA: Western Psychological Services.

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 (4), 726.

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 & Adolescent Psychiatry 48 (4), 404–412.

Stringaris, A., Goodman, R., 2009b. Three dimensions of oppositionality in youth. Journal of Child Psychology and Psychiatry 50 (3), 216–223. https://doi.org/10.1111/j.1469-7610.2008.01989.x.

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 (9), 1048–1054. https://doi.org/10.1176/appi.ajp.2009.08121849.

Stringaris, A., Goodman, R., Ferdinando, S., Razdan, V., Muhrer, E., Leibenluft, E., Brotman, M.A., 2012. The Affective Reactivity Index: A concise irritability scale for clinical and research settings. Journal of Child Psychology and Psychiatry 53 (11), 1109–1117. https://doi.org/10.1111/j.1469-7610.2012.02561.x.

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 (7), 721–739. https://doi.org/10.1111/jcpp.12823.

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 & Adolescent Psychiatry 55 (7), 556–570. https://doi.org/10.1016/j.jaac.2016.04.014.

Wakschlag, L.S., Briggs-Gowan, M.J., Choefi, S.W., Nichols, S.R., Kestler, J., Burns, J.L., Carter, A.S., Henry, D., 2014. Advancing a multidimensional, developmental spectrum approach to preschool disruptive behavior. Journal of the American Academy of Child and Adolescent Psychiatry 53 (1), 82–96. https://doi.org/10.1016/j.jaac.2013.10.011.

Zeman, J., Cassano, M., Suveg, C., Shipman, K., 2010. Initial validation of the children's worry management scale. Journal of Child and Family Studies 19 (4), 381–392.

Zeman, J., Shipman, K., Penza-Clyve, S., 2001. Development and initial validation of the children's sadness management scale. Journal of Nonverbal Behavior 25 (3), 187–205. https://doi.org/10.1023/A:1010623226626.

Zeman, J., Shipman, K., Suveg, C., 2002. Anger and sadness regulation: Predictions to internalizing and externalizing symptoms in children. Journal of Clinical Child and Adolescent Psychology 31 (3), 393–398.