Journal of Affective Disorders 332 (2023) 185–193

Contents lists available at ScienceDirect

Journal of Affective Disorders

journal homepage: www.elsevier.com/locate/jad

Research paper

Robust caregiver-youth discrepancies in irritability ratings on the affective reactivity index: An investigation of its origins

Ajitha Mallidi [a], Tatiana Meza-Cervera [a], Katharina Kircanski [a], Argyris Stringaris [b], Melissa A. Brotman [a], Daniel S. Pine [a], Ellen Leibenluft [a], Julia O. Linke [a,*]

[a] Emotion and Development Branch, National Institute of Mental Health, National Institutes of Health, Bethesda, MD, United States of America
[b] Divisions of Psychiatry and Psychology and Language Sciences, University College London, United Kingdom

ARTICLE INFO

Keywords:
Irritability
Informant discrepancy
Affective reactivity index
Measurement invariance
Retest-reliability
Bifactor model

ABSTRACT

Objective: The Affective Reactivity Index (ARI) is widely used to assess young people's irritability symptoms, but youth and caregivers often diverge in their assessments. Such informant discrepancy might be rooted in poor psychometric properties, the differential conceptualization of irritability across informants, or reflect socio-demographic and clinical characteristics. We use an out-of-sample replication approach and leverage longitudinal data, available for a subset of the participants, to test these hypotheses.

Method: Across two independent samples ($N_{Cohort\text{-}1}$ = 765, 8–21 years; $N_{Cohort\text{-}2}$ = 1910, 6–21 years), we investigate the reliability and measurement invariance of the ARI, examine sociodemographic and clinical predictors of discrepant reporting and probe the utility of a bifactor model for cross-informant integration.

Results: Despite good internal consistency and 6-week-retest-reliability of parent (Cohort-1: $\alpha = 0.92$, ICC = 0.85; Cohort-2: $\alpha = 0.93$) and youth forms (Cohort-1: $\alpha = 0.88$, ICC = 0.78; Cohort-2: $\alpha = 0.82$), we confirm substantial informant discrepancy in ARI ratings (3 points on a scale from 0 to 12), which is stable over six weeks (ICC = 0.53). Measurement invariance across informants was weak, indicating that parents and youth may interpret ARI items differently. Irritability severity and diagnostic status predicted informant-discrepancy, albeit in opposing directions: higher severity was linked to relative, higher irritability-ratings by youth (Cohort-1: $\beta = -0.06$, $p < .001$; Cohort-2: $\beta = -0.06$, $p < .001$), while diagnoses of Disruptive Mood Dysregulation Disorder (Cohort-1: $\beta = 0.44$, $p < .001$; Cohort-2: $\beta = 0.84$, $p < .001$) and Oppositional Defiant Disorder (Cohort-1: $\beta = 0.41$, $p < .001$; Cohort-2: $\beta = 0.42$, $p < .001$) predicted relative higher irritability-ratings by caregivers. In both datasets, a bifactor model parsing informant-specific from shared irritability-related variance fit the data well (CFI = 0.99, RMSEA = 0.05; N_2: CFI = 0.99; RMSEA = 0.04).

Conclusion: Parent and youth ARI reports and their discrepancy are reliable and reflect different interpretations of the scale items; hence they should not be averaged. This finding also suggests that irritability is not a unitary construct. Future work should investigate and model how different aspects of irritability might differ in their impact on the responses of specific informants.

1. Introduction

Irritability, defined as increased proneness to anger, is among the most common reasons for psychiatric consultation in youth (Brotman et al., 2017). The Affective Reactivity Index (ARI; Stringaris et al., 2012) completed by youth and caregivers is widely used to assess irritability (Dougherty et al., 2021). However, prior work indicates significant discrepancies between youth and caregiver ARI ratings (Evans et al., 2020a; Evans et al., 2020b; Stoddard et al., 2014), with estimates of agreement ranging between r ~ 0.23–0.76 (Evans et al., 2020a; Pan and Yeh, 2019; Stringaris et al., 2012; Zik et al., 2021). Informants contribute differentially to response variance in the ARI (Zik et al., 2021), suggesting that informants differ in their sensitivity to different aspects of irritability (e.g., phasic and tonic components, Vidal-Ribas et al., 2016). This implies that informant discrepancies should not be discarded as noise but instead be viewed as meaningful, for example, for the identification of subtypes with treatment implications (De Los Reyes, 2011). However, work investigating the validity and utility of informant

* Corresponding author.
E-mail address: dr.julia.linke@gmail.com (J.O. Linke).

https://doi.org/10.1016/j.jad.2023.03.091
Received 15 December 2022; Received in revised form 3 March 2023; Accepted 29 March 2023
Available online 6 April 2023

perspectives in irritability research needs to be preceded by a detailed description of informant discrepancies across different populations and an evaluation of different potential causes of these discrepancies that include, aside from differential conceptualization of irritability across informants, psychometric properties of the scale, and clinical and sociodemographic characteristics of the participants.

Discrepant reporting of irritability could be driven by poor psychometric properties of the ARI, which asks caregivers and youth to rate the child's irritability over the past six months. Previous research in modestly sized (N ~ 88–237) community- and clinically-referred youth samples renders this explanation unlikely, as it shows adequate internal consistency for parent and youth ARI reports (Evans et al., 2020a; Mulraney et al., 2014; Stringaris et al., 2012; Zik et al., 2021). Further, the two-week retest reliability of parent ratings was good (Pan and Yeh, 2019: $N = 62$, intraclass correlation coefficient [ICC] = 0.87; Tseng et al., 2017: $N = 99$, ICC = 0.90) and ranged from moderate to good for youth reports (Pan and Yeh, 2019: ICC = 0.60; Tseng et al., 2017: ICC = 0.88). However, youth ratings were less reliable than parent ratings, particularly in youth with very low or very high irritability levels (Evans et al., 2020a). A large meta-analysis showed a negative relationship between reliability and informant discrepancy (De Los Reyes et al., 2015). Here we seek to replicate previous findings of satisfactory internal consistency and test-retest reliability of the ARI in two large samples covering the spectrum of irritability to rule out poor psychometric properties as a source of informant discrepancy. In addition, we extend the current literature by determining the test-retest reliability of the ARI over the course of six weeks, a period approximating the length of a short psychosocial intervention.

A second possibility is that informant discrepancy originates from differential interpretations of ARI items or divergent conceptualization of irritability between caregivers and youth. This can be assessed with tests of measurement invariance, which quantify whether items of a multi-item instrument (i.e., the ARI) load similarly onto the assumed latent factor of interest (i.e., irritability) for different groups (i.e., parents and youth; (Dimitrov, 2010; Millsap, 2011). Measurement invariance of the ARI would suggest that informant discrepancy of ARI ratings reflects noise as it does, for example, for anxiety ratings in youth (Behrens et al., 2019); in that case, the common practice of averaging parent and youth reports is justified. On the other hand, measurement noninvariance of the ARI would suggest that the scale operates differently across informants, supporting the hypothesis of a multifaceted irritability phenotype. For example, tonic (i.e., irritable mood) and phasic (i.e., temper outbursts) irritability appear to have distinct genetic underpinnings (Moore et al., 2019), respond differentially to treatment (Towbin et al., 2020), and might not be equally accessible to all informants. In addition, it is essential to establish measurement invariance of ARI ratings within informants across sexes, ages, and diagnostic groups to guide clinicians and researchers seeking to integrate and compare ARI ratings within one informant.

Finally, informant discrepancies might be driven by specific sociodemographic or clinical variables. A prior study suggests that ARI informant discrepancy is more extensive in younger children and youth diagnosed with disruptive mood dysregulation disorder (DMDD; (Zik et al., 2021). In other psychopathologies, larger informant discrepancy has been associated with lower socioeconomic status (Robinson et al., 2019) and higher intelligence (Behrens et al., 2019). Further, male youth report fewer externalizing symptoms than their parents (Bajeux et al., 2018; Penney and Skilling, 2012; Robinson et al., 2019), while female youth report more internalizing symptoms than their parents. (Bajeux et al., 2018; Behrens et al., 2019; Penney and Skilling, 2012) Building on this literature, we examined the effects of age, sex, socioeconomic status, intelligence, and child diagnosis on ARI informant discrepancy.

In addition to understanding sources of discrepant reporting, it is crucial from both the clinical and research perspectives to investigate how ARI ratings can best be integrated across informants. For example, a recent meta-analysis noted a lack of convergence of neuroimaging findings in irritability (Lee et al., 2023). Roughly one-third of the work included in this study used the ARI; of these, approximately two-thirds averaged across informants (Crum et al., 2021; Kryza-Lacombe et al., 2021; Liuzzi et al., 2020; Pagliaccio et al., 2017; Stoddard et al., 2017; Tseng et al., 2019; Wiggins et al., 2016). Thus, the highly prevalent practice of averaging might contribute to the heterogeneous findings, particularly if parents and youth indeed conceptualize irritability differently. Some more clinically oriented studies rely on clinicians to integrate parent- and child-report of irritability (Kircanski et al., 2018a; Miller et al., 2018; Perepletchikova et al., 2017; Towbin et al., 2020; Waxmonsky et al., 2016). However, treatment research might also benefit from a formalized integration of multi-informant perspectives since some evidence suggests different trajectories in parent- vs. youth ratings of irritability during treatment (Evans et al., 2020b; Linke et al., 2020). To this end, advanced latent variable approaches, which can parse irritability-related variability from informant-specific variability, have been proposed (Zik et al., 2021). Specifically, recent work suggests the validity of a bifactor approach to weight input from youth and parents (Kircanski et al., 2018b). Thus, in addition to characterizing and identifying sources of informant discrepancy, we sought to investigate whether a bifactor model can be applied across samples to integrate parent and youth ARI ratings.

The goals of the present study were to (1) describe discrepancies between caregiver- and youth ARI reports, (2) examine potential sources of such discrepant reporting, and (3) explore a bifactor approach to integrating irritability ratings across informants. To this end, we conducted four sets of analyses in two large pediatric samples: community-based and clinically-referred. First, we quantified informant agreement and discrepancy to provide two complementary perspectives on informant effects. Second, we examined scale characteristics to rule out that discrepant reporting merely reflects poor internal consistency and test-retest reliability of the ARI. Further, discrepancies might originate from differential interpretations of ARI items or a diverging conceptualization of irritability between informants. To explore this possibility, we assessed measurement invariance between both informants (Dimitrov, 2010; Putnick and Bornstein, 2016). Third, we tested the possible effects of sociodemographic and clinical variables on informant discrepancy. Finally, we examined whether a bifactor model with a parent-specific, youth-specific, and shared latent factor might be suited to integrate ARI ratings across informants (Kircanski et al., 2018b).

2. Methods

2.1. Recruitment

This study leveraged data from two independent datasets. One sample was recruited at the National Institute of Mental Health (NIMH, $N = 765$) as part of ongoing efforts to understand the neurobiological mechanisms of pediatric anxiety, bipolar disorder (BD), and irritability, which is the hallmark feature of disruptive mood dysregulation disorder (DMDD, American Psychiatric Association, 2013) but is also highly comorbid with attention-deficit/hyperactivity disorder (ADHD, Eyre et al., 2019). Recruitment included several strategies, including direct advertisements to families directly and referrals from local healthcare providers. In this sample, participants with an anxiety disorder (ANX) diagnosis and some with DMDD were treatment-seeking. The second sample was recruited through the Child Mind Institute (CMI, $N = 1910$, Alexander et al., 2017) as part of a multicenter study. Parent-child dyads for this study were recruited by advertising diagnostic evaluations for children who may have psychiatric concerns or need school-based accommodations among community members, educators, local care providers, and the parents directly.

Inclusion criteria for the NIMH sample were the absence of a psychiatric disorder diagnosis (i.e., healthy volunteer [HV] status) or the presence of at least one of the following psychiatric diagnoses: ADHD,

ANX, BD, or DMDD. Exclusion criteria comprised the presence of a psychotic disorder, trauma-related disorders, pervasive developmental disorders, or full-scale intelligence quotient (IQ) < 70. To ensure comparability across cohorts, the same inclusion and exclusion criteria were applied to the CMI data set, except for the age cut-off (NIMH: 8–21 years; CMI: 6–21 years). Consent and assent were obtained, and procedures were approved by the review boards of the respective institutions.

2.2. Measures

Diagnoses were made by trained clinicians using the traditional (NIMH; Kaufman et al., 1997) or computerized (CMI; Kaufman et al., 2017) Kiddie Schedule for Affective Disorders and Schizophrenia. Across samples, youth irritability over the last six months was assessed by parents and youth using the ARI (Stringaris et al., 2012). This questionnaire comprises six items rated on a three-point Likert scale ranging from 0 (not true) to 2 (certainly true). The total score ranges from 0 to 12; a seventh item that assesses impairment does not contribute to the total score. In the NIMH sample, IQ was measured with the Wechsler Abbreviated Scale of Intelligence (WASI; Wechsler, 1999), while in the CMI sample, the Wechsler Intelligence Scale for Children (WISC; Wechsler, 2003) and the Adult Intelligence Scale (WAIS; Wechsler, 2008) were used. Impairment was assessed by trained clinicians using the Children's Global Assessment Scale (CGAS), which assigns each participant a value between 1 and 100 (Shaffer et al., 1983).

2.3. Informant agreement and discrepancy

We examined parent-youth agreement and discrepancy, as these metrics provide a complementary perspective on informant effects. We calculated the parent-youth agreement as ICCs (R: psych package). Informant discrepancy is reported as standardized mean difference score (SMDS; parent − youth z-score of the ARI) and the absolute value of SMDS (absSMDS). SMDS provides information on both the direction and magnitude of the effect, while the absSMDS characterizes the overall discrepancy. We focus on SMDS because, compared to other indices such as the raw difference scores (RDS; parent − youth total ARI score), or residual difference scores, the SMDS only correlates equally with both informants' ratings from which it originates (De Los Reyes and Kazdin, 2004). However, Table 1 also shows raw difference scores (RDS; parent − youth ARI score) and absolute raw difference scores (absRDS) to facilitate comparisons with other studies.

We show between-informant ICCs and their confidence intervals (CI) across samples, by age (<8 [CMI only]/ 8–12.4/ >12.4 years), sex, and within each diagnostic group, and explore age-by-sex interactions. Further, we used independent samples *t*-test to compare discrepancy measures between samples and diagnostic groups. Sex and age effects on informant discrepancy were determined separately for each sample using a two-factor analysis of variance. Informant discrepancies on the item level were tested using McNemar Bowker tests (Table S1).

2.4. Internal consistency

We examined the internal consistency of the ARI-parent and -youth in both samples by deriving Cronbach's alpha (R: psych package).

2.5. Test-retest reliability

Within the NIMH sample, 177 parent-youth-dyads completed the ARI twice within an average of 45.3 days. In this subset, we examined the test-retest reliability of parent ratings, youth ratings, and the informant discrepancy measured as SMDS and absSMDS using linear mixed-effects models (R: lme4 package). These models included within-subjects effects for participants and timepoint while controlling for the number of days between the two assessments. The ICC values derived

Table 1
Demographics and clinical characteristics of the two samples.

	NIMH Sample (*N* = 765)	CMI Sample (*N* = 1910)	Test Statistic	p-value	Effect Sizes[b]
Male/female	408/357	1186/724	$\chi^2_{(1)} = 16.42$	<0.001	V = 0.08
Sociodemographics	mean (SD)	mean (SD)			
Age	12.7 (2.74)	10.2 (2.85)	$t_{(1444.8)} = 21.07$	<0.001	d = 0.89
IQ	111.8 (13.63)	100.9 (15.26)	$t_{(1429.8)} = 17.67$	<0.001	d = 0.74
Annual Income	%	%			
< $90.000	26.2	36.4	$\chi^2_{(1)=}20.51$	<0.001	V = 0.10
$90.000–150.000	36.9	26.0	$\chi^2_{(1)=}23.30$	<0.001	V = 0.11
>$150.000	36.9	37.6	$\chi^2_{(1)=}0.01$	<0.94	V < 0.00
Race, Ethnicity	%	%			
African American	10.9	14.7	$\chi^2_{(1)=}6.28$	<0.05	V = 0.05
Caucasian	77.7	47.0	$\chi^2_{(1)=}130.72$	<0.001	V = 0.23
Asian	2.6	3.7	$\chi^2_{(1)=}1.79$	0.181	V = 0.04
Hispanic	8.3	6.4	$\chi^2_{(1)=}2.90$	0.088	V = 0.03
Diagnoses	%	%			
ADHD	37.8	60.4	$\chi^2_{(1)} = 112.17$	<0.001	V = 0.39
ANX	35.8	32.5	$\chi^2_{(1)} = 2.77$	0.096	V = 0.11
BD[a]	7.5	0.2		<0.001	
DMDD[a]	19.1	1.6		<0.001	
ODD[a]	7.6	14.0		<0.001	
None	39.4	12.3	$\chi^2_{(1)} = 250.50$	<0.001	V = 0.31
Medication	%	%			
ADHD-Stimulant	17.4	15.8	$\chi^2_{(1)} = 51.09$	<0.001	V = 0.16
ADHD-Nonstimulant	6.9	4.5	$\chi^2_{(1)} = 30.03$	<0.001	V = 0.12
Antidepressants	10.9	6.4	$\chi^2_{(1)} = 73.50$	<0.001	V = 0.19
Antipsychotics	8.0	1.5	$\chi^2_{(1)} = 78.73$	<0.001	V = 0.20
Mood stabilizer	4.7	1.4	$\chi^2_{(1)} = 7.53$	<0.001	V = 0.06
Anticonvulsive	4.4	0.1	$\chi^2_{(1)} = 131.98$	<0.001	V = 0.26
Anxiolytics	1.7	1.4	$\chi^2_{(1)} = 0.20$	0.65	V = 0.01
Symptom Measures	mean (SD)	mean (SD)			
ARI-P	3.9 (3.81)	3.0 (3.21)	W = 809,017	<0.001	d = 0.26
ARI-Y	3.3 (3.09)	3.5 (3.14)	W = 694,742	<0.05	d = −0.08
CGAS	59.7 (17.31)	66.6 (11.42)	$t_{(173.92)} = -4.95$	<0.001	d = −0.57
Discrepancy					
RDS	0.57 (3.45)	−0.54 (3.82)	$t_{(1550.1)} = 7.29$	<0.001	d = 0.30
absRDS	2.54 (2.40)	2.88 (2.56)	$t_{(1497.1)} = -3.29$	<0.01	d = −0.14
SMDS	0.00 (0.98)	0.00 (1.20)	$t_{(1711.5)} = 0.00$	0.999	d = 0.00
absSMDS	0.71 (0.67)	0.91 (0.79)	$t_{(1631.6)} = -6.47$	<0.001	d = −0.26

Abbreviations: absRDS, absolute raw difference score; absSMDS, absolute standardized mean difference score; ADHD, Attention-Deficit/Hyperactivity Disorder; ANX, anxiety disorder; ARI—P, Affective Reactivity Index completed by parent; ARI—Y, Affective Reactivity Index completed by youth; CGAS, Children's Global Assessment Scale (Shaffer et al., 1983); CMI, Child Mind

Institute; DMDD, Disruptive Mood Dysregulation Disorder; NIMH, National Institute of Mental Health; ODD, Oppositional Defiant Disorder; RDS, raw difference score; SD, standard deviation; SMDS, standardized mean difference score; d, Cohen's d; V, Cramer's V.

[a] Data has been compared using Fisher's exact test, which does not provide a test statistic.

[b] Cohen's d may be interpreted as large (>0.8), medium (0.5–0.8) or small (0.2–0.5) effect. Similarly, Cramer's V can be interpreted as large (>0.5), medium (0.3–0.5), or small (0.1–0.3) effect.

from these models indicate the ratio of participant-specific to the total variance.

2.6. Measurement invariance

We first assessed measurement invariance between parent and youth for the complete samples and subsets (i.e., boys, girls, youth <8 [CMI only]/ 8–12.4/ >12.4 years, and the combinations such as boys >12.4 years). Second, we assessed measurement invariance within each informant by testing for potential effects of sex, youth age (<8 [CMI only]/ 8–12.4/ >12.4 years), and diagnosis (yes/no).

We conducted four confirmatory factor analyses (CFA) with increasing stringency, testing for configural, measurement, structural, and residual invariance implemented in MPlus 8 (Muthén and Muthén, 2017). First, the comparability of the factor structure across informants/groups was determined (configural invariance). Then factor loadings (metric/weak invariance), item thresholds (scalar/strong invariance), and item residuals (residual/strict invariance) were set to equal across informants/groups. To account for the categorical nature of the ARI items, all CFAs used a mean- and variance-adjusted weighted least squares (WLSMV) estimator. Good model fit was established using two criteria: comparative fit index (CFI) > 0.95, and root mean square error of approximation (RMSEA) < 0.06. To compare models, we did not rely on the χ^2 difference test (DIFFTEST in MPlus) as it is overly sensitive in large samples such as ours. Instead, we consider measurement invariance to be present if changes in fit indices did not exceed the following two thresholds: $\Delta CFI \leq -0.002$ and $\Delta RMSEA \geq 0.007$ (Cheung and Rensvold, 2002; Hu and Bentler, 1999).

2.7. Sociodemographic and clinical variables

In both samples, we conducted multiple linear regression analyses to assess the effects of sociodemographic variables (sex, age [here as continuous variable], income), youth characteristics (IQ, diagnostic status), and irritability severity (mean of ARI-parent and ARI-youth total scores) on informant-discrepancy (i.e., SMDS, absSMDS). We applied a threshold of $p < .025$, thereby correcting for the two discrepancy measures (SMDS, absSMDS) that were predicted.

2.8. Bifactor model

CFA was conducted in MPlus 8 (Muthén and Muthén, 2017) on the ARI-youth and ARI-parent to extract a general factor from the first six items in both questionnaires. Parameters were freed for both questionnaires, and each factor was fixed to 1. To determine the model fit, we used the RMSEA, CFI, and standardized root mean square residual (SRMR). Cut-offs were used following Hu and Bentler (1999) i.e., RMSEA<0.06, SRMR<0.08, CFI ≥ 0.95. Standardized factor loadings are reported. To examine how age, sex, and youth diagnosis related to the latent factors, Spearman correlations, and independent sample *t*-tests were conducted. To test robustness, we fit the bifactor model obtained in the NIMH sample to the CMI sample and vice versa. We also determined the correlation between the latent factors obtained from both models across samples.

3. Results

3.1. Participants

Samples differed vastly in age, intelligence, and impairment, although the latter information was missing for a substantial proportion of participants and should be interpreted cautiously. CMI participants were younger, had lower intelligence scores, and were more impaired. The CMI sample had a higher income, was racially more diverse, included more participants with an annual household income below $90.000, higher proportions of ADHD and ODD cases, lower proportions of diagnosis-free, DMDD, and BD cases, and consumed fewer psychotropic medications. These latter effects were small. The CMI sample also included more boys, but this effect was negligible. Parent-rated irritability was lower in the CMI sample, while youth-rated irritability was comparable between samples (Table 1). While these profound between-sample differences complicate the explanation of discrepant findings, they inspire confidence in the robustness of converging cross-sample results.

3.2. Informant agreement

In both samples, the parent-youth agreement was significant but low to moderate in the NIMH (ICC = 0.50, $F_{(764,765)} = 3.0$, $p < .001$, CI[0.45, 0.54]) and low in the CMI sample (ICC = 0.27, $F_{(1909,1910)} = 1.7$, $p < .001$, CI[0.23, 0.30], Fig. 1). In the NIMH sample, the agreement between girls and their parents was moderate and higher than in boys, while in the CMI sample, confidence intervals of ICCs for girls and boys largely overlapped. Within each sample, ICC confidence intervals for younger (8–12.4 years) and older parent-youth dyads (>12.4 years) largely overlapped (Fig. 1A, Table S2), suggesting no age-related increase in agreement. We also found no evidence of an age-by-sex interaction (Fig. S1). In the presence of a DMDD or ODD diagnosis, ICCs were non-significant in both samples (all *p*-values > .06).

3.3. Informant-discrepancy

Across samples, parent and youth ratings diverged by approximately three points (absRDS in Table 1), with approximately 30 % of the parent-youth dyads disagreeing by four or more points (Fig. 1B). The absRDS was slightly larger in the CMI sample. Further, there were moderate differences between samples regarding the direction of the discrepancy, with higher parent relative to youth rating in the NIMH sample but the opposite pattern in the CMI sample (RDS in Table 1, Fig. 1B). AbsSMDS was higher in younger parent-youth dyads across samples (NIMH: $F_{(1,754)} = 6.17$, $p = .013$; CMI: $F_{(1,1906)} = 9.5$, $p = .002$). In both samples, there was neither a main effect of sex nor a sex-by-age interaction on SMDS and absSMDS (all *p*-values > .076). The diagnostic status of the child had a significant effect on informant discrepancy; in both samples, informant discrepancy was larger in youth diagnosed with a psychiatric disorder than HVs (all p-values < .01; Table S3). Further, in youth diagnosed with a psychiatric disorder, the discrepancy was driven by higher parent ratings, an effect that was strongest in parent-youth dyads with a child diagnosed with DMDD or ODD, while in HVs, elevated youth ratings contributed to the discrepancy.

3.4. Internal consistency

The ARI-parent and -youth showed good to excellent internal consistency in both samples. Parents showed excellent consistency (NIMH: α = 0.92 [0.92, 0.93], CMI: α = 0.90 [0.89, 0.90]) and youth showed good consistency (NIMH: α = 0.88 [0.86, 0.89], CMI: α = 0.82 [0.81, 0.83]).

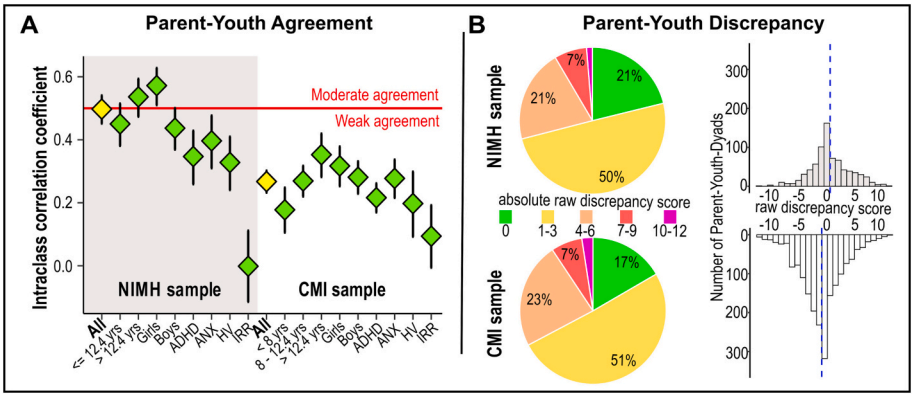

Fig. 1. Parent-youth agreement and discrepancy across samples. Panel A shows intraclass correlation coefficients with confidence intervals for the full and subsamples based on age, sex, and diagnostic status. Panel B shows the absolute raw discrepancy between parent and youth ratings as frequency (pie chart) and the raw discrepancy (bar chart). The dashed blue line signifies the mean discrepancy for each sample. Abbreviations: ADHD, Attention-Deficit/ Hyperactivity Disorder; ANX, anxiety disorder; CMI, Child Mind Institute; HV, healthy volunteers; ICC, intraclass correlation coefficient; IRR, diagnosis, where irritability is a core criterion (i.e., Disruptive Mood Dysregulation Disorder and Opposition Defiant Disorder); NIMH, National Institute of Mental Health. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

3.5. Test-retest reliability

Both parent (ICC = 0.85) and youth reports (ICC = 0.78) showed good test-retest reliability. The test-retest reliability of the informant-discrepancy was moderate (SMDS: ICC = 0.53; absSMDS: ICC = 0.48).

3.6. Measurement invariance

Across samples, the fit of the configural model assuming one irritability factor across informants was acceptable. Still, we could only establish weak/metric measurement invariance between parent and youth ARI ratings (Table 2). When factor loadings were freed for items two ("often loses temper", CMI), three ("stays angry for a long time", both datasets), five ("gets angry frequently", NIMH), and six ("loses temper easily", CMI), fit of the strong/scalar model improved with an increase of CFI = 0.001, suggesting that these items are interpreted differently by youth and parents. Across samples, exploratory analyses confirmed weak measurement invariance for subgroups of girls and boys (Table S4), different age groups (6–8, 8–12.4, >12.4 years, Table S5), young/ old girls (Table S6), and young/ old boys (Table S7). There was one exception: for girls older than 12.4 years in the NIMH subsample, strict measurement invariance was found.

Across samples, strict invariance of the parent report was established for age (Table S8), sex (Table S9), and diagnostic status (Table S10). Youth-report displayed only weak invariance for age (Table S8), strong to strict invariance for sex (Table S9), and strict invariance for diagnostic status (Table S10).

3.7. Sociodemographic and clinical variables

In the NIMH sample, greater SMDS was predicted by irritability severity ($\beta = -0.06$, $t = -4.32$, $p < .001$), diagnosis of DMDD ($\beta = 0.44$, $t = 8.11$, $p < .001$), ODD ($\beta = 0.41$, $t = 5.68$, $p < .001$), and BD ($\beta = 0.16$, $t = 2.32$, $p = .020$; $R^2_{adj} = 0.10$, $F_{(4,755)} = 20.05$, $p < .001$). In the CMI sample, variables predicting SMDS also included irritability severity ($\beta = -0.06$, $t = -4.88$, $p < .001$), diagnosis of DMDD ($\beta = 0.84$, $t = 7.85$, $p < .001$), and ODD ($\beta = 0.42$, $t = 10.45$, $p < .001$), as well as diagnosis of ADHD ($\beta = 0.11$, $t = 3.88$, $p < .001$), an anxiety disorder ($\beta = 0.09$, $t = 3.27$, $p = .001$), IQ ($\beta = 0.01$, $t = 4.31$, $p < .001$) and age ($\beta = 0.04$, $t = 4.43$, $p < .001$; $R^2_{adj} = 0.06$, $F_{(7,1902)} = 19.32$, $p < .001$). When participants 8 years and younger were excluded from the CMI dataset, which reduced the sample to $n = 1415$, irritability severity ($\beta = -0.03$, $t = -2.04$, $p = .041$), diagnosis of DMDD ($\beta = 0.75$, $t = 5.92$, $p < .001$), ODD ($\beta = 0.38$, $t = 7.75$, $p < .001$) and ADHD ($\beta = 0.09$, $t = 2.95$, $p = .003$), and IQ ($\beta = 0.01$, $t = 4.25$, $p < .001$; $R^2_{adj} = 0.07$, $F_{(5,1409)} = 24.22$, $p < .001$) remained significant predictors of SMDS. Greater absSMDS was predicted by greater irritability severity ($\beta = 0.10$, $t = 13.09$, $p < .001$) in the NIMH sample ($R^2_{adj} = 0.18$, $F_{(1,763)} = 171.5$, $p < .001$), and greater irritability severity ($\beta = 0.14$, $t = 21.72$, $p < .001$) and a diagnosis of ADHD ($\beta = 0.05$, $t = 3.02$, $p < .001$) in the CMI sample ($R^2_{adj} = 0.21$, $F_{(2,1907)} = 257.8$, $p < .001$). For full regression models, see Table S11.

Across samples, higher irritability severity predicted higher ARI ratings in youth relative to parents, while all other predictors indicated the opposite pattern (parent-ratings > youth-ratings). Effects of predictors that did not replicate across samples (i.e., diagnosis of BD, ADHD, ANX, IQ, and age) were relatively small, as signified by the beta-weights, compared to the effects of the replicated predictors (i.e., DMDD

Table 2
Measurement invariance for parent and youth informants across the NIMH and CMI sample.***

	Model Fit							Fit Difference		
Model	χ^2	df	CFI	TLI	RMSEA	Lower CI	Upper CI	$\Delta\chi^2$	ΔCFI	ΔRMSEA
NIMH Sample										
Configural	68.06***	18	1.00	1.00	0.06	0.05	0.08	–	–	–
Metric	96.34***	23	1.00	1.00	0.07	0.05	0.08	29.06***	−0.001	0.005
Scalar	210.86***	34	1.00	1.00	0.08	0.07	0.09	140.54***	−0.002	0.017
Residual	214.49***	34	1.00	1.00	0.09	0.08	0.11	29.73***	0.000	0.011
CMI Sample										
Configural	251.53***	18	1.00	0.99	0.08	0.07	0.09	–	–	–
Metric	257.30***	23	1.00	1.00	0.07	0.07	0.08	50.13***	0.000	−0.009
Scalar	1105.28***	34	0.98	0.98	0.13	0.12	0.14	1010.13***	−0.015	0.055
Residual	554.58***	28	0.99	0.99	0.10	0.09	0.11	490.72***	0.010	−0.029

Abbreviations: χ2, chi-square; Δ, change; CI, confidence interval; CMI, Child Mind Institute; CFI, comparative fit index; df, degree of freedom; NIMH, National Institute of Mental Health; RMSEA, root mean square error of approximation; TLI, Tucker Lewis Index.
*** $p < .001$.
** $p < .01$.
* $p < .05$.

or ODD diagnosis).

3.8. Bifactor model

Model fit for the NIMH sample (CFI = 0.99; RMSEA = 0.05; SRMR = 0.02) and the CMI sample (CFI = 0.99; RMSEA = 0.04; SRMR = 0.02; Fig. 2) was excellent. In the NIMH sample, all items loaded significantly on the shared and the respective informant-specific factors. In the CMI sample, youth-rated items one (easily annoyed by others), two (lose temper often), and six (lose temper easily) loaded strongly on the shared but not the youth-specific factor. However, the fit of the NIMH model to the CMI sample (CFI = 0.95; RMSEA = 0.10; SRMR = 0.12), and vice versa (CFI = 0.97; RMSEA = 0.10; SRMR = 0.12) was acceptable. Cross-correlations for the shared (NIMH: $r_s = 0.87$, $p < .001$; CMI: $r_s = 0.78$, $p < .001$) and parent-specific factors (NIMH: $r_s = 0.88$, $p < .001$; CMI: $r_s = 0.89$, $p < .001$) were substantial, but weaker for the youth-specific factors (NIMH: $r_s = 0.43$, $p < .001$; CMI: $r_s = 0.47$, $p < .001$).

Across samples, shared and parent-specific factor scores were lowest in HVs, and highest in individuals with DMDD or ODD (Fig. 3). The youth-specific factor did not differ between diagnostic groups. Across samples, youth age was negatively associated with the shared (NIMH: $r_S = -0.23$, $p < .001$; CMI: $r_S = -0.14$, $p < .001$) and youth-specific factor scores (NIMH: $r_S = -0.14$, $p < .001$; CMI: $r_S = -0.17$, $p < .001$, Fig. S2). There were no sex differences (all p-values $> .07$).

4. Discussion

Using two large pediatric samples, we replicated previous findings regarding informant discrepancy in ARI ratings, a widely used parent- and youth-report measure for irritability. First, in cross-sectional data, we confirmed significant but low parent-youth agreement and substantial parent-youth discrepancy for ARI ratings. Second, we demonstrated good internal consistency for each informant. In a subsample, where repeated measures were available over an average of six weeks, we showed good test-retest reliability for both informants. These findings argue against inadequate reliability as a potential cause of informant discrepancy. Interestingly, test-retest reliability for the informant discrepancy was also good, indicating that the discrepancy is relatively stable over this interval. Third, we found that irritability severity was a robust predictor of informant discrepancy, but the direction of the discrepancy varied by diagnosis, such that parent ratings were higher than youth ratings only in youth with DMDD and ODD. Fourth, measurement noninvariance across informants indicates that the ARI does not function similarly across parents and youth. Thus, ratings from the two informants should not be averaged but might be integrated at the level of variances. Indeed, a bifactor approach, which parses informant-specific from shared irritability-related variance, may be an option to integrate ARI ratings across informants.

Although recent evidence suggests that measurement invariance alone might be insufficient to substantiate the claim that the same construct is assessed across groups (Protzko, 2022), we interpret our findings of measurement invariance such that within informants, irritability is conceptualized similarly across youth's sex and diagnostic status. Thus, parent- and youth ratings might be used separately to compare these groups. However, while parents' conceptualization of irritability is independent of the child's age, youth's conceptualization of irritable behavior changes from childhood through adolescence. Thus, parent ratings only might be used for comparisons across age groups.

Consistent with previous studies examining the reliability of the ARI (Evans et al., 2020a; Mulraney et al., 2014; Pan and Yeh, 2019; Stringaris et al., 2012; Tseng et al., 2017; Zik et al., 2021), we found good internal consistency and six-week test-retest reliability suggesting that parent- and youth ratings might be used separately to evaluate treatment effects over that time period. In line with the literature on diverging ratings among informants (De Los Reyes et al., 2015), informant discrepancy was also moderately reliable. The combination of (a) stability within informants, (b) stability of informant discrepancy, and (c) measurement non-invariance across informants suggest that, while reliable, parent and youth ratings may reflect different components of the irritability phenotype. For example, phasic (i.e. temper outbursts) and tonic (i.e. grouchy mood) components of irritability might be more or less salient to differing informants. Indeed, a recent report supports this hypothesis linking youth irritability-ratings to the connectivity of default mode, limbic and temporal brain regions associated with internal states. At the same time, parent-report was related to somatomotor and prefrontal regions relevant to overt behavior (Linke et al., 2022).

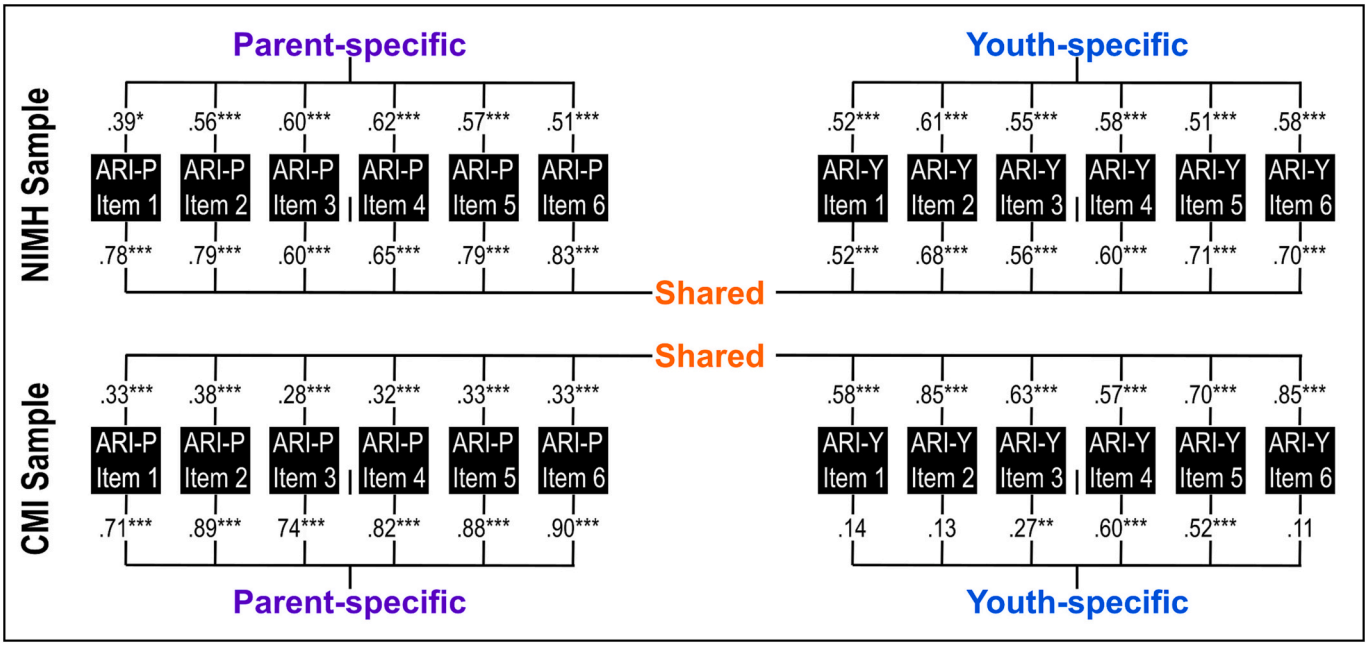

Fig. 2. Bifactor model of parent-youth irritability. Paths represent the loadings of each item on the parent-specific, youth-specific, and shared factor. Abbreviations: ARI—P, Affective Reactivity Index completed by parent; ARI—Y, Affective Reactivity Index completed by youth; CMI, Child Mind Institute; NIMH, National Institute of Mental Health, *** $p < .001$, ** $p < .01$, *$p < .05$.

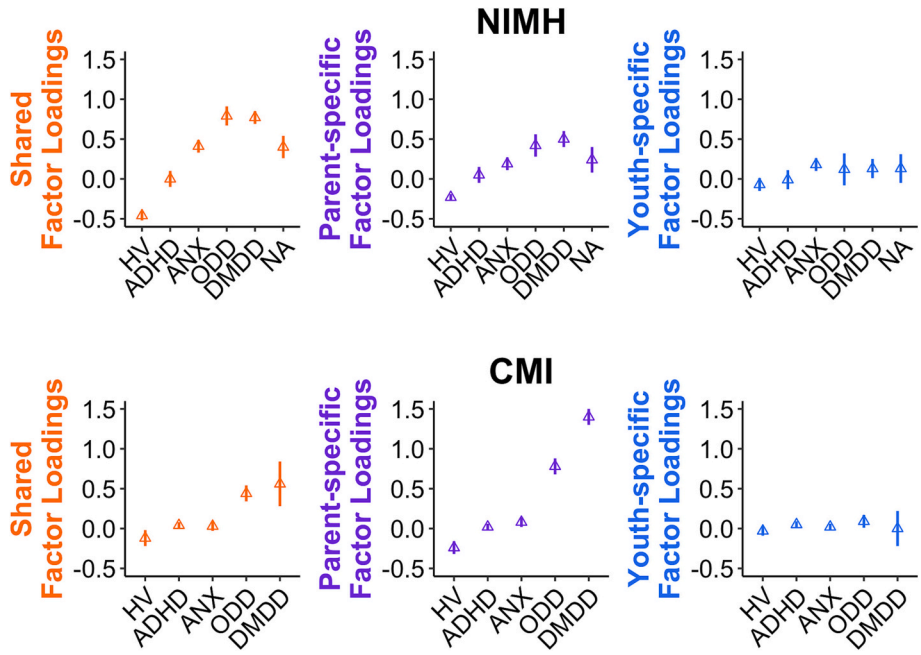

Fig. 3. Shared, parent-, and youth-specific factor loadings across diagnostic groups.
Abbreviations: ADHD, Attention-Deficit/Hyperactivity Disorder; ANX, anxiety disorder; BD, Bipolar Disorder; CMI, Child Mind Institute; DMDD, Disruptive Mood Dysregulation Disorder; NIMH, National Institute of Mental Health; ODD, Oppositional Defiant Disorder.

Further, irritability might occur predominantly in specific contexts (i.e., with peers, school, alone, …) that are accessible to one but not another informant. Thus, the divergence between informants should not be discarded as "noise". This has important implications for current and future work. First, it suggests that the informant is a considerable source of heterogeneity in irritability assessment, which might contribute to the current lack of convergence of, for example, imaging findings in irritability (Lee et al., 2023). Second, findings show that future research should focus on measurement validity and examine whether these somewhat complementary perspectives can be leveraged to identify clinically meaningful subtypes of irritability, with the ultimate goal of refining the current conceptualization of the irritability phenotype.

Ecological momentary assessment (EMA) obtained from parents and youth is a potential avenue to gain insights into determinants of parental and youth assessment of irritability (Naim et al., 2021b). EMA coupled with remote physiological measures such as actigraphy or heart rate variability (Naim et al., 2021a) might address, for example, whether child ratings are driven by internal emotional experience and arousal while parent ratings are driven by observed behavior. EMA studies could also help identify contextual factors that drive the discrepant reporting of irritability. Further, examination of informant discrepancies with a person-centered statistical approach might identify subgroups of youth that are of practical relevance. In the context of internalizing and externalizing disorders (Curhan et al., 2020), such an approach has been validated, demonstrating that such subgroups differ in the treatment and social services they receive and thus might eventually aid in clinical decision-making. Such work should also consider a wider variety of key informants, such as teachers and clinicians.

While our data do not support averaging, cross-informant information might be integrated at the level of variances. We tested this using a bifactor model, which parsed parent- and youth-specific from shared irritability-related variance. In both samples, the model fit the data well, but cross-validation analyses showed the robustness of the shared and parent-specific factors only. Shared factor scores were highest in younger children and those with DMDD or ODD, suggesting sensitivity of the shared factor to syndromal context and developmental processes. The parent-specific factor scores also varied by diagnosis. Future work examining the validity of the shared and informant-specific latent factors concerning informant-independent criteria (e.g., neurobiology, natural courses, or treatment responses) and comparisons of the bifactor approach to other recently proposed strategies such as the calculation of a trait score across informants (Makol et al., 2020) is needed.

Across samples, irritability severity was identified as the most robust, albeit weak, predictor of informant-discrepancy metrics, with higher severity predicting higher absolute informant-discrepancy, specifically with higher ratings in youth than parents. Two other predictors of informant discrepancy replicated across samples: the presence of DMDD or ODD (diagnoses characterized by severe irritability) were related to an opposite pattern than irritability severity i.e., elevated parent- relative to youth ratings. Thus, the higher prevalence of these two diagnostic groups in the NIMH (26.7 %) compared to the CMI (15.6 %) sample in the context of moderate but comparable irritability severity might explain why, when comparing discrepant reporting across samples, NIMH-parents, but CMI-youth, report higher irritability relative to the other informant. While these findings are consistent with a previous report of increased informant discrepancy in the context of DMDD (Zik et al., 2021), further work examining informant discrepancy and its direction across diagnoses and irritability severity is warranted.

There were also a few sample-specific findings. While the agreement was weak across samples, it was generally higher in the NIMH sample, a difference that was not explained by age, sex, or diagnostic status. Further, variables such as age and IQ, which differed largely between cohorts, did not robustly explain variance in informant discrepancy. Thus, future studies examining moderators of informant agreement/discrepancy are warranted. Such work could include the study of parental psychopathology (Pugh and Farrell, 2011) and psychotropic medication.

The current manuscript presents a first step in advancing our understanding of informant discrepancies in irritability ratings in two large cohorts. However, results must be viewed in light of some limitations. While we focus on findings that replicate across samples despite between-sample differences in sociodemographic and clinical measures, it is nonetheless puzzling that obvious differences in sociodemographic variables (i.e., age, IQ) cannot explain between-sample differences in the level of informant agreement. Thus, future work examining determinants of measurement validity is needed (De Los Reyes et al.,

2019). We also note that using difference scores (Laird, 2020) and measurement invariance to characterize informant discrepancy (De Los Reyes et al., 2022) is controversial in the literature. However, these critiques and our conclusions converge in suggesting that future work should examine the valid perspectives of different informants and their impacts on clinical outcomes. A multi-method assessment of irritability should complement this work to understand the structure of the phenotype beyond specific assessment methods. Future work should examine the generalizability of the present findings by including additional diagnostic categories, such as major depression. Further, the annual household income in both samples was upwards skewed, and so future work is needed to test whether findings generalize low-income households. An extension to non-Western cultural backgrounds is also warranted.

In sum, our study confirms parent-youth discrepancy in ARI scores from middle childhood to adolescence. This effect might be accentuated in youth with severe irritability and appears to be driven by informant-specific interpretation of ARI items. Our data suggest that averaging parent and youth ARI ratings is inadvisable, but latent variable approaches such as bifactor models might facilitate integration across informants. While further work on the validity and utility of such latent variable approaches is needed, parent and youth ARI ratings might be used separately to study irritability across sexes and diagnostic groups and to evaluate treatment response. Parent ratings can also be used to compare different age groups. Finally, the good to excellent reliability of parent ratings, youth ratings, and informant discrepancy coupled with weak measurement invariance suggest that parents and youth are both reliable informants that report on different aspects of irritability. Future research should characterize these different perspectives further and explore their associations with the phenomenology of irritability and its neurobiological underpinnings.

CRediT authorship contribution statement

Conceptualization: AM, JOL; Assessment and preprocessing: AM, TMC; Supervision: JOL, EL; Formal analysis: AM, TMC, JOL; Writing – original draft preparation: AM, JOL, TMC, EL; Writing – review and editing: AM, JOL, TMC, EL, AS, DSP, MAB, KK; Visualization: AM, TMC, JOL; Funding acquisition: EL, DSP.

Funding sources

The authors' research is supported by the NIMH Intramural Research Program (ZIAMH002786, ZIAMH002778, ZIAMH002782).

Conflict of interest

None.

Data availability

The NIMH data that support the findings of this study are available upon reasonable request from the corresponding author (JOL). The CMI data are publicly available.

Acknowledgements

No additional acknowledgements are noted.

Appendix A. Supplementary information

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jad.2023.03.091.

References

Alexander, L.M., Escalera, J., Ai, L., Andreotti, C., Febre, K., Mangone, A., Vega-Potler, N., Langer, N., Alexander, A., Kovacs, M., Litke, S., O'Hagan, B., Andersen, J., Bronstein, B., Bui, A., Bushey, M., Butler, H., Castagna, V., Camacho, N., Milham, M. P., 2017. An open resource for transdiagnostic research in pediatric mental health and learning disorders. Sci. Data 4 (1), 170181. https://doi.org/10.1038/sdata.2017.181.

American Psychiatric Association, 2013. Diagnostic and Statistical Manual of Mental Disorders, (Fifth Edition). American Psychiatric Association. https://doi.org/10.1176/appi.books.9780890425596.

Bajeux, E., Klemanski, D.H., Husky, M., Leray, E., Chan Chee, C., Shojaei, T., Fermanian, C., Kovess-Masfety, V., 2018. Factors associated with parent-child discrepancies in reports of mental health disorders in young children. Child Psychiatry Hum. Dev. 49 (6), 1003–1010. https://doi.org/10.1007/s10578-018-0815-7.

Behrens, B., Swetlitz, C., Pine, D.S., Pagliaccio, D., 2019. The screen for child anxiety related emotional disorders (SCARED): informant discrepancy, measurement invariance, and test-retest reliability. Child Psychiatry Hum. Dev. 50 (3), 473–482. https://doi.org/10.1007/s10578-018-0854-0.

Brotman, M.A., Kircanski, K., Stringaris, A., Pine, D.S., Leibenluft, E., 2017. Irritability in youths: a translational model. Am. J. Psychiatr. 174 (6), 520–532. https://doi.org/10.1176/appi.ajp.2016.16070839.

Cheung, G.W., Rensvold, R.B., 2002. Evaluating goodness-of-fit indexes for testing measurement invariance. Struct. Equ. Model. Multidiscip. J. 9 (2), 233–255. https://doi.org/10.1207/S15328007SEM0902_5.

Crum, K.I., Hwang, S., Blair, K.S., Aloi, J.M., Meffert, H., White, S.F., Tyler, P.M., Leibenluft, E., Pope, K., Blair, R.J.R., 2021. Interaction of irritability and anxiety on emotional responding and emotion regulation: a functional MRI study. Psychol. Med. 51 (16), 2778–2788. https://doi.org/10.1017/S0033291720001397.

Curhan, A.L., Rabinowitz, J.A., Pas, E.T., Bradshaw, C.P., 2020. Informant discrepancies in internalizing and externalizing symptoms in an at-risk sample: the role of parenting and school engagement. J. Youth Adolesc. 49 (1), 311–322. https://doi.org/10.1007/s10964-019-01107-x.

De Los Reyes, A., 2011. Introduction to the special section: more than measurement error: discovering meaning behind informant discrepancies in clinical assessments of children and adolescents. J. Clin. Child Adolesc. Psychol. 40 (1), 1–9. https://doi.org/10.1080/15374416.2011.533405.

De Los Reyes, A., Kazdin, A.E., 2004. Measuring informant discrepancies in clinical child research. Psychol. Assess. 16 (3), 330–334. https://doi.org/10.1037/1040-3590.16.3.330.

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. Psychol. Bull. 141 (4), 858–900. https://doi.org/10.1037/a0038498.

De Los Reyes, A., Lerner, M.D., Keeley, L.M., Weber, R.J., Drabick, D.A.G., Rabinowitz, J., Goodman, K.L., 2019. Improving interpretability of subjective assessments about psychological phenomena: a review and cross-cultural meta-analysis. Rev. Gen. Psychol. 23 (3), 293–319. https://doi.org/10.1177/1089268019837645.

De Los Reyes, A., Tyrell, F.A., Watts, A.L., Asmundson, G.J.G., 2022. Conceptual, methodological, and measurement factors that disqualify use of measurement invariance techniques to detect informant discrepancies in youth mental health assessments. Front. Psychol. 13, 931296 https://doi.org/10.3389/fpsyg.2022.931296.

Dimitrov, D.M., 2010. Testing for factorial invariance in the context of construct validation. Meas. Eval. Couns. Dev. 43 (2), 121–149. https://doi.org/10.1177/0748175610373459.

Dougherty, L.R., Galano, M.M., Chad-Friedman, E., Olino, T.M., Bufferd, S.J., Klein, D.N., 2021. Using item response theory to compare irritability measures in early adolescent and childhood samples. Assessment 28 (3), 918–927. https://doi.org/10.1177/1073191120936363.

Evans, S.C., Abel, M.R., Doyle, R.L., Skov, H., Harmon, S.L., 2020. Measurement and correlates of irritability in clinically referred youth: further examination of the affective reactivity index. J. Affect. Disord. https://doi.org/10.1016/j.jad.2020.11.002. S0165032720329189.

Evans, S.C., Weisz, J.R., Carvalho, A.C., Garibaldi, P.M., Bearman, S.K., Chorpita, B.F., Research Network on Youth Mental Health, 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.

Eyre, O., Riglin, L., Leibenluft, E., Stringaris, A., Collishaw, S., Thapar, A., 2019. Irritability in ADHD: association with later depression symptoms. Eur. Child Adolesc. Psychiatry 28 (10), 1375–1384. https://doi.org/10.1007/s00787-019-01303-x.

Hu, L., Bentler, P.M., 1999. Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. Struct. Equ. Model. Multidiscip. J. 6 (1), 1–55. https://doi.org/10.1080/10705519909540118.

Kaufman, J., Birmaher, B., Brent, D., Rao, U., Flynn, C., Moreci, P., Williamson, D., Ryan, N., 1997. Schedule for affective disorders and schizophrenia for school-age children-present and lifetime version (K-SADS-PL): initial reliability and validity data. J. Am. Acad. Child Adolesc. Psychiatry 36 (7), 980–988. https://doi.org/10.1097/00004583-199707000-00021.

Kaufman, J., Townsend, L.D., Kobak, K., 2017. The computerized kiddie schedule for affective disorders and schizophrenia (KSADS): development and administration guidelines. J. Am. Acad. Child Adolesc. Psychiatry 56 (10), S357. https://doi.org/10.1016/j.jaac.2017.07.770.

Kircanski, K., Clayton, M.E., Leibenluft, E., Brotman, M.A., 2018. Psychosocial treatment of irritability in youth. Curr. Treat. Opt. Psychiatry 5 (1), 129–140. https://doi.org/10.1007/s40501-018-0141-5.

Kircanski, K., White, L.K., Tseng, W.-L., Wiggins, J.L., Frank, H.R., Sequeira, S., Zhang, S., Abend, R., Towbin, K.E., Stringaris, A., Pine, D.S., Leibenluft, E., Brotman, M.A., 2018. A latent variable approach to differentiating neural mechanisms of irritability and anxiety in youth. JAMA Psychiatry 75 (6), 631. https://doi.org/10.1001/jamapsychiatry.2018.0468.

Kryza-Lacombe, M., Hernandez, B., Owen, C., Reynolds, R.C., Wakschlag, L.S., Dougherty, L.R., Wiggins, J.L., 2021. Neural mechanisms of reward processing in adolescent irritability. Dev. Psychobiol. 63 (5), 1241–1254. https://doi.org/10.1002/dev.22090.

Laird, R.D., 2020. Analytical challenges of testing hypotheses of agreement and discrepancy: comment on Campione-Barr, Lindell, and Giron (2020). Dev. Psychol. 56 (5), 970–977. https://doi.org/10.1037/dev0000763.

Lee, K.S., Hagan, C.N., Hughes, M., Cotter, G., McAdam Freud, E., Kircanski, K., Leibenluft, E., Brotman, M.A., Tseng, W.-L., 2023. Systematic review and meta-analysis: task-based fMRI studies in youths with irritability. J. Am. Acad. Child Adolesc. Psychiatry 62 (2), 208–229. https://doi.org/10.1016/j.jaac.2022.05.014.

Linke, J.O., Kircanski, K., Brooks, J., Perhamus, G., Gold, A.L., Brotman, M.A., 2020. Exposure-based cognitive-behavioral therapy for disruptive mood dysregulation disorder: an evidence-based case study. Behav. Ther. 51 (2), 320–333. https://doi.org/10.1016/j.beth.2019.05.007.

Linke, J.O., Haller, S.P., Xu, E.P., Nguyen, L.T., Chue, A.E., Botz-Zapp, C., Revzina, O., Perlstein, S., Ross, A.J., Tseng, W.-L., Shaw, P., Brotman, M.A., Pine, D.S., Gotts, S.J., Leibenluft, E., Kircanski, K., 2022. Persistent frustration-induced reconfigurations of brain networks predict individual differences in irritability. J. Am. Acad. Child Adolesc. Psychiatry. https://doi.org/10.1016/j.jaac.2022.11.009. S0890856722019827.

Liuzzi, M.T., Kryza-Lacombe, M., Christian, I.R., Palumbo, D.E., Amir, N., Wiggins, J.L., 2020. Neural and behavioral correlates of inhibitory control in youths with varying levels of irritability. J. Affect. Disord. 273, 567–575. https://doi.org/10.1016/j.jad.2020.04.049.

Makol, B.A., Youngstrom, E.A., Racz, S.J., Qasmieh, N., Glenn, L., De Los Reyes, A., 2020. Integrating multiple informants' reports: how conceptual and measurement models may address long-standing problems in clinical decision-making. Clin. Psychol. Sci. 8 (6), 953–970. https://doi.org/10.1177/2167702620924439.

Miller, L., Hlastala, S.A., Mufson, L., Leibenluft, E., Yenokyan, G., Riddle, M., 2018. Interpersonal psychotherapy for mood and behavior dysregulation: pilot randomized trial. Depress. Anxiety 35 (6), 574–582. https://doi.org/10.1002/da.22761.

Millsap, R.E., 2011. Statistical Approaches to Measurement Invariance, 1st ed. Routhledge.

Moore, A.A., Lapato, D.M., Brotman, M.A., Leibenluft, E., Aggen, S.H., Hettema, J.M., York, T.P., Silberg, J.L., Roberson-Nay, R., 2019. Heritability, stability, and prevalence of tonic and phasic irritability as indicators of disruptive mood dysregulation disorder. J.Child Psychol.Psychiatry Allied Discip. 60 (9), 1032–1041. https://doi.org/10.1111/jcpp.13062.

Mulraney, M.A., Melvin, G.A., Tonge, B.J., 2014. Psychometric properties of the affective reactivity index in australian adults and adolescents. Psychol. Assess. 26 (1), 148–155. https://doi.org/10.1037/a0034891.

Muthén, L.K., Muthén, B.O., 2017. Mplus User's Guide, (8th ed.). Muthén & Muthén.

Naim, R., Goodwin, M.S., Dombek, K., Revzina, O., Agorsor, C., Lee, K., Zapp, C., Freitag, G.F., Haller, S.P., Cardinale, E., Jangraw, D., Brotman, M.A., 2021. Cardiovascular reactivity as a measure of irritability in a transdiagnostic sample of youth: preliminary associations. Int. J. Methods Psychiatr. Res. 30 (4), e1890 https://doi.org/10.1002/mpr.1890.

Naim, R., Smith, A., Chue, A., Grassie, H., Linke, J., Dombek, K., Shaughnessy, S., McNeil, C., Cardinale, E., Agorsor, C., Cardenas, S., Brooks, J., Subar, A.R., Jones, E. L., Do, Q.B., Pine, D.S., Leibenluft, E., Brotman, M.A., Kircanski, K., 2021. Using ecological momentary assessment to enhance irritability phenotyping in a transdiagnostic sample of youth. Dev. Psychopathol. 33 (5), 1734–1746. https://doi.org/10.1017/S0954579421000717.

Pagliaccio, D., Wiggins, J.L., Adleman, N.E., Curhan, A., Zhang, S., Towbin, K.E., Brotman, M.A., Pine, D.S., Leibenluft, E., 2017. Behavioral and neural sustained attention deficits in disruptive mood dysregulation disorder and attention-Deficit/Hyperactivity disorder. J. Am. Acad. Child Adolesc. Psychiatry 56 (5), 426–435. https://doi.org/10.1016/j.jaac.2017.02.008.

Pan, P.-Y., Yeh, C.-B., 2019. Irritability and maladaptation among children: the utility of chinese versions of the affective reactivity index and aberrant behavior checklist-irritability subscale. J. Child Adolesc. Psychopharmacol. 29 (3), 213–219. https://doi.org/10.1089/cap.2018.0070.

Penney, S.R., Skilling, T.A., 2012. Moderators of informant agreement in the assessment of adolescent psychopathology: extension to a forensic sample. Psychol. Assess. 24 (2), 386–401. https://doi.org/10.1037/a0025693.

Perepletchikova, F., Nathanson, D., Axelrod, S.R., Merrill, C., Walker, A., Grossman, M., Rebeta, J., Scahill, L., Kaufman, J., Flye, B., Mauer, E., Walkup, J., 2017. Randomized clinical trial of dialectical behavior therapy for preadolescent children with disruptive mood dysregulation disorder: feasibility and outcomes. J. Am. Acad. Child Adolesc. Psychiatry 56 (10), 832–840. https://doi.org/10.1016/j.jaac.2017.07.789.

Protzko, J., 2022. Invariance: What Does Measurement Invariance Allow us to Claim? PsyArXiv. https://doi.org/10.31234/osf.io/r8yka.

Pugh, K.L., Farrell, A.D., 2011. In: The Impact of Maternal Depressive Symptoms on Adolescents' Aggression: Role of Parenting and Family Mediators, 21, pp. 589–602.

Putnick, D.L., Bornstein, M.H., 2016. Measurement invariance conventions and reporting: the state of the art and future directions for psychological research. Dev. Rev. 41, 71–90. https://doi.org/10.1016/j.dr.2016.06.004.

Robinson, M., Doherty, D.A., Cannon, J., Hickey, M., Rosenthal, S.L., Marino, J.L., Skinner, S.R., 2019. Comparing adolescent and parent reports of externalizing problems: a longitudinal population-based study. Br. J. Dev. Psychol. 37 (2), 247–268. https://doi.org/10.1111/bjdp.12270.

Shaffer, D., Gould, M.S., Brasic, J., Ambrosini, P., Fisher, P., Bird, H., Aluwahlia, S., 1983. A children's global assessment scale (CGAS). Arch. Gen. Psychiatry 40 (11), 1228–1231. https://doi.org/10.1001/archpsyc.1983.01790100074010.

Stoddard, J., Stringaris, A., Brotman, M.A., Montville, D., Pine, D.S., Leibenluft, E., 2014. Irritability in child and adolescent anxiety disorderS. Depress. Anxiety 31 (7), 566–573. https://doi.org/10.1002/da.22151.

Stoddard, J., Tseng, W.-L., Kim, P., Chen, G., Yi, J., Donahue, L., Brotman, M.A., Towbin, K.E., Pine, D.S., Leibenluft, E., 2017. Association of Irritability and Anxiety with the neural mechanisms of implicit face emotion processing in youths with psychopathology. JAMA Psychiatry 74 (1), 95. https://doi.org/10.1001/jamapsychiatry.2016.3282.

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: the affective reactivity index. J. Child Psychol. Psychiatry 53 (11), 1109–1117. https://doi.org/10.1111/j.1469-7610.2012.02561.x.

Towbin, K., Vidal-Ribas, P., Brotman, M.A., Pickles, A., Miller, K.V., Kaiser, A., Vitale, A. D., Engel, C., Overman, G.P., Davis, M., Lee, B., McNeil, C., Wheeler, W., Yokum, C. H., Haring, C.T., Roule, A., Wambach, C.G., Sharif-Askary, B., Pine, D.S., Stringaris, A., 2020. A double-blind randomized placebo-controlled trial of citalopram adjunctive to stimulant medication in youth with chronic severe irritability. J. Am. Acad. Child Adolesc. Psychiatry 59 (3), 350–361. https://doi.org/10.1016/j.jaac.2019.05.015.

Tseng, W.-L., Moroney, E., Machlin, L., Roberson-Nay, R., Hettema, J.M., Carney, D., Stoddard, J., Towbin, K.A., Pine, D.S., Leibenluft, E., Brotman, M.A., 2017. Test-retest reliability and validity of a frustration paradigm and irritability measures. J. Affect. Disord. 212, 38–45. https://doi.org/10.1016/j.jad.2017.01.024.

Tseng, W.-L., Deveney, C.M., Stoddard, J., Kircanski, K., Frackman, A.E., Yi, J.Y., Hsu, D., Moroney, E., Machlin, L., Donahue, L., Roule, A., Perhamus, G., Reynolds, R.C., Roberson-Nay, R., Hettema, J.M., Towbin, K.E., Stringaris, A., Pine, D.S., Brotman, M.A., Leibenluft, E., 2019. Brain mechanisms of attention orienting following frustration: associations with irritability and age in youths. Am. J. Psychiatr. 176 (1), 67–76. https://doi.org/10.1176/appi.ajp.2018.18040491.

Vidal-Ribas, P., Brotman, M.A., Valdivieso, I., Leibenluft, E., Stringaris, A., 2016. The status of irritability in psychiatry: a conceptual and quantitative review. J. Am. Acad. Child Adolesc. Psychiatry 55 (7), 556–570. https://doi.org/10.1016/j.jaac.2016.04.014.

Waxmonsky, J.G., Waschbusch, D.A., Belin, P., Li, T., Babocsai, L., Humphery, H., Pariseau, M.E., Babinski, D.E., Hoffman, M.T., Haak, J.L., Mazzant, J.R., Fabiano, G. A., Pettit, J.W., Fallahazad, N., Pelham, W.E., 2016. A randomized clinical trial of an integrative group therapy for children with severe mood dysregulation. J. Am. Acad. Child Adolesc. Psychiatry 55 (3), 196–207. https://doi.org/10.1016/j.jaac.2015.12.011.

Wechsler, D., 1999. Wechsler Abbreviated Scale of Intelligence. Psychological Corporation.

Wechsler, D., 2003. Wechsler Intelligence Scale for Children—Fourth Edition (WICS-IV). The Psychological Cooperation.

Wechsler, D., 2008. Wechsler Adult Intelligence Scale—Fourth Edition (WAIS-IV). John Wiley & Sons Inc.

Wiggins, J.L., Brotman, M.A., Adleman, N.E., Kim, P., Oakes, A.H., Reynolds, R.C., Chen, G., Pine, D.S., Leibenluft, E., 2016. Neural correlates of irritability in disruptive mood dysregulation and bipolar disorders. Am. J. Psychiatry 173 (7), 722–730. https://doi.org/10.1176/appi.ajp.2015.15060833.

Zik, J., Deveney, C.M., Ellingson, J.M., Haller, S.P., Kircanski, K., Cardinale, E.M., Brotman, M.A., Stoddard, J., 2021. Understanding irritability in relation to anger, aggression, and informant in a pediatric clinical population. J. Am. Acad. Child Adolesc. Psychiatry. https://doi.org/10.1016/j.jaac.2021.08.012. S0890856721013563.