

Transdiagnostic Irritability in a Clinical Sample Aged 7–18 Years: Diagnostic Correlates, Parent– Child Perspectives, and Psychometric Properties of the Turkish Tonic and Phasic Irritability Scale (TAPIS-TR)

Makbule Esen Öksüzoğlu

mesenoksuzoglu@kastamonu.edu.tr

Kastamonu University

Elif Anaç

Kastamonu Education and Research Hospital

Özge Gözaçanlar Özkan

Uşak University

Buse Akıncı

Uşak Training and Research Hospital

Hilal Subaşı

Torbalı State Hospital

Nagihan Özer

Şırnak State Hospital

Duygu Kınay Ermiş

İstanbul Başakşehir Çam ve Sakura Şehir Hastanesi

Ebrar Onar

Sincan Training and Research Hospital

Zeliha Büşra Özdemir

Bitlis Tatvan State Hospital

Hande Günal Okumuş

Uşak Training and Research Hospital

Jamilah Silver

University of North Carolina at Chapel Hill

Research Article

Keywords: transdiagnostic irritability, multi-informant assessment, psychiatric comorbidity, tonic and phasic irritability

Posted Date: October 18th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10838451/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Additional Declarations: No competing interests reported.

Abstract

Background: Irritability is common across child and adolescent psychiatric disorders, but distribution in routine clinical populations, informant-specific value, and tonic–phasic structure remain uncertain.

Methods: We studied 542 consecutively referred child–parent dyads aged 7–18 years across seven public hospitals in Türkiye. Diagnoses were established with a semi-structured psychiatric interview, and severity was assessed with a clinician-rated global severity scale. We examined child and parent scores on the Turkish Tonic and Phasic Irritability Scale (TAPIS-TR) alongside symptom, self-esteem, and distress measures. Diagnostic models were adjusted for age, gender, comorbidity, and recruiting site.

Results: Irritability was raised across the clinic but unevenly distributed, rising with diagnosis count and clinician-rated severity. Depressive disorders, attention-deficit/hyperactivity disorder (ADHD), oppositional defiant disorder (ODD) and conduct disorder (CD) were associated with higher irritability (aORs per SD 1.44–3.49, $q < .01$), independently of comorbidity and recruiting site; anxiety, obsessive–compulsive and learning disorders were not. Agreement was moderate (ICC = .65): child report carried depressive disorders; parent report carried ODD and CD; both carried ADHD. The TAPIS-TR was reliable ($\alpha = .94–.95$) and its two-factor structure fitted better than a single factor; the facets correlated .85–.86 under a restrictive confirmatory model but .74 when cross-loadings were permitted, and differed in external correlates ($q \leq .002$), while a general factor accounted for most common variance.

Conclusion: Irritability functions primarily as a transdiagnostic marker of overall clinical burden, while child and parent reports provide complementary disorder-specific information. The TAPIS-TR reliably assesses general irritability, with tonic and phasic facets best interpreted as secondary components rather than independent indices.

1. INTRODUCTION

Irritability is characterised by an increased tendency to experience anger, particularly in response to frustration, blocked goals, or perceived provocation [1, 2]. Although related to anger, aggression, and emotion dysregulation, irritability is conceptually distinct: anger refers to an emotional state, aggression to behaviour intended to cause harm, and emotion dysregulation to broader difficulties in modulating emotional states, whereas irritability specifically reflects a heightened proneness to anger in response to frustration [1, 3, 4]. Some degree of irritability is developmentally normative, particularly in younger children. Still, it becomes clinically significant when it is persistent, severe, developmentally or contextually inappropriate, difficult to control, and associated with functional impairment [3, 5, 6]. Irritability is among the most common presenting concerns in child and adolescent mental health services. It occurs across a broad range of conditions, including depressive and anxiety disorders, attention-deficit/hyperactivity disorder (ADHD), oppositional defiant disorder (ODD), conduct disorder (CD), autism spectrum disorder, and disruptive mood dysregulation disorder (DMDD) [3, 7].

Irritability also carries prognostic weight. It shows moderate stability over time and predicts later psychopathology: irritability in childhood and adolescence has been linked to subsequent depressive and anxiety disorders [2], and preschool irritability to later internalizing and externalizing problems [8]. In a United Kingdom birth cohort of 7,225 children, persistently high irritability between ages 3 and 7 years remained associated with depressive symptoms and self-harm at age 14 after adjustment for sociodemographic characteristics, early mental health problems, and cognitive development [9]. Trajectory studies have identified low, decreasing, increasing, and high-persistent patterns, so the course is clearly not uniform [10, 11].

Its associations with both internalizing and externalizing psychopathology have led to irritability being treated as a transdiagnostic dimension rather than a symptom belonging to one diagnosis [2, 12]. Klein et al. [7] described this as multifinality: comparable levels of irritability may precede quite different outcomes in different children. Most of the evidence for that position, however, comes from community samples and longitudinal cohorts, in which irritability is common, comorbidity is comparatively rare, and diagnostic groups can be kept reasonably clean. The clinic is a different setting. Children arrive already impaired, most carry more than one diagnosis, and the practical question is not whether irritability is transdiagnostic but how it is distributed once everyone in the room is unwell: whether it distinguishes between diagnoses, whether it simply tracks how ill the child is, and how much of it is explained by the comorbidity that accompanies each diagnosis. Those questions are difficult to answer in community data and have rarely been addressed in a referred sample large enough to hold comorbidity constant.

A further complication is internal to the construct. Current models distinguish two temporally related manifestations: tonic irritability, a persistent angry, grouchy, or easily annoyed mood, and phasic irritability, brief and intense temper outbursts that are developmentally or contextually inappropriate [1, 13]. These are not simply the internalizing and externalizing faces of irritability, but longitudinal work suggests partly separate correlates: tonic irritability has been linked more closely to depressive and anxiety disorders and to suicidal ideation, and phasic irritability to ADHD, ODD, CD, substance use, and other externalizing outcomes [14, 15]. Empirically, though, the two overlap heavily. In the Great Smoky Mountains Study, they frequently co-occurred and predicted one another over time [5], and factor analytic studies support their separability while reporting correlations high enough to question their practical value [14, 16]. Progress has been slowed by measurement: many studies derived tonic and phasic scores from selected items in diagnostic interviews or from general measures of psychopathology and temperament that were never designed for the purpose [2, 13]. Irritability-specific instruments such as the Affective Reactivity Index (ARI) improved matters by focusing on the threshold, frequency, and duration of irritable feelings and behaviours rather than on aggression [17], but the ARI was not built to assess the two dimensions separately.

Whose account is used matters as well. Young people have privileged access to persistent inner tension and to their own experience of frustration, whereas parents observe outbursts and their consequences; discrepancies between them are therefore likely to carry information rather than only error [18]. Klein et al. [7] made much the same point, suggesting that youth may be the better reporters of tonic irritability and parents the better reporters of phasic irritability. In this regard, the Tonic and Phasic Irritability Scale (TAPIS) was developed as the first multi-informant measure designed specifically to capture both dimensions in school-age children and adolescents [13]. Its clinical validation sample comprised 92 parent–youth dyads, and the tonic–phasic correlations were .85 for youth and .90 for parent report, so how the two dimensions behave in a large and diagnostically mixed clinical population remains open. No Turkish instrument assesses tonic and phasic irritability separately with parallel youth- and parent-report forms, although a clinician-rated measure of general irritability is available [19], and cross-cultural use of any such measure requires local evaluation rather than translation alone [20, 21].

The present study therefore examined irritability across the diagnostic spectrum of a large, consecutively referred clinical sample of children and adolescents aged 7 to 18 years, using parallel child and parent reports. We asked four questions. First, how is irritability distributed across the diagnoses seen in routine practice, and which diagnoses remain associated with it once age, gender, and comorbid diagnoses are taken into account? Second, does irritability track the number of concurrent diagnoses and clinician-rated illness severity, as would be expected if it indexes overall psychiatric burden rather than any particular disorder? Third, what does each informant contribute, and are child and parent reports interchangeable? Fourth, beyond distinguishing a one-factor general model from a correlated two-factor tonic–phasic structure, does separating tonic from phasic irritability add clinically useful information, tested formally by comparing their associations with external validators through direct coefficient-difference tests

rather than inferred from numerical differences in correlations or odds ratios alone? Answering them required an instrument that was not previously available in Turkish. We therefore adapted both TAPIS forms into Turkish (TAPIS-TR) and, as a fifth aim, examined their content validity, factor structure, reliability, convergent validity, and measurement invariance in the same sample, so that the clinical findings rest on a documented instrument and the Turkish forms become available for routine clinical and research use.

2. METHODS

2.1. Participants

This multicentre study recruited child–parent dyads consecutively from child and adolescent psychiatry outpatient clinics at seven public hospitals in Türkiye: Uşak Training and Research Hospital, Kastamonu Training and Research Hospital, Şırnak State Hospital, İzmir Torbalı State Hospital, Bitlis Tatvan State Hospital, Sincan Training and Research Hospital, and İstanbul Başakşehir Çam ve Sakura City Hospital. Data were collected between May 1 and August 1, 2026. The number of dyads contributed per centre ranged from 25 to 169.

Inclusion criteria were (1) age between 7 and 18 years; (2) presentation to a child and adolescent psychiatry outpatient clinic; (3) sufficient proficiency in Turkish by both the child or adolescent and the accompanying parent to complete the instruments; and (4) written informed consent from a parent or legal guardian. Participants were excluded if a severe neurological or developmental condition precluded reliable completion of the assessments, or if either the young person or the parent declined to take part. Participants with substantial missing data were excluded from analysis; this included six children with DMDD, none of whom yielded a complete child–parent dyad (see Strengths and Limitations). Because recruitment was consecutive and no diagnostic group was pre-specified or over-sampled, the resulting sample reflects the case mix of general child psychiatry outpatient practice.

Clinicians established psychiatric diagnoses using the Schedule for Affective Disorders and Schizophrenia for School-Age Children–Present and Lifetime Version (K-SADS-PL) [22, 23]. Eligibility and exclusion criteria were assessed during the diagnostic interview.

A minimum of 300 child–parent dyads was set a priori. Because the 19-item child/adolescent and parent forms were to be evaluated separately, this ensured at least 300 responses for the factor-analytic evaluation of each form, a target informed by simulation evidence that roughly 300 participants are generally adequate for recovering the factor structure of psychiatric rating scales, with requirements varying by model complexity, number of factors and indicators, loading magnitude, estimator, and missingness [24, 25]. Recruitment continued well beyond that minimum in order to obtain usable numbers in the smaller diagnostic categories and to estimate diagnosis-specific and informant-specific effects with reasonable precision.

The final analytic sample comprised 542 child–parent dyads. Mean age was 12.90 years (SD = 2.98); 260 (48.0%) were boys and 282 (52.0%) girls. Sociodemographic and clinical characteristics are given in Table 1.

Table 1
Sociodemographic and clinical characteristics of the sample (N = 542)

Characteristic	n or M	% or SD
Age, years, M (SD)	12.90	2.98
Girls / boys, n	282 / 260	52.0 / 48.0%
Children (7–12) / adolescents (13–18), n	256 / 286	47.2 / 52.8%
Intact (nuclear) family, n (%)	452	83.4
Number of siblings, M (SD)	1.72	1.20
BMI z-score, M (SD)	0.47	1.17
Maternal higher education ($\geq$ high school), n (%)	257	47.6
Paternal higher education ($\geq$ high school), n (%)	291	55.7
Higher socioeconomic status (level 4–5), n (%)	353	65.1
Family psychiatric history, n (%)	228	42.1
Past history		
Past psychiatric (K-SADS) diagnosis, n (%)	278	51.3
Past psychotropic medication, n (%)	210	38.7
Past non-psychiatric medical illness, n (%)	62	11.4
Current non-psychiatric medical comorbidity, n (%)	70	12.9
Clinical characteristics		
Any current K-SADS diagnosis, n (%)	496	91.5
No current psychiatric diagnosis, n (%)	46	8.5
Number of diagnoses, M (SD)	1.59	1.03
One / two / ≥ 3 diagnoses, n	251 / 162 / 83	
ADHD, n (%)	260	48.0
Any anxiety disorder, n (%)	163	30.1
Any depressive disorder, n (%)	95	17.5
Oppositional defiant disorder, n (%)	80	14.8
OCD and related disorders, n (%)	56	10.3
Specific learning disorder, n (%)	41	7.6

Note. ADHD = attention-deficit/hyperactivity disorder; OCD = obsessive-compulsive disorder; CGI-S = Clinical Global Impression–Severity; K-SADS = Kiddie Schedule for Affective Disorders and Schizophrenia; BMI = body-mass index. Parental higher education denotes completion of high school or above; higher socioeconomic status denotes the top two of five ordinal levels. Family psychiatric history and current/past medical comorbidities are detailed in Supplementary Table S1. "No current psychiatric diagnosis" (n = 46) denotes participants with no current K-SADS diagnosis of any kind. In the mutually exclusive diagnostic cluster classification, "Other non-indexed diagnosis" (n = 27) denotes participants whose only diagnosis fell outside the seven indexed categories.

Characteristic	n or M	% or SD
Conduct disorder, n (%)	26	4.8
Bipolar / trauma-related, n	2 / 9	
CGI-S, median (IQR)	4	3–5
Diagnostic cluster (mutually exclusive)		
No current diagnosis, n (%)	46	8.5
Other non-indexed diagnosis, n (%)	27	5.0
Internalizing only, n (%)	189	34.9
Neurodevelopmental only, n (%)	136	25.1
Externalizing only, n (%)	7	1.3
Mixed across clusters, n (%)	137	25.3
Current medication		
Any psychotropic, n (%)	281	51.8
SSRI/SNRI / methylphenidate / antipsychotic, n	138 / 129 / 106	
<i>Note.</i> ADHD = attention-deficit/hyperactivity disorder; OCD = obsessive–compulsive disorder; CGI-S = Clinical Global Impression–Severity; K-SADS = Kiddie Schedule for Affective Disorders and Schizophrenia; BMI = body-mass index. Parental higher education denotes completion of high school or above; higher socioeconomic status denotes the top two of five ordinal levels. Family psychiatric history and current/past medical comorbidities are detailed in Supplementary Table S1. "No current psychiatric diagnosis" (n = 46) denotes participants with no current K-SADS diagnosis of any kind. In the mutually exclusive diagnostic cluster classification, "Other non-indexed diagnosis" (n = 27) denotes participants whose only diagnosis fell outside the seven indexed categories.		

2.2. Procedure

Permission to translate and use the TAPIS [13] was obtained from Dr Jamilah Silver, on behalf of the original development team, on April 7, 2026. The Turkish adaptation followed a forward- and back-translation procedure. Two bilingual clinicians specialising in child and adolescent mental health independently translated the child/adolescent and parent forms into Turkish, and discrepancies were resolved by consensus with priority given to conceptual equivalence, linguistic clarity, cultural relevance, and developmental appropriateness for the 7- to 18-year age range. A third bilingual translator who had not taken part in the forward translation back-translated the reconciled version into English, and remaining discrepancies were reviewed with Dr Silver and the development team. Before data collection, an expert panel of 22 child and adolescent psychiatrists rated both forms for item clarity, conceptual equivalence, cultural relevance, developmental appropriateness, and overall content validity, after which the preliminary forms were piloted in 10 child–parent dyads to check comprehensibility and ease of completion. Minor wording revisions followed the pilot.

The protocol was approved by the Non-Interventional Clinical Research Ethics Committee of Uşak University (Approval No. 1235-1235-14; April 28, 2026) and the study was conducted in accordance with institutional and national regulations and the Declaration of Helsinki. Guardians and young people received age-appropriate verbal and written information about the aims, procedures, duration, risks and benefits, data confidentiality, and the right to withdraw without penalty. Written informed consent was obtained from a parent or legal guardian and assent from the child or adolescent.

Sociodemographic and clinical data were recorded by the attending clinician. Diagnoses were made with the Turkish adaptation of the K-SADS-PL [23], integrating information from the young person and the parent by clinical judgement. After the diagnostic assessment, parents completed the parent forms of the TAPIS, the Strengths and Difficulties Questionnaire (SDQ), and the ARI. Young people completed the self-report TAPIS and ARI and, where age-appropriate, the SDQ, the Rosenberg Self-Esteem Scale (RSES), and the Brief Symptom Inventory (BSI). The clinician completed the sociodemographic and clinical form, the K-SADS-PL, and the Clinical Global Impression–Severity scale (CGI-S). A subsample of dyads repeated the TAPIS after a median of 30 days (IQR 24–41; range 7–70); complete retest data were available for 59 child and 59 parent reports.

2.3. Measures

2.3.1. Sociodemographic and Clinical Information Form

This clinician-completed form recorded age, gender, school grade, parental sociodemographic characteristics, developmental and medical history, psychiatric and treatment history, age at symptom onset, symptom duration and severity, comorbid conditions, and family psychiatric history.

2.3.2. Schedule for Affective Disorders and Schizophrenia for School-Age Children–Present and Lifetime Version (K-SADS-PL)

The K-SADS-PL is a semi-structured, clinician-administered interview for current and lifetime psychiatric disorders in children and adolescents [22]. The Turkish DSM-5 adaptation [23] was used to establish diagnoses and determine eligibility. Children, adolescents, and parents were interviewed separately, and the information was integrated by clinical judgement. For analysis, seven diagnostic categories with adequate cell sizes were indexed a priori: any depressive disorder, any anxiety disorder, obsessive–compulsive and related disorders, ADHD, ODD, CD, and specific learning disorder (SLD).

2.3.3. Clinical Global Impression–Severity Scale (CGI-S)

The CGI-S [26] is a clinician-rated 7-point global severity scale ranging from 1 (normal, not at all ill) to 7 (among the most extremely ill patients). Child and adolescent psychiatrists at each of the seven centres rated overall symptom severity after a full clinical evaluation, and a pre-study meeting was held to standardise the approach across sites. The CGI-S served as the index of overall clinical severity, as a covariate in the parent–child discrepancy model, and as a control in the adjusted medication analyses.

2.3.4. Tonic and Phasic Irritability Scale (TAPIS)

The TAPIS [13] is a multi-informant measure of irritability in children and adolescents with two related dimensions: tonic irritability (8 items), reflecting a persistently angry or irritable mood, and phasic irritability (11 items), reflecting brief, intense temper outbursts. Parallel child/adolescent self-report and parent-report forms are rated on a five-point scale. Subscale scores are the means of the relevant items, and the general irritability score is the mean of all 19 items, with higher scores indicating greater irritability. The general score was the primary index throughout, and the tonic and phasic subscales were examined secondarily. Internal consistency in the present sample was excellent for both forms (Cronbach's $\alpha = .94$ child, $.95$ parent).

2.3.5. Affective Reactivity Index (ARI)

The ARI is a brief dimensional measure of irritability [17] comprising six symptom items and one impairment item, rated 0 (not true) to 2 (certainly true), with a symptom total ranging from 0 to 12. Parallel child/adolescent and parent forms have shown good psychometric properties in clinical and community samples [17, 27, 28], and the Turkish

adaptation was undertaken by Kocael [29]. The ARI total and the impairment item served as the principal convergent validators of the TAPIS. Internal consistency here was good ($\alpha = .88$ child, $.90$ parent).

2.3.6. Strengths and Difficulties Questionnaire (SDQ)

The SDQ is a 25-item screening measure of emotional and behavioural difficulties with five subscales: emotional symptoms, conduct problems, hyperactivity/inattention, peer relationship problems, and prosocial behaviour [30]. The Turkish version has established validity and reliability [31]. Both child and parent versions were used to characterise the symptom context of irritability, with the expectation that tonic irritability would relate comparatively more strongly to emotional symptoms and phasic irritability to conduct problems and hyperactivity/inattention. In the present study, internal consistency was acceptable for both forms ($\alpha = .79$).

2.3.7. Rosenberg Self-Esteem Scale (RSES)

The RSES is a 10-item self-report measure of global self-worth [32], rated on a 4-point scale with five items reverse-scored. The Turkish adaptation of the 10-item self-esteem subscale was carried out by Çuhadaroğlu [33], who reported a validity coefficient of $.71$ against interview-based classifications and test–retest reliability of $.75$; it has since been widely used in adolescent and young adult samples [34]. Scores were coded so that higher values indicate lower self-esteem. In the present study, internal consistency was good ($\alpha = .83$).

2.3.8. Brief Symptom Inventory (BSI)

The BSI is a 53-item self-report measure covering nine symptom dimensions and three global indices, rated 0 (not at all) to 4 (extremely) [35]. The Turkish adaptation was by Şahin and Durak [36], with psychometric evaluation in adolescents by Şahin et al. [37]. Selected indices were used here: the Global Severity Index and the hostility subscale in relation to parent-reported irritability, adolescent-reported hostility as a predictor of the parent-minus-child discrepancy, and the depression subscale both as a validator of the tonic and phasic subscales and as a covariate in that model. In the present study, internal consistency was excellent for the Global Severity Index ($\alpha = .97$) and good for the depression ($\alpha = .89$) and hostility ($\alpha = .85$) subscales.

2.4. Statistical Analysis

Analyses were carried out in IBM SPSS Statistics 28.0, Python 3.12, and R 4.6.0. All scores were recomputed from item-level data. The general irritability score, the mean of all 19 items, was the primary index; the tonic and phasic subscales were secondary. Analyses were complete-case. The analytic n was 542 for the dual-informant and diagnostic models and ranged from 59 (test–retest) to 542 elsewhere, depending on the measure and the covariates available.

Diagnostic groups overlap, so each of the seven indexed diagnoses was first compared present versus absent with Mann–Whitney U tests and rank-biserial effect sizes. Mutually exclusive diagnostic-cluster groupings and depression-by-anxiety groups were compared with Kruskal–Wallis tests, Dunn post-hoc comparisons, and epsilon-squared. An ordered increase along the externalizing spectrum was tested by correlating irritability (Spearman) with a sequence fixed in advance of inspecting the irritability data. The association of the general score (per 1 SD) with each diagnosis was estimated with Firth penalized logistic regression adjusted for age and gender (Model A). Because comorbidity is pervasive in a referred sample and may act as a mediator or a collider, two sensitivity models were specified before these outcomes were examined, although the analysis plan was not preregistered: Model B added the number of other diagnoses, and Model C added the specific comorbid diagnoses relevant to each outcome. Because recruitment spanned seven centres, a further sensitivity model added fixed effects for the recruiting centre (six

indicator terms), so that between-centre heterogeneity in case mix and rating practice was held constant. Model A remained the primary specification.

The tonic and phasic subscales were examined within the same framework. Each was entered alone and then jointly with the other in the same Firth model, so that single-predictor and mutually adjusted estimates could be read side by side, and the contrast between the two mutually adjusted coefficients was tested with a Wald statistic. Differences between the tonic and phasic correlations with a common validator were tested with Williams's test for dependent correlations sharing one variable rather than inferred from the size of the two coefficients. Whether separating the components improved model fit was judged by the change in AIC relative to the corresponding general-score model; a two-point difference was used as a conventional benchmark for improved fit rather than as a formal decision threshold.

Comorbidity burden was modelled with a linear mixed-effects model (irritability = burden $\times$ informant + age + gender, burden coded 0–3) and, as a check, a negative-binomial model of diagnosis count. Associations with clinician-rated severity used Spearman correlations for each informant.

Parent–child agreement was quantified with absolute-agreement ICC(2,1) and Bland–Altman analysis. Informant differences by diagnosis were estimated with per-diagnosis linear mixed-effects models (irritability = informant $\times$ diagnosis + age + gender, random participant intercept), with a sensitivity model adding informant $\times$ comorbid–diagnosis interactions. Predictors of the parent-minus-child discrepancy, including the adolescent's self-reported hostility subscale of BSI, were examined in HC3 robust regression controlling for dyad-mean irritability, depression subscale of BSI, CGI-S severity, ODD, depressive disorder, age, and gender. The incremental value of each informant was tested with nested Firth models and penalized likelihood-ratio tests.

Measurement analyses were secondary. Content validity had been rated a priori by the 22 expert panellists on four criteria using a 4-point relevance scale, summarised with item- and scale-level content validity indices (I-CVI, S-CVI) and modified kappa. Item-level analyses covered floor and ceiling effects, corrected item–subscale correlations, and distributional statistics. Reliability used Cronbach's α , McDonald's ω , and test–retest ICC(2,1) with the standard error of measurement and the minimal detectable change (MDC95). Structure was examined with exploratory factor analysis (principal-axis extraction, oblimin rotation, parallel analysis) and confirmatory WLSMV models (CFI, TLI, RMSEA, SRMR), in which a single-factor solution and the hypothesised correlated two-factor solution were compared for each informant with a robust scaled difference test. Because a confirmatory model fixes every cross-loading to zero and can therefore inflate the estimated correlation between two facets whose items legitimately tap both, the two-factor solution was re-estimated as an exploratory structural equation model (ESEM) with Browne's partially specified oblique target rotation, in which each item's loading on its non-assigned factor was targeted to zero but left free to depart from it. A Schmid–Leiman bifactor solution (ECV, ω_H , ω_{HS} , H) was retained as a decomposition of general and specific variance rather than as the test of facet separability. Convergent validity used Spearman correlations with the ARI, and factor distinctiveness the average variance extracted, the Fornell–Larcker criterion, and the heterotrait–monotrait ratio. Configural and metric invariance across gender, age group, and current medication status were tested by constraining loadings ($\Delta CFI \geq -.010$, $\Delta RMSEA \leq .015$); scalar invariance was not evaluated, so the invariance results support comparison of structure and loadings but not of latent means.

Firth models used a Jeffreys-penalty iteratively reweighted least-squares implementation and were re-estimated with an independent optimizer, agreeing to seven decimal places across all 14 diagnosis-by-informant models (7 diagnoses $\times$ 2 informants). The false discovery rate was controlled (Benjamini–Hochberg) within ten families defined by the analytic question each set of tests addressed, and fixed before any test statistic or p value was inspected: general-score diagnosis, informant $\times$ diagnosis, subscale–diagnosis, tonic-versus-phasic subscale contrast,

incremental informant, unadjusted diagnosis– and medication–subscale, medication, depression × anxiety, severity correlation, and convergent validator. Each family groups tests that share a hypothesis and an outcome metric, so that correction is applied across the comparisons a reader would interpret together rather than across the study as a whole; families were analysed in a confirmatory-to-exploratory order, and FDR was applied within each. Sensitivity models and the discrepancy regression are reported with nominal p values; $q < .05$ was taken as significant.

3. RESULTS

3.1. Sample and Diagnostic Composition

Five hundred and forty-two children and adolescents (282 girls, 52.0%) aged 7 to 18 years ($M = 12.90$, $SD = 2.98$) took part with a parent. A current K-SADS diagnosis was present in 496 (91.5%); 46 had none, and a further 27 had only a diagnosis outside the seven indexed categories. Comorbidity was common: 251 children carried a single diagnosis, 162 met criteria for two disorders, and 83 for three or more. ADHD was the commonest diagnosis (48.0%), followed by anxiety disorders (30.1%), depressive disorders (17.5%), ODD (14.8%), obsessive–compulsive and related disorders (OCRD) (10.3%), SLD (7.6%) and CD (4.8%). The median CGI-S was 4 (IQR 3–5), and 281 children (51.8%) were taking a psychotropic medication. A family psychiatric history, recorded in 228 children (42.1%), was associated with higher irritability on both reports ($p < .001$); current non-psychiatric medical comorbidity, present in 70 (12.9%), was not ($p > .10$; Supplementary Table S1). Full characteristics are given in Table 1.

Table 1

3.2. Distribution of Irritability Across the Clinic

Both informants used the full range of the scale, and the score distributions were broad rather than concentrated at the lower end (child self-report $M = 1.50$, $SD = 0.97$; parent report $M = 1.46$, $SD = 0.94$; floor 3.5% and 4.1%, ceiling 0.4% for both; Supplementary Figure S1 and Supplementary Table S2). Floor endorsement was most pronounced for the two items describing the most severe functional consequences of temper, namely removal from activities and the reorganisation of family routines around the child, although several other severe phasic items also showed substantial floor endorsement. Girls rated themselves as more irritable than boys ($M = 1.65$ vs 1.35 , $p < .001$, $r = .18$); the corresponding parent-reported difference was a non-significant trend (1.53 vs 1.39 , $p = .07$). Self-reported irritability increased modestly with age ($\rho = .15$, $p < .001$), whereas parent-reported irritability did not ($\rho = .02$, $p = .73$; Supplementary Table S3).

Across diagnoses, differences took the form of a gradient rather than discrete elevations. At the unadjusted subscale level, depressive disorders and CD showed the broadest elevations, whereas the ADHD and ODD differences were more concentrated in phasic and/or parent-reported scores (Supplementary Table S4); the mutually exclusive groups nonetheless overlapped substantially (Fig. 1). Setting aside the eight children whose only diagnosis was SLD, the lowest scores were in children with no current diagnosis and the highest in the disruptive-behaviour and depressive groups, with the intermediate groups closely spaced. Grouping by broad diagnostic cluster gave the same result, with the lowest scores again in children with no current diagnosis (child $M = 0.96$, parent $M = 0.94$) and small effect sizes (child $\epsilon^2 = .10$, $H(5) = 56.4$; parent $\epsilon^2 = .12$, $H(5) = 67.3$; both $p < .001$); the highest-scoring group was not a single cluster but the children whose diagnoses spanned more than one (child $M = 1.88$, parent $M = 1.96$; Supplementary Table S5). Along the predefined externalizing sequence – no disruptive-behaviour disorder, ADHD alone, ODD, then CD – irritability increased steadily, more so on parent than on child report (child $\rho = .12$, $p = .005$; parent $\rho = .26$, $p < .001$; Supplementary Table S6).

Figure 1

3.3. Diagnosis-Specific Associations

Four diagnoses were associated with general irritability after adjustment for age and gender (Table 2, Fig. 2A). Depressive disorders, ADHD, ODD, and CD each became more likely as irritability rose, on child report (aORs per SD: 1.82 [1.42, 2.34], 1.44 [1.18, 1.75], 1.63 [1.27, 2.08] and 2.28 [1.53, 3.38]) and on parent report (1.48 [1.17, 1.88], 1.55 [1.28, 1.89], 2.29 [1.76, 2.96] and 3.49 [2.24, 5.44]; all $q < .001$ apart from parent-reported depressive disorder, $q < .01$). Anxiety disorders, OCD, and SLD showed no association on either report.

Table 2

Parent- and child-reported general irritability by diagnosis: descriptive scores, informant agreement, adjusted associations, informant discrepancy, and incremental clinical information (N = 542)

Diagnosis	n	Child M (SD)	Parent M (SD)	ICC (95% CI)	Inf. × dx b (95% CI)	Child aOR (95% CI)	Parent aOR (95% CI)	Joint aOR (C / P)	Parent adds q	Child adds q
Depressive	95	2.07 (0.90)	1.76 (0.90)	.59 (.41–.72)	–.33 (–.50, –.15)**	1.82 (1.42–2.34)***	1.48 (1.17–1.88)**	1.77 / 1.04	.058	< .001
Anxiety	163	1.62 (0.93)	1.41 (0.81)	.54 (.42–.65)	–.24 (–.39, –.10)**	1.05 (0.86–1.27)	0.88 (0.72–1.07)	1.24 / 0.77	.012	.019
OCD/related	56	1.74 (1.01)	1.66 (0.98)	.63 (.44–.76)	–.04 (–.26, .18)	1.24 (0.95–1.63)	1.24 (0.95–1.63)	1.14 / 1.15	.058	.058
ADHD	260	1.57 (0.97)	1.61 (0.94)	.63 (.55–.69)	+.16 (.02, .29)*	1.44 (1.18–1.75)***	1.55 (1.28–1.89)***	1.15 / 1.43	.002	.038
ODD	80	1.81 (0.97)	2.07 (0.84)	.38 (.18–.55)	+.36 (.17, .55)**	1.63 (1.27–2.08)***	2.29 (1.76–2.96)***	0.92 / 2.40	< .001	.058
Conduct	26	2.23 (1.07)	2.56 (0.92)	.70 (.42–.86)	+.39 (.07, .70)*	2.28 (1.53–3.38)***	3.49 (2.24–5.44)***	1.03 / 3.42	< .001	.094
SLD	41	1.44 (1.05)	1.67 (1.03)	.67 (.46–.81)	+.30 (.04, .55)*	1.01 (0.73–1.40)	1.30 (0.95–1.77)	0.76 / 1.55	.014	.044

Note. General (total) irritability = mean of all 19 TAPIS-TR items (0–4). Diagnostic groups overlap; participants with comorbidity may contribute to more than one row (contrast Fig. 1, mutually exclusive groups). Child M (SD) / Parent M (SD) = group means. ICC(2,1) = two-way random-effects, absolute-agreement, single-measures intraclass correlation (parent–child concordance within the diagnosis-positive group). Inf. × dx b = parent-minus-child difference in general irritability associated with the diagnosis, from separate linear mixed-effects models (age + gender, random participant intercept; negative = children report relatively higher, positive = parents higher; FDR across seven interactions). Child aOR / Parent aOR = per-SD odds ratios from separate Firth penalized logistic regressions adjusted for age and gender (Model A; FDR within the 14-coefficient general-score family); comorbidity-adjusted Models B and C are in Table 3. Joint aOR (C / P) = child and parent odds ratios when both are entered together. Parent adds q / Child adds q = FDR-corrected penalized likelihood-ratio tests of adding one informant to a model already containing the other (14-test incremental family; seven diagnoses × two directions). * $q < .05$; ** $q < .01$; *** $q < .001$. Comorbidity-interaction sensitivity analyses and full joint-model statistics are in Supplementary Tables S7 and S9.

Table 2

Figure 2

All four associations were retained after adjustment for the number of other diagnoses and for the specific comorbidities relevant to each outcome, and several increased slightly (Table 3). CD produced the largest estimates on both reports; with 26 children in this category, the corresponding confidence intervals were the widest in the analysis. Two further effects reached nominal significance only in the comorbidity-adjusted models: child-reported irritability and anxiety disorders (aOR = 1.27 [1.03, 1.57], $p = .029$), and parent-reported irritability and OCRD (aOR = 1.33 [1.00, 1.76], $p = .048$). Adding fixed effects for the recruiting centre left the pattern essentially unchanged: all four associations remained significant on both reports (centre-adjusted aORs 1.42–3.05, all $q < .01$), with estimates shifting by no more than 13% and by less than 5% in five of the eight primary comparisons. Anxiety disorders and SLD remained null on both reports. Parent-reported OCRD crossed the significance threshold in this model (aOR = 1.37 [1.03, 1.82], $q = .049$), as it had in the comorbidity-adjusted model, and is treated as marginal throughout. Between-centre differences therefore do not account for the diagnostic associations reported here.

Table 3
General-score diagnosis associations across comorbidity-adjusted models (Models A–C; N = 542)

Diagnosis / informant	n	Model A aOR (95% CI), q	Model B aOR (95% CI), p	Model C aOR (95% CI), p
Depressive (child)	95	1.82 (1.42–2.34)***	2.24 (1.70–2.96), < .001	2.20 (1.67–2.90), < .001
Depressive (parent)	95	1.48 (1.17–1.88)**	1.71 (1.32–2.20), < .001	1.69 (1.30–2.18), < .001
ADHD (child)	260	1.44 (1.18–1.75)***	1.60 (1.30–1.97), < .001	1.59 (1.26–2.02), < .001
ADHD (parent)	260	1.55 (1.28–1.89)***	1.72 (1.40–2.12), < .001	1.41 (1.12–1.77), .003
ODD (child)	80	1.63 (1.27–2.08)***	1.64 (1.28–2.12), < .001	1.44 (1.10–1.88), .007
ODD (parent)	80	2.29 (1.76–2.96)***	2.34 (1.80–3.05), < .001	2.14 (1.61–2.86), < .001
Conduct (child)	26	2.28 (1.53–3.38)***	2.38 (1.59–3.56), < .001	2.11 (1.41–3.15), < .001
Conduct (parent)	26	3.49 (2.24–5.44)***	3.71 (2.37–5.82), < .001	3.31 (2.10–5.19), < .001
Anxiety (child)	163	1.05 (0.86–1.27)	1.08 (0.88–1.33), .450	1.27 (1.03–1.57), .029
Anxiety (parent)	163	0.88 (0.72–1.07)	0.90 (0.73–1.10), .299	1.02 (0.83–1.26), .839
OCD/related (child)	56	1.24 (0.95–1.63)	1.27 (0.96–1.69), .096	1.29 (0.97–1.72), .077
OCD/related (parent)	56	1.24 (0.95–1.63)	1.27 (0.96–1.67), .093	1.33 (1.00–1.76), .048
SLD (child)	41	1.01 (0.73–1.40)	1.01 (0.72–1.42), .942	0.93 (0.67–1.29), .660
SLD (parent)	41	1.30 (0.95–1.77)	1.33 (0.97–1.83), .081	1.20 (0.87–1.64), .264
<i>Note.</i> General (total) irritability, per 1 SD, Firth penalized logistic regression. Model A: adjusted for age and gender (primary; FDR-corrected q; * q < .05, ** q < .01, *** q < .001). Model B: + number of OTHER (non-target) diagnoses (nominal p). Model C: + the relevant comorbid diagnoses per outcome (depressive: ADHD, ODD, anxiety; anxiety: depressive, ADHD; OCD/related: anxiety, depressive; ADHD: ODD, CD, anxiety, depressive; ODD: ADHD, CD; conduct: ADHD, ODD; SLD: ADHD; nominal p). Because burden and comorbid diagnoses may be mediators or colliders, Models B and C are sensitivity analyses and Model A is primary. The depressive, ADHD, ODD, and conduct-disorder associations all remained significant across models.				

Table 3

3.4. Comorbidity Burden and Clinical Severity

Irritability increased almost linearly with the number of concurrent diagnoses on both reports (Fig. 3A), by .37 of a scale point per additional diagnostic category ($p < .001$), with no difference between informants in the size of that increase (interaction $p = .63$); modelling the diagnosis count directly gave the same result (IRR = 1.12 per SD, $p =$

.011). Both reports were also correlated with clinician-rated severity (p with CGI-S: child .38, parent .39; both $q < .001$; Table 4, Fig. 3B).

Table 4

Clinical and treatment correlates of general irritability: comorbidity burden, clinical severity, and medication associations (N = 542)

Effect / correlate	n	Estimate (95% CI)	p / q		
Comorbidity burden (linear mixed-effects model)					
Diagnosis burden main effect (b per level)	542	+ .37 (.28, .46)	< .001		
Burden × informant interaction	542	– .02 (–.10, .06)	.63		
Clinical severity (Spearman ρ)	n	Child ρ	Child q	Parent ρ	Parent q
CGI-S severity	542	.38***	< .001	.39***	< .001
Number of diagnoses	542	.34***	< .001	.33***	< .001
Medication associations (Firth; base / adjusted aOR per SD)	n	Base aOR (95% CI)	q	Adjusted aOR (95% CI)	p
Any psychotropic medication (child)	281	1.28 (1.07–1.53)*	.012	0.99 (0.81–1.22)	.953
Any psychotropic medication (parent)	281	1.29 (1.08–1.54)*	.011	1.01 (0.82–1.24)	.917
SSRI/SNRI (child)	138	1.19 (0.97–1.46)	.123	0.96 (0.74–1.26)	.783
SSRI/SNRI (parent)	138	1.06 (0.87–1.30)	.542	0.98 (0.76–1.28)	.902
Antipsychotic (child)	106	1.74 (1.39–2.17)***	< .001	1.33 (1.03–1.73)	.032
Antipsychotic (parent)	106	1.86 (1.49–2.31)***	< .001	1.26 (0.96–1.65)	.094
Stimulant (child)	129	1.19 (0.96–1.47)	.126	1.00 (0.76–1.33)	.979
Stimulant (parent)	129	1.29 (1.05–1.59)*	.024	1.00 (0.74–1.36)	.975
<i>Note.</i> General (total) irritability. Comorbidity burden was modelled as a linear mixed-effects model (irritability = burden × informant + age + gender; burden coded ordinally 0, 1, 2, and 3 for three or more diagnoses); the burden × informant interaction was non-significant (p values shown). Clinical severity and diagnosis-count associations are Spearman correlations with FDR-corrected q values (child and parent separately). Medication associations are per-SD Firth penalized logistic regressions: the base model adjusts for age and gender (q = FDR within the 8-coefficient medication family), and the adjusted model additionally controls for CGI-S severity, number of diagnoses, and class-specific indication, with nominal adjusted p values. After adjustment, the any-medication, SSRI/SNRI, and stimulant associations were null and the antipsychotic association was attenuated, remaining nominally significant only for child report. Medication categories were not mutually exclusive (polypharmacy). * q < .05; ** q < .01; *** q < .001.					

Table 4

Figure 3

Within internalizing psychopathology, children with anxiety alone did not differ from children with neither internalizing disorder ($q = .146$), whereas those with depression alone and those with both disorders were elevated (both $q < .001$; $\epsilon^2 = .08$).

3.5. Parent–Child Agreement and the Contribution of Each Informant

Parent–child agreement was moderate ($\text{ICC}(2,1) = .65$, 95% CI [.59, .69]). Mean bias was negligible ($-.04$), but the limits of agreement were wide, from -1.62 to 1.54 , with no proportional bias across the range ($p = .34$; Supplementary Figure S2). Agreement varied by diagnosis (Table 2), from $.38$ in ODD to $.70$ in CD ($n = 26$).

The direction of the discrepancy also varied by diagnosis (Fig. 2B; Supplementary Table S7). Children with depressive and anxiety disorders reported more irritability than their parents reported for them ($b = -.33$ and $-.24$, both $q < .01$), whereas parents of children with ODD, CD, ADHD, and SLD reported more irritability than the children reported for themselves ($b = +.16$ to $+.39$). Among adolescents, higher self-reported hostility was associated with a more child-elevated discrepancy, with dyad-mean irritability, depressive symptoms, illness severity, diagnosis, age, and gender controlled ($b = -.39$, $p < .001$; VIFs ≤ 3.08 ; Supplementary Table S8). Self-reported hostility correlated $\rho = .77$ with the child's irritability rating and $\rho = .57$ with the parent's.

The incremental contribution of each informant depended on the diagnosis (Table 2; Supplementary Table S9). For depressive disorders, only the child report retained an association when both were entered together ($aOR = 1.77$ vs 1.04) and only the child report improved model fit ($q < .001$). For ODD and CD, only the parent report did so ($aOR = 2.40$ and 3.42 ; both $q < .001$). For ADHD, both informants contributed independently (parent $q = .002$, child $q = .038$). For anxiety disorders each report improved model fit when added to the other (parent adds $q = .012$; child adds $q = .019$), but the two coefficients ran in opposite directions (child 1.24 , parent 0.77), so this is incremental model improvement under suppression rather than two independent positive contributions; SLD produced coefficients in opposite directions (0.76 vs 1.55); and for OCD neither report improved fit (both $q = .058$).

3.6. Tonic and Phasic Irritability

Secondary analyses examined the two subscales separately. Across the mutually exclusive groups (Fig. 4), tonic scores were highest in depressive disorders and CD, and phasic scores highest in CD, followed by depressive disorders and ODD; the ordering of the intermediate groups differed little between the two panels.

Against dimensional symptom measures the two facets behaved differently, and the differences held when tested directly rather than read from the size of the coefficients (Supplementary Table S10). Tonic irritability correlated more strongly than phasic irritability with SDQ emotional symptoms (child $\rho = .53$ vs $.42$, $t = 4.52$; parent $.53$ vs $.43$, $t = 4.25$) and with self-reported depressive symptoms on the BSI ($.54$ vs $.45$, $t = 3.47$), whereas phasic irritability correlated more strongly with SDQ conduct problems (child $.55$ vs $.64$, $t = -3.77$; parent $.58$ vs $.66$, $t = -3.52$); all five contrasts survived FDR correction ($q \leq .002$). The facets did not differ in their correlations with SDQ hyperactivity or peer problems, with the ARI, or with clinician-rated severity (all $q > .30$).

Against categorical diagnoses the picture was less stable. Entered alone, each subscale reproduced the pattern of the general score; entered together, 8 of the 28 coefficients changed sign and the mutually adjusted estimates diverged sharply from the single-predictor ones (Supplementary Table S11, Supplementary Figure S3). On child report for CD, for instance, tonic irritability was associated with the diagnosis when entered alone ($aOR = 1.74$) but fell below

unity when phasic irritability was entered alongside it (aOR = 0.61, phasic aOR = 3.44). None of the 14 tonic-versus-phasic contrasts survived FDR correction (smallest $q = .058$), so differences between the facets at the level of individual diagnoses cannot be separated from the instability their collinearity produces.

Splitting the general score into its two components lowered the AIC in every diagnostic model, by between 0.6 and 8.4 points, and exceeded the conventional two-point benchmark in 10 of the 14. The fit gain was modest and variable in magnitude, and it came at the cost of unstable subscale coefficients: as observed scores the two subscales correlated .75 for both informants, and subscale-specific reliability after accounting for the general factor was low ($\omega_{HS} \leq .25$).

Figure 4

3.7. Medication Correlates

Medication was recorded as prescribed in routine care, and these cross-sectional associations were treated as exploratory (Table 4). Irritability was higher in children taking antipsychotics on both reports and marginally higher in those taking any psychotropic medication. After adjustment for illness severity, diagnosis count, and the indication for each class, the any-medication and stimulant associations were null, and the antipsychotic association was attenuated, remaining nominally significant only on child report (aOR = 1.33, $p = .032$).

3.8. Psychometric Properties of the TAPIS-TR

Measurement properties are summarised here and reported in full in the supplement (Table 5). Content validity was excellent: every item met the I-CVI standard on all four criteria, the lowest single value was .909, and both forms reached S-CVI/Ave $\approx .99$ (Supplementary Tables S12 and S13). Parallel analysis supported two factors for both informants (Supplementary Figure S4). A single-factor solution fitted the item pool less well than the hypothesised correlated two-factor solution for both informants (CFI .96 vs .98 and RMSEA .087 vs .068 on child report; .96 vs .98 and .098 vs .073 on parent report; robust scaled $\Delta\chi^2(1) = 39.0$ and 53.8, both $p < .001$), and the two-factor solution fitted well (child CFI = .98, RMSEA = .068, SRMR = .056; parent CFI = .98, RMSEA = .073, SRMR = .058) with uniformly high loadings (Supplementary Table S14).

Table 5
Summary psychometric properties of the general score and tonic/phasic subscales (N = 542)

Index	Child	Parent
Internal consistency		
General score α / ω	.94 / .94	.95 / .95
Tonic / phasic α	.90 / .91	.90 / .93
Tonic / phasic composite ω	.92 / .93	.92 / .94
Test-retest ICC(2,1)		
General score (child n = 59, parent n = 59)	.64 (.42–.78)	.73 (.57–.83)
General score SEM / MDC95	.56 / 1.54	.50 / 1.39
Parent-child agreement, general (dyadic)	.65 (.59–.69), across informants	
Convergent validity (ARI, ρ)	.83	.81
Dimensionality		
Factor correlation (CFA, WLSMV)	.86	.85
Factor correlation (target-rotated ESEM, WLSMV)	.74	.74
AVE tonic / phasic	.52 / .50	.54 / .56
HTMT ratio	.83	.81
Bifactor ECV / ω H	.80 / .86	.80 / .88
ω HS tonic / phasic	.03 / .25	.17 / .15

Note. General score = mean of all 19 items; tonic = items 1–8; phasic = items 9–19; all scored 0–4. α = Cronbach's alpha; ω = McDonald's omega (composite reliability). ICC(2,1) = two-way random-effects, absolute-agreement, single-measures intraclass correlation (child retest n = 59, parent retest n = 59; 30-day interval). SEM = standard error of measurement; MDC95 = minimal detectable change at 95% confidence. Parent-child agreement is the dyadic inter-informant ICC(2,1) for the general score (N = 542). ARI = Affective Reactivity Index (convergent validity, Spearman ρ). WLSMV = weighted least squares mean- and variance-adjusted estimator; AVE = average variance extracted; HTMT = heterotrait-monotrait ratio; ECV = explained common variance; ω H = omega hierarchical (general factor); ω HS = omega hierarchical subscale (subscale-specific reliability). Bifactor indices are reported as a decomposition of general and specific variance; the comparison of one-factor, correlated two-factor and target-rotated ESEM solutions, together with the cross-loadings freed under rotation, is given in Supplementary Table S15. HTMT should be read alongside the latent factor correlations and ω HS rather than on its own. The two factor correlations come from the same items and estimator and differ only in specification: the confirmatory model fixes every cross-loading to zero, whereas the target-rotated ESEM leaves each item's loading on its non-assigned factor free.

In the confirmatory solution, which fixes every cross-loading to zero, the two factors correlated .86 (child) and .85 (parent). Permitting small cross-loadings under target rotation reduced that correlation to .74 for both informants while the primary loadings remained coherent and the model fitted better on every index (Supplementary Table S15). The cross-loadings were few and interpretable: three items on the child form and two on the parent form exceeded |.30|, all of them phasic items describing the onset or intensity of anger rather than its behavioural consequences – being easily enraged (child .49, parent .40) and having frequent temper outbursts (.37, .39) – together with one item on physical aggression that loaded negatively on the tonic factor in child report (–.34). No tonic item exceeded |.15| on the phasic factor in child report or |.23| in parent report. Part of the association between the facets in the

confirmatory model is therefore attributable to the restriction that no item may load on both, indicating that the facets are less strongly correlated than the strict CFA alone implies.

A bifactor decomposition nonetheless placed 80% of the common variance on a single general dimension ($\omega_H = .86-.88$), leaving little reliable variance in the specific factors ($\omega_{HS} \leq .25$; Supplementary Table S16). The general score was used as the primary index throughout, and the subscales were treated as facets nested within it rather than as independent scales.

Table 5

Internal consistency of the general score was excellent ($\alpha = .94$ child, $.95$ parent) and convergence with the ARI was strong ($\rho = .83$ child, $.81$ parent). Test–retest reliability over a median interval of 30 days was acceptable for the child form and good for the parent form ($n = 59$ for both; $ICC = .64$ [.42, .78] and $.73$ [.57, .83]; Supplementary Table S17), with corresponding MDC95 values of 1.54 and 1.39 on the 0–4 metric. Several mean scores also declined significantly across the retest interval, which was itself variable (7–70 days), while participants remained in care, so these estimates reflect both measurement stability and possible clinical change. HTMT ratios between the tonic and phasic facets were $.83$ (child) and $.81$ (parent), below the conventional $.85$ threshold, although at latent correlations of this magnitude the ratio is better read alongside the factor correlations and ω_{HS} than on its own. Configural fit was good across age, gender, and current medication status ($CFI .980-.982$, $RMSEA .063-.069$), and constraining the loadings produced changes in fit within conventional cutoffs, although for child-report age the margin was narrow ($\Delta CFI = -.010$, $\Delta RMSEA = .011$); scalar invariance was not evaluated (Supplementary Table S18).

4. DISCUSSION

This study examined the validity and reliability of the TAPIS-TR in children and adolescents, while also exploring the clinical correlates of irritability and its relationship with concurrent psychiatric diagnoses. Content validity was excellent, the two-factor structure identified in the original scale was reproduced in both forms and fitted better than a single-factor solution, and internal consistency was high ($\alpha = .94-.95$). It also converged strongly with the ARI ($\rho = .81-.83$), while HTMT ratios for the tonic and phasic dimensions stayed below the conventional threshold. The facets were more separable than a confirmatory model alone suggested, their correlation falling from $.85-.86$ to $.74$ once small cross-loadings were permitted; even so, the residual overlap remains substantial and ω_{HS} was low, so the facets are not reliable as independent scores once their shared general variance is removed. Test–retest reliability was acceptable for the child form and good for the parent form, and metric invariance was supported across age, gender, and medication status within conventional cutoffs; scalar invariance was not tested. Taken together, these findings support the TAPIS-TR as a reliable measure of general irritability whose two facets are reproducible and differentially correlated, while suggesting caution in interpreting tonic and phasic subscales as independent clinical scores.

Beyond its psychometric properties, irritability itself was elevated throughout the clinic rather than being confined to a particular diagnosis, and the pattern across disorders was one of degree rather than kind. Depressive disorders, ADHD, ODD, and CD were each associated with higher irritability independent of comorbidity, whereas anxiety disorders, OCD, and SLD carried no such association on either report. Irritability increased with the number of concurrent diagnoses and with clinician-rated severity, and was highest in children whose difficulties cut across more than one diagnostic cluster. Parent and child reports agreed only moderately, and the degree of agreement itself depended on diagnosis. Tonic and phasic dimensions overlapped substantially and retained little reliable variance of their own once the general factor was accounted for, yet they were more separable than a restrictive confirmatory model implied and their external correlates diverged reliably, tonic and phasic irritability tracking emotional and

behavioural symptoms respectively, and this may matter for clinical interpretation in ways the factor structure alone would not capture.

The clinical findings rest on an instrument that performed well: content validity, established beforehand with a large expert panel, and internal consistency were excellent, the intended two-factor structure was supported and fitted well for both informants, and convergence with the ARI was strong [17]. Metric invariance was supported across gender, age group, and current medication status within conventional cutoffs, though the margin was narrow for child-report age and scalar invariance was not evaluated. Since measurement equivalence cannot be assumed from translation alone [21], it is notable that the tonic–phasic configuration of the source instrument held in Turkish [13]. Floor effects were strongest for the two items marking the most severe behavioural consequences of temper, but they were not confined to them: several other severe phasic items also drew floor endorsement above 50%, which is what one would expect of behaviours that are uncommon even in a referred sample and which limits the information those items carry at the lower end of the scale.

Temporal stability over a median one-month interval was acceptable for the child form and good for the parent form, but the minimal detectable change was wide relative to the range of the scale. This is stated deliberately: the TAPIS-TR suits between-group comparison, cross-sectional characterisation and research, but individual change smaller than a substantial shift cannot be distinguished from measurement noise, so it should not yet be used alone to judge treatment response in a single patient. Part of that width is probably intrinsic to a construct that fluctuates with triggers and mood state as well as reflecting a trait-like propensity [38, 39].

Irritability increased almost linearly with the number of concurrent diagnostic categories on both reports, peaked where diagnoses crossed clusters, and correlated moderately with clinician-rated severity. It therefore indexes the overall weight of psychopathology a child carries, as expected of a phenotype loading on a general psychopathology dimension [39–41]. Clinically, a high score should prompt a search for accumulated comorbidity and impairment rather than for a specific diagnosis; methodologically, relating irritability to a single disorder without holding comorbidity constant risks attributing to that disorder what belongs to the burden accompanying it. The correlation with severity was moderate rather than strong, however, so irritability is not reducible to global severity but carries its own information about the child’s frustration threshold and the family’s daily experience [4, 6]. The negative findings carry equal weight, since in a clinical sample they are the more informative half of the result. Anxiety, OCD, and SLD were not distinguished by irritability, and the associations appearing only in the comorbidity-adjusted models were nominal and not robust across informants. Irritability here is therefore not merely a proxy for being unwell: it discriminates, if coarsely, between conditions in which frustration proneness is central and those in which it is not, which also supports the construct validity of the Turkish forms. CD produced the largest estimates but was the smallest group, with the widest intervals, so that association is directionally consistent rather than precise.

The transdiagnostic case rests largely on community and longitudinal cohorts, where irritability is common, comorbidity uncommon and diagnostic groups reasonably clean [2, 5, 12]. Asked among children already impaired and mostly carrying more than one diagnosis, the same question produced a continuum: mutually exclusive groups overlapped substantially, and although the disruptive-behaviour and depressive groups occupied the upper end and undiagnosed children the lower end, the intermediate groups were closely spaced. This qualifies the intuition, still common in practice, that pronounced irritability points towards a particular diagnosis. Associations with internalizing and externalizing conditions alike survived adjustment for comorbidity, so irritability bridges the two spectra rather than belonging to either — as expected from the irritable dimension of oppositionality predicting emotional disorders while the headstrong and hurtful dimensions predict behavioural ones [42, 43], and from reviews positioning it as a shared regulatory difficulty [3, 39]. Agreement between parents and children was moderate [44], but its structure

mattered more than its level: mean bias was negligible while the limits of agreement were wide, so individual dyads can diverge substantially and one informant cannot stand in for the other. The direction of divergence was systematic. Children with depressive and anxiety disorders described more irritability than their parents observed, whereas parents of children with ODD, CD, ADHD, and SLD described more than the children did; entered together, only the child's account retained an association with depressive disorders, only the parent's with ODD and CD, and both contributed independently for ADHD. For anxiety disorders the joint model produced coefficients in opposing directions, consistent with suppression, so the two reports capture partly different information rather than parental report being protective. Agreement was lowest in ODD, the very condition in which the parent's report proved the informative one.

This matches the expectation that young people have privileged access to persistent inner tension while parents observe outbursts and their consequences [7, 18], and is consistent with treating discrepancy as potentially informative rather than as pure error [44]: in this sample, parent-only assessment captured less of the depression association, whereas child-only assessment captured less of the disruptive-behaviour association. Adolescents reporting higher hostility also rated their own irritability higher relative to their parents, with dyad-mean irritability, depressive symptoms, severity, diagnosis, age, and gender controlled; because both came from the same informant with overlapping content, this finding is hypothesis-generating pending replication with an independent index of hostility.

Current models distinguish persistent irritable mood from discrete temper outbursts, with partly separate correlates reported longitudinally [14–16]. Both halves of that claim found support here, but at different levels of analysis. Structurally, the two-factor configuration fitted better than a single dimension, and the facets proved more distinct than the confirmatory model implied: the cross-loadings that a strict specification suppresses fell on precisely the phasic items describing the onset and intensity of anger, and freeing them lowered the factor correlation from .85–.86 to .74 while improving fit. Against dimensional symptom measures the facets diverged as theory predicts, tonic irritability tracking emotional symptoms and phasic irritability conduct problems, and these contrasts were statistically reliable rather than merely numerical. Against categorical diagnoses, however, the two competed for the same variance: subscale-specific reliability was negligible, several subscale-level associations reversed direction under mutual adjustment, and no tonic-versus-phasic contrast survived correction – consistent with substantial collinearity and limited unique score variance rather than with opposite substantive effects, and in keeping with the development study and wider factor-analytic evidence [13, 16]. The most defensible reading is therefore neither that the facets are independent traits nor that the distinction is empty, but that they are heavily overlapping facets of a broader irritability liability which still carry somewhat different phenomenological information. The general score should be the primary clinical index, the facets should be read descriptively rather than scored as competing predictors, and whether they predict course or treatment response differently requires longitudinal data [14, 15].

Girls rated themselves as more irritable than boys while parents reported no corresponding difference, and self-reported irritability rose with age while parent-reported irritability did not. Both patterns invite an informant-by-group reading, but the informants were not contrasted in a single model and scalar invariance was untested, so differences in latent means cannot be separated from differences in item functioning or reporting threshold [44]; they are descriptive and require an explicit informant-by-gender model. Irritability was higher in children prescribed antipsychotics, attenuating after adjustment for severity, diagnosis count and indication; since prescribing was as delivered in routine care, this almost certainly reflects channelling rather than a medication effect.

Strengths and Limitations

The strengths of the study include a large, consecutively recruited clinical sample from public hospitals in different regions of Türkiye, diagnoses established with a semi-structured gold-standard interview [22, 23], simultaneous validation of both forms in the same dyads, and an analytic approach designed for the structure of clinical data. To our knowledge this is the first psychometric evaluation of both TAPIS forms in Turkish and the first in a large, diagnostically mixed clinical population.

Several limitations qualify the findings. The design is cross-sectional, so no causal inference is warranted and the prognostic questions motivating much of this literature [8–11] cannot be addressed. Without a healthy comparison group, diagnostic accuracy could not be assessed and no cut-off proposed; scalar invariance was not evaluated, limiting mean comparisons across gender and age. Recruitment spanned seven centres, and although a sensitivity model with centre fixed effects left the diagnostic associations unchanged, the smallest centre contributed 25 children, so centre-specific patterns within the rarer diagnoses cannot be excluded. Because K-SADS diagnoses integrated child and parent information, the criterion itself was not independent of either informant, and the apparent contribution of each TAPIS informant should therefore be understood as incremental clinical information rather than incremental validity established against a fully independent criterion; some shared-method contribution is possible. The retest subsample was modest, not randomly selected, assessed over a variable interval, and followed while participants continued to receive care, so the stability estimates incorporate real clinical change as well as measurement error. CD was represented by few children, so estimates there are imprecise, and the child forms, which require reading comprehension, were completed by children as young as school-entry age without formal comprehension testing. Over half the sample was medicated at assessment, and the sample is a referred one, so the findings should not be generalised to community populations.

A further limitation concerns DMDD, the one diagnosis in which irritability is itself the defining feature. Six children meeting criteria for it presented during the recruitment period and none entered the analytic sample: two attended with a grandparent rather than a parent, so no parent report could be obtained; one child declined to complete the forms; and three were too agitated at assessment to finish them. Since analyses were complete-case and participants with substantial missing data were excluded, the same consideration applies more broadly – the sample probably under-represents the most severely and acutely irritable children, and the findings characterise irritability across the ordinary case mix of outpatient practice rather than at its extreme. How the TAPIS-TR behaves in this disorder therefore remains to be established.

Priorities for further work follow directly: examination of the Turkish forms in community samples, testing of scalar invariance, a design including non-referred participants to establish diagnostic thresholds, and longitudinal data on sensitivity to change and on whether the facets predict course and treatment response differently. Characterising the scale in DMDD will require deliberate over-sampling and an assessment procedure adapted to children who cannot complete a standard battery in one session. Comparison with the clinician-rated Turkish irritability measure already available [19] and with Turkish adaptation work on the Affective Reactivity Index [29] would locate the TAPIS-TR among instruments available to Turkish clinicians [20, 27, 28].

5. CONCLUSION

In a consecutively referred clinical population, irritability was a graded transdiagnostic dimension tracking comorbidity burden and illness severity more closely than any single diagnosis, while still discriminating conditions in which frustration proneness is central from those in which it is not; its two theoretical facets were reproducible and differentially correlated yet supported primarily as facets within a dominant general irritability dimension rather than as independent clinical scores, and parent and child accounts were complementary rather than interchangeable. The

TAPIS-TR showed strong reliability and convergent and structural validity in this clinical sample, suitable for clinical characterisation and research though not yet for tracking individual change or for diagnostic classification, and a TAPIS-TR profile is best read as an indicator of overall psychiatric burden obtained from both informants.

Declarations

Ethics approval

This study was approved by the Non-Interventional Clinical Research Ethics Committee of Uşak University (Approval No. 1235-1235-14; April 28, 2026) and was conducted in accordance with institutional and national regulations and the Declaration of Helsinki.

Consent to participate

Written informed consent was obtained from a parent or legal guardian, and assent was obtained from each child or adolescent.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Competing interests

The authors declare that they have no competing interests.

Data availability

The datasets analysed in the current study are available from the corresponding author upon reasonable request.

Consent to publish

Not applicable. No individually identifying information is reported.

Author contributions

All authors contributed to the conceptualization of the study. M.E.Ö. and H.G.O. developed the methodology, led the formal analysis, and supervised the research process. All authors contributed to data collection. D.K.E., N.Ö., M.E.Ö., and Ö.G.Ö. prepared the initial draft of the manuscript. H.G.O. and N.Ö. reviewed and edited the manuscript and provided critical feedback on earlier versions. J.I.S. supervised the psychometric and analytic strategy, advised on the structural and differential-validity analyses, and critically revised the manuscript. All authors read and approved the final manuscript.

Additional Declarations

All authors have completed and submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest.

References

1. Brotman MA, Kircanski K, Leibenluft E (2017) Irritability in children and adolescents. *Annu Rev Clin Psychol* 13:317–341. <https://doi.org/10.1146/annurev-clinpsy-032816-044941>

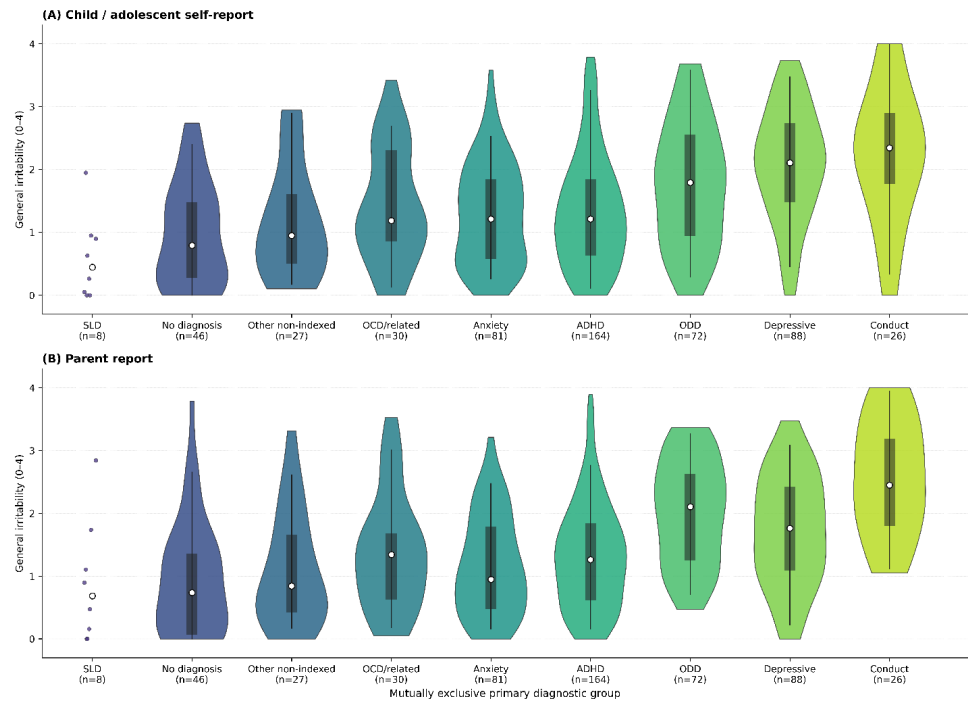
2. Vidal-Ribas P, Brotman MA, Valdivieso I, Leibenluft E, Stringaris A (2016) The status of irritability in psychiatry: A conceptual and quantitative review. *J Am Acad Child Adolesc Psychiatry* 55:556–570.
<https://doi.org/10.1016/j.jaac.2016.04.014>
3. Leibenluft E, Allen LE, Althoff RR, Brotman MA, Burke JD, Carlson GA, Dickstein DP, Dougherty LR, Evans SC, Kircanski K, Klein DN, Malone EP, Mazefsky CA, Nigg J, Perlman SB, Pine DS, Roy AK, Salum GA, Shakeshaft A, Stringaris A (2024) Irritability in youths: A critical integrative review. *Am J Psychiatry* 181:275–290.
<https://doi.org/10.1176/appi.ajp.20230256>
4. Toohey MJ, DiGiuseppe R (2017) Defining and measuring irritability: Construct clarification and differentiation. *Clin Psychol Rev* 53:93–108. <https://doi.org/10.1016/j.cpr.2017.01.009>
5. Copeland WE, Brotman MA, Costello EJ (2015) Normative irritability in youth: Developmental findings from the Great Smoky Mountains Study. *J Am Acad Child Adolesc Psychiatry* 54:635–642.
<https://doi.org/10.1016/j.jaac.2015.05.008>
6. Saatchi B, Olshansky EF, Fortier MA (2023) Irritability: A concept analysis. *Int J Ment Health Nurs* 32:1193–1210.
<https://doi.org/10.1111/inm.13140>
7. Klein DN, Dougherty LR, Kessel EM, Silver J, Carlson GA (2021) A transdiagnostic perspective on youth irritability. *Curr Dir Psychol Sci* 30:437–443. <https://doi.org/10.1177/09637214211035101>
8. Finlay-Jones AL, Ang JE, Brook J, Lucas JD, MacNeill LA, Mancini VO, Kottampally K, Elliott C, Smith JD, Wakschlag LS (2024) Systematic review and meta-analysis: Early irritability as a transdiagnostic neurodevelopmental vulnerability to later mental health problems. *J Am Acad Child Adolesc Psychiatry* 63:184–215. <https://doi.org/10.1016/j.jaac.2023.01.018>
9. Srinivasan R, Flouri E, Lewis G, Solmi F, Stringaris A, Lewis G (2024) Changes in early childhood irritability and its association with depressive symptoms and self-harm during adolescence in a nationally representative United Kingdom birth cohort. *J Am Acad Child Adolesc Psychiatry* 63:39–51.
<https://doi.org/10.1016/j.jaac.2023.05.027>
10. Riglin L, Eyre O, Thapar AK, Stringaris A, Leibenluft E, Pine DS, Tilling K, Davey Smith G, O'Donovan MC, Thapar A (2019) Identifying novel types of irritability using a developmental genetic approach. *Am J Psychiatry* 176:635–642. <https://doi.org/10.1176/appi.ajp.2019.18101134>
11. Yu Q, Hodgdon EA, Kryza-Lacombe M, Osuna L, Bozzetto LE, Ciro D, Wakschlag LS, Wiggins JL (2023) Roads diverged: Developmental trajectories of irritability from toddlerhood through adolescence. *J Am Acad Child Adolesc Psychiatry* 62:457–471. <https://doi.org/10.1016/j.jaac.2022.07.849>
12. Chin M, Robson DA, Woodbridge H, Hawes DJ (2025) Irritability as a transdiagnostic construct across childhood and adolescence: A systematic review and meta-analysis. *Clin Child Fam Psychol Rev* 28:101–124.
<https://doi.org/10.1007/s10567-024-00512-4>
13. Silver J, Carlson GA, Farquharson WH, Atlas J, Klein DN (2025) Development and validation of a multi-informant scale for assessing youth tonic and phasic irritability. *Res Child Adolesc Psychopathol* 53:1281–1296.
<https://doi.org/10.1007/s10802-025-01318-w>
14. Silver J, Carlson GA, Olino TM, Perlman G, Mackin D, Kotov R, Klein DN (2021) Differential outcomes of tonic and phasic irritability in adolescent girls. *J Child Psychol Psychiatry* 62:1220–1227.
<https://doi.org/10.1111/jcpp.13402>
15. Silver J, Mackin DM, Bufferd SJ, Dougherty LR, Goldstein BL, Carlson GA, Klein DN (2023) Tonic and phasic irritability in 6-year-old children: Differential correlates and outcomes. *J Child Psychol Psychiatry* 64:234–243.
<https://doi.org/10.1111/jcpp.13688>

16. Sorcher LK, Silver J, Chad-Friedman E, Carlson GA, Klein DN, Dougherty LR (2024) Early predictors and concurrent correlates of tonic and phasic irritability in adolescence. *Res Child Adolesc Psychopathol* 52:1105–1117. <https://doi.org/10.1007/s10802-024-01185-x>
17. Stringaris A, Goodman R, Ferdinando S, Razdan V, Muhrer E, Leibenluft E, Brotman MA (2012) The Affective Reactivity Index: A concise irritability scale for clinical and research settings. *J Child Psychol Psychiatry* 53:1109–1117. <https://doi.org/10.1111/j.1469-7610.2012.02561.x>
18. Zik J, Deveney CM, Ellingson JM, Haller SP, Kircanski K, Cardinale EM, Brotman MA, Stoddard J (2022) Understanding irritability in relation to anger, aggression, and informant in a pediatric clinical population. *J Am Acad Child Adolesc Psychiatry* 61:711–720. <https://doi.org/10.1016/j.jaac.2021.08.012>
19. Turan S, Ermiş Ç, Eray Ş, Yavuz BE, Uzman S, Ozbek MM, Tunçtürk M, Çıray RO, İnal N (2022) Psychometric properties of the Clinician Affective Reactivity Index for assessment of irritability in a clinical sample of Turkish children and adolescents. *Scand J Child Adolesc Psychiatr Psychol* 10:24–32. <https://doi.org/10.2478/sjcapp-2022-0002>
20. Osada M, Mori H, Takahashi M, Shinkawa H, Adachi M, Adachi M, Saito T, Nakamura K (2025) Psychometric properties and cross-cultural adaptation of the Japanese version of the Affective Reactivity Index: A community-based study of adolescent irritability. *Psychiatry Clin Neurosci Rep* 4:e70169. <https://doi.org/10.1002/pcn5.70169>
21. Stevanovic D, Jafari P, Knez R, Franic T, Atilola O, Davidovic N, Bagheri Z, Lakic A (2017) Can we really use available scales for child and adolescent psychopathology across cultures? A systematic review of cross-cultural measurement invariance data. *Transcult Psychiatry* 54:125–152. <https://doi.org/10.1177/1363461516689215>
22. 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* 36:980–988. <https://doi.org/10.1097/00004583-199707000-00021>
23. Ünal F, Öktem F, Çetin Çuhadaroğlu F, Çengel Kültür SE, Akdemir D, Foto Özdemir D, Çak HT, Ünal D, Tıraş K, Aslan C, Kalaycı BM, Aydos BS, Kütük F, Taşyürek E, Karaokur R, Karabucak B, Karakök B, Karaer Y, Artık A (2019) Okul çağı çocukları için duygulanım bozuklukları ve şizofreni görüşme çizelgesi-şimdi ve yaşam boyu şekli-DSM-5 Kasım 2016-Türkçe uyarlamasının (ÇDŞG-ŞY-DSM-5-T) geçerlik ve güvenirliği [Reliability and validity of the Schedule for Affective Disorders and Schizophrenia for School-Age Children—Present and Lifetime Version, DSM-5 November 2016—Turkish adaptation (K-SADS-PL-DSM-5-T)]. *Turk Psikiyatri Derg* 30:42–50. <https://doi.org/10.5080/u23408>
24. Rouquette A, Falissard B (2011) Sample size requirements for the internal validation of psychiatric scales. *Int J Methods Psychiatr Res* 20:235–249. <https://doi.org/10.1002/mpr.352>
25. Wolf EJ, Harrington KM, Clark SL, Miller MW (2013) Sample size requirements for structural equation models: An evaluation of power, bias, and solution propriety. *Educ Psychol Meas* 73:913–934. <https://doi.org/10.1177/0013164413495237>
26. Guy W (1976) Clinical global impressions. ECDEU assessment manual for psychopharmacology (Rev. U.S. Department of Health, Education, and Welfare, pp 217–222
27. DeSousa DA, Stringaris A, Leibenluft E, Koller SH, Manfro GG, Salum GA (2013) Cross-cultural adaptation and preliminary psychometric properties of the Affective Reactivity Index in Brazilian youth: Implications for DSM-5 measured irritability. *Trends Psychiatry Psychother* 35:171–180. <https://doi.org/10.1590/S2237-60892013000300004>

28. Mulraney MA, Melvin GA, Tonge BJ (2014) Psychometric properties of the Affective Reactivity Index in Australian adults and adolescents. *Psychol Assess* 26:148–155. <https://doi.org/10.1037/a0034891>
29. Kocael Ö (2015) Çocuk ve ergenlerde irritabilite: Duygusal Reaktivite İndeksi'nin Türkçe geçerlilik güvenilirlik çalışması [Irritability in children and adolescents: Turkish validity and reliability study of the Affective Reactivity Index] [Medical specialty thesis, Uludağ University]. Ulusal Tez Merkezi. <https://tez.yok.gov.tr/UlusalTezMerkezi>
30. Goodman R (1997) The Strengths and Difficulties Questionnaire: A research note. *J Child Psychol Psychiatry* 38:581–586. <https://doi.org/10.1111/j.1469-7610.1997.tb01545.x>
31. Güvenir T, Özbek A, Baykara B, Arkar H, Şentürk B, İncekaş S (2008) Güçler ve Güçlükler Anketi'nin (GGA) Türkçe uyarlamasının psikometrik özellikleri [Psychometric properties of the Turkish version of the Strengths and Difficulties Questionnaire (SDQ)]. *Çocuk ve Gençlik. Ruh Sağlığı Dergisi* 15:65–74
32. Rosenberg M (1965) *Society and the adolescent self-image*. Princeton University Press. <https://doi.org/10.1515/9781400876136>
33. Çuhadaroglu F (1986) Adölesanlarda benlik saygısı [Self-esteem in adolescents] [Unpublished medical specialty thesis]. Hacettepe University
34. Çeçen AR (2008) Üniversite öğrencilerinde yaşam doyumunu yordamada bireysel bütünlük (tutarlılık) duygusu, aile bütünlük duygusu ve benlik saygısı [Sense of coherence, family sense of coherence, and self-esteem in predicting life satisfaction among university students]. *Eğitimde Kuram ve Uygulama* 4:19–30
35. Derogatis LR (1993) *Brief Symptom Inventory (BSI): Administration, scoring, and procedures manual*, 4th edn. National Computer Systems
36. Şahin NH, Durak A (1994) Kısa Semptom Envanteri: Türk gençleri için uyarlanması [Brief Symptom Inventory: Adaptation for Turkish youth]. *Türk Psikoloji Dergisi* 9:44–56
37. Şahin NH, Durak Batıgün A, Uğurtaş S (2002) Kısa Semptom Envanteri (KSE): Ergenler için kullanımının geçerlik, güvenilirlik ve faktör yapısı [Validity, reliability, and factor structure of the Brief Symptom Inventory in adolescents]. *Türk Psikiyatri Derg* 13:125–135
38. Stringaris A (2011) Irritability in children and adolescents: A challenge for clinicians and researchers. *Eur Child Adolesc Psychiatry* 20:61–63. <https://doi.org/10.1007/s00787-011-0160-x>
39. Stringaris A, Vidal-Ribas P, Brotman MA, Leibenluft E (2018) Practitioner Review: Definition, recognition, and treatment challenges of irritability in young people. *J Child Psychol Psychiatry Allied Discip* 59:721–739
40. Caspi A, Moffitt TE (2018) All for one and one for all: Mental disorders in one dimension. *Am J Psychiatry* 175:831–844. <https://doi.org/10.1176/appi.ajp.2018.17121383>
41. Caspi A, Houts RM, Belsky DW, Goldman-Mellor SJ, Harrington H, Israel S, Meier MH, Ramrakha S, Shalev I, Poulton R, Moffitt TE (2014) The p factor: One general psychopathology factor in the structure of psychiatric disorders? *Clin Psychol Sci* 2:119–137. <https://doi.org/10.1177/2167702613497473>
42. Stringaris A, Goodman R (2009) Longitudinal outcome of youth oppositionality: irritable, headstrong, and hurtful behaviors have distinctive predictions. *J Am Acad Child Adolesc Psychiatry* 48:404–412
43. Stringaris A, Goodman R (2009) Three dimensions of oppositionality in youth. *J Child Psychol Psychiatry* 50:216–223. <https://doi.org/10.1111/j.1469-7610.2008.01989.x>
44. De Los Reyes A, Augenstein TM, Wang M, Thomas SA, Drabick DA, Burgers DE, Rabinowitz JA (2015) The validity of the multi-informant approach to assessing child and adolescent mental health. *Psychol Bull* 141:858–900. <https://doi.org/10.1037/a0038498>

Figures

Figure 1. General irritability across mutually exclusive primary diagnostic groups, by informant

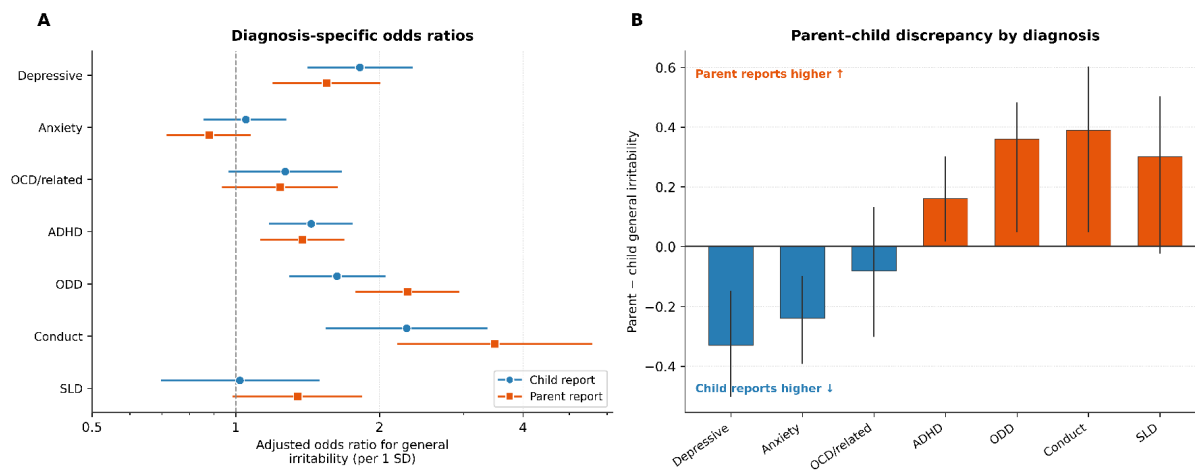


Note. Vertical violins of general irritability (0-4) for the nine mutually exclusive primary diagnostic groups, shown separately for (A) child/adolescent self-report and (B) parent report; each group is shown in a distinct colour and ordered by its child-report median; white dots = group medians, dark bars = interquartile ranges, whiskers = 5th-95th percentiles. Groups with $n < 10$ (SLD) are shown as individual points.

Figure 1

See image above for figure legend

Figure 2. Diagnosis-specific associations and parent-child discrepancy in general irritability

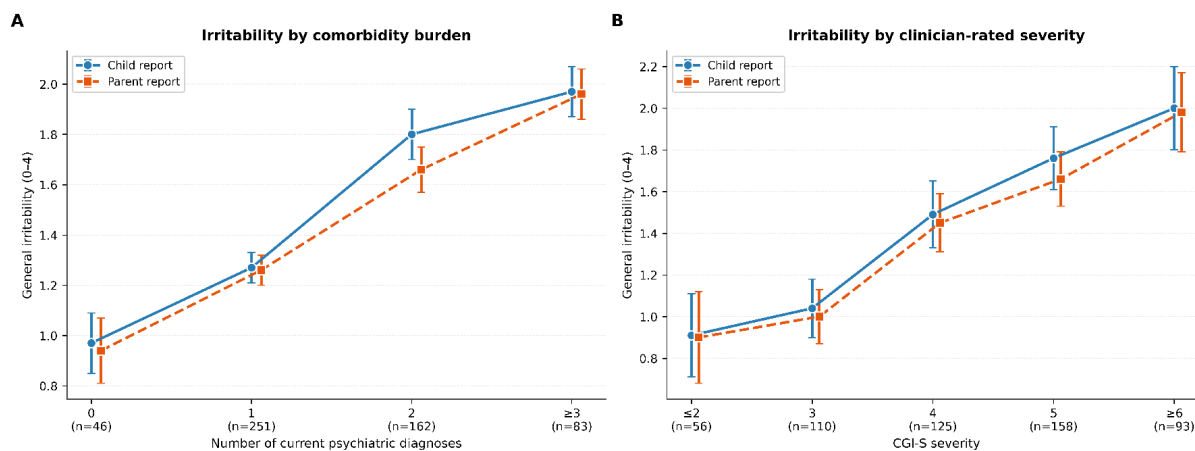


Note. (A) Adjusted odds ratios (per 1 SD of general irritability) for each diagnosis from Firth penalized logistic regressions controlling for age and gender; circles = child report, squares = parent report; horizontal lines = 95% CIs; dashed reference at OR = 1. (B) Parent-minus-child difference in mean general irritability by diagnosis; positive values (orange) = higher parent report, negative (blue) = higher child report; whiskers = 95% CIs.

Figure 2

See image above for figure legend

Figure 3. General irritability by comorbidity burden and clinician-rated severity

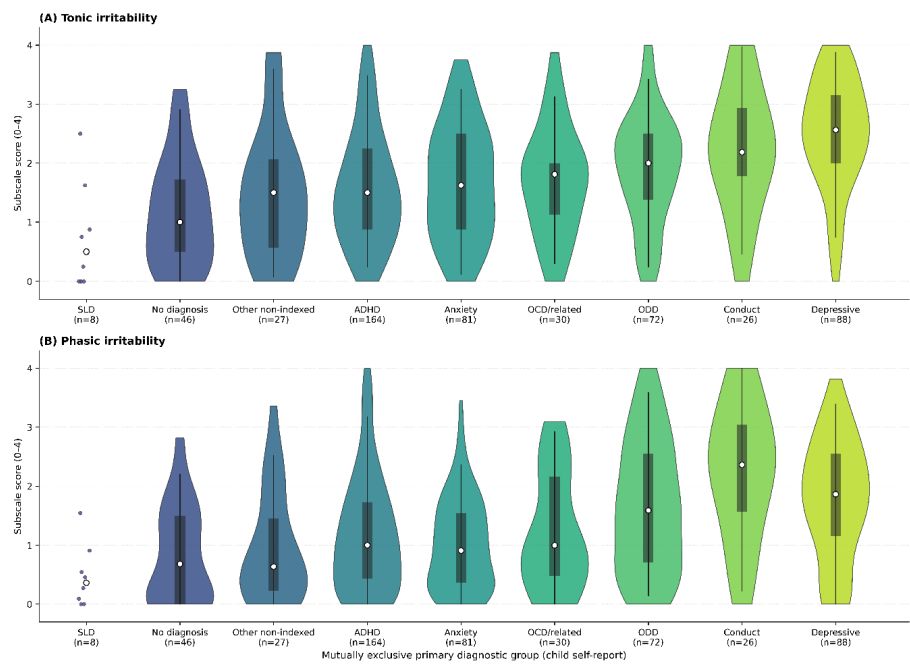


Note. (A) Mean general irritability across the number of current psychiatric diagnoses (0, 1, 2, ≥3) and (B) across CGI-S severity strata, for child report (blue circles) and parent report (orange squares); whiskers = 95% CIs; group sizes below each point.

Figure 3

See image above for figure legend

Figure 4. Tonic and phasic irritability across mutually exclusive primary diagnostic groups (child self-report)



Note. Vertical violins of the (A) tonic and (B) phasic subscale scores (0-4, child self-report) for the nine mutually exclusive primary diagnostic groups; each group is shown in a distinct colour and ordered by its subscale median; white dots = group medians, dark bars = interquartile ranges, whiskers = 5th-95th percentiles. Groups with n < 10 (SLD) are shown as individual points.

Figure 4

See image above for figure legend

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SUPPLEMENTARYMATERIALS.docx](#)