Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Development and Initial Evaluation of an Adolescent Depression Screening Tool: Integrating Factor Analytic, Network, and Clinical Evidence

Zane Gulbe

zane.gulbe@rsu.lv

Riga Stradiņš University

Viktorija Perepjolkina

Riga Stradiņš University

Jeļena Koļesņikova

Riga Stradiņš University

Ilona Krone

Riga Stradiņš University

Ņikita Bezborodovs

Riga Stradiņš University

Linda Nauzere

Riga Stradiņš University

Ainārs Stepens

Riga Stradiņš University

Research Article

Keywords: adolescent depression, depression screening, instrument development, short form, psychometric evaluation, symptom network analysis, classification uncertainty, screening thresholds

Posted Date: November 2nd, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11291346/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Authors: Gulbe, Z. [1]***, Perepjolkina, V.*** [1]***, Kolesnikova, J.*** [1,2]***, Krone, I.*** [1,2]***, Bezborodovs, N.*** [3,4]***, Nauzere. L.*** [2]***, Stepens, A.*** [5]

[1] *Department of Psychology, Rīga Stradiņš University, Riga, Latvia*

[2] *Laboratory of Psychology, Rīga Stradiņš University, Riga, Latvia*

[3] *Department of Psychiatry and Narcology, Rīga Stradiņš University, Riga, Latvia*

[4] *Children's Clinical University Hospital, Child and Youth Mental Health Centre, Riga, Latvia*

[5] *Military Medicine Research and Study Centre, Rīga Stradiņš University, Riga, Latvia*

Corresponding author: Zane Gulbe (zane.gulbe@rsu.lv)

DEVELOPMENT AND INITIAL EVALUATION OF AN ADOLESCENT DEPRESSION SCREENING TOOL: INTEGRATING FACTOR ANALYTIC, NETWORK, AND CLINICAL EVIDENCE

Abstract

Background

Adolescent depression screening requires brief instruments that capture clinically relevant symptoms and acknowledge inconclusive results. This study developed and initially evaluated an ICD-11-anchored screening instrument comprising a full form and an abbreviated short form, integrating structural, network, and clinical evidence.

Methods

The development sample included 828 adolescents aged 13–19: 661 community respondents, 87 other clinical respondents, and 80 depression-symptom respondents. Item evaluation combined exploratory factor analysis of polychoric correlations, regularized symptom-network analysis, response distributions, symptom overlap, and clinical content review. Unweighted, provisionally weighted, and ridge-regression scoring were compared, including nested cross-validation for regression evaluation. Discrimination between the depression-symptom and each comparison group was assessed using receiver operating characteristic areas under the curve (AUCs) and paired bootstrap intervals. Uncertain intervals and common raw-score

boundaries were explored. Internal consistency, classification agreement, subgroup differences, and WHO-5 correlations were examined.

Results

The retained 23-item full form showed four related symptom domains. A 13-item short form was selected, although its one- and two-factor solutions retained substantial misfit. McDonald's omega was .960 for the full form and .935 for the short form. Short- and full-form AUCs were .848 and .846 against the other clinical group and .825 and .824 against the community group; paired difference intervals included zero. Weighting provided no consistent advantage across contrasts. Revised short-form bands were 0–18, 19–22, and 23–52; full-form bands were 0–31, 32–44, and 45–92. At the short-form lower boundary, sensitivity was 83.8% and specificity was 72.4% and 72.5% against the other clinical and community groups, respectively. Classification agreement was 90.2%. Sensitivity was lower among boys and particularly adolescents aged 13–14 in both forms. WHO-5 correlations were −.631 and −.649 for the short and full forms.

Conclusions

The proposed Adolescent Depressive Symptoms Screener offers a brief, simply scored form with promising initial discrimination. Its provisional bands support further evaluation of symptom enquiry and follow-up decisions rather than diagnosis. Development-sample optimism and subgroup limitations require independent validation before routine screening recommendations.

Keywords: *adolescent depression; depression screening; instrument development; short form; psychometric evaluation; symptom network analysis; classification uncertainty; screening thresholds.*

Introduction

The Clinical Challenge of Adolescent Depression in Help-Seeking Contexts

Depression during adolescence represents a major global public health concern, with approximately one in seven individuals aged 10–19 years experiencing a diagnosed mental health condition [1]. This escalating burden is acutely mirrored across the wider European region and within specific national contexts; for instance, population surveys in Northeastern and Central Europe, including Latvia, report that up to 38% of adolescents experience recurrent weekly depressive symptoms, accompanied by a documented two-fold increase in officially registered youth mood disorders over the past decade [2, 3]. Because young people frequently conceal emotional distress due to perceived stigma or limited mental health literacy, identifying at-risk youth remains difficult. This challenge is further compounded in clinical and educational settings by the substantial diagnostic difficulty of distinguishing normative developmental turbulence and emotional reactivity from clinically meaningful depressive pathology [4–6]. When distress goes undetected, unaddressed symptom cascades risk rapid functional decline and chronicity.

The clinical presentation of adolescent depressive pathology is markedly heterogeneous, manifesting across affective, cognitive, interpersonal, and somatic domains. Crucially, young people presenting to primary care or mental health services rarely frame their complaints in formal diagnostic terms. Instead, initial consultations are frequently driven by nonspecific difficulties – including sleep disruption, school-related concentration impairment, behavioral withdrawal, or diffuse somatic complaints – long before depressed mood is explicitly verbalized. Delayed identification of these symptom constellations carries substantial long-term risks, including educational disruption, interpersonal impairment, and increased suicide risk [7]. Consequently,

effective triage requires tools capable of capturing these masked, distributed presentations rather than relying solely on spontaneous endorsements of low mood.

In primary care and outpatient psychological settings, brief screening tools are essential for triaging adolescents whose distress warrants detailed clinical evaluation. According to the *Standards for Educational and Psychological Testing* [8] and contemporary test-development standards [9], instrument validation must explicitly define what is measured, the intended context of use, and the target population. The present instrument is developed specifically for adolescents aged 13–19 years presenting to primary care, school counselling, or outpatient psychological services with emotional, somatic, or behavioral complaints. Its intended function is to provide an interpretable initial screening evaluation – informing the urgency of follow-up and the necessity of a comprehensive diagnostic interview – rather than serving as an automated, stand-alone diagnostic threshold.

Operationalizing Symptom Expression: An ICD-11 Framework and Clinical Gradients

To achieve diagnostic and screening utility, item selection must be firmly anchored in modern nosological frameworks while capturing adolescent-specific symptom expression. In the International Classification of Diseases, Eleventh Revision (ICD-11), depressive episodes (mild, moderate, and severe) require core manifestations of depressed mood or marked anhedonia, accompanied by additional neurovegetative, cognitive, and psychomotor symptoms [10]. However, the psychometric weight of individual symptoms varies substantially between community detection and clinical discrimination.

Empirical evidence demonstrates that symptoms operate across distinct diagnostic gradients. In community-based adolescent studies, loss of interest or pleasure, fatigue, sleep disturbance, impaired concentration, and suicidal ideation serve as primary discriminators

differentiating depressed adolescents from non-depressed controls, whereas guilt, marked psychomotor changes, and appetite or weight fluctuations demonstrate clearer gradients with episode severity [11]. Furthermore, vegetative and physical complaints are exceptionally common among adolescents with major depression; despite their limited diagnostic specificity when evaluated in isolation, their inclusion is vital to prevent under-identification in medical and somatic presentations [12]. Interpersonal dynamics also provide indispensable diagnostic context: adolescent depressive presentations frequently involve social withdrawal and negative self-attributions, including perceptions of being a burden to others – a cognitive distortion strongly linked to suicidal ideation [13]. Finally, in accordance with clinical guidance, thoughts of death, passive death wishes, and explicit suicidal ideation must be evaluated independently of aggregated depression scores to ensure immediate patient safety [7].

Current Landscape of Adolescent Screening: Methodological Trade-offs

To situate current instrument development within the broader diagnostic landscape, a systematic synthesis of widely recognized screening and assessment measures across five methodological axes (clinical gold standards, functional impairment, internalizing differentiation, developmentally tailored constructs, and population surveillance) is provided in Table S1. A critical evaluation of these established instruments reveals distinct psychometric philosophies and operational trade-offs:

1. **Direct Diagnostic Criteria Adaptations:** Tools such as the Patient Health Questionnaire for Adolescents (PHQ-A / PHQ-9; [14] offer rapid administration aligned with classical criteria. However, because they map directly onto adult diagnostic manuals, they can overlook adolescent-specific symptom manifestations, such as irritable mood reactivity or somatic masks.

2. **Comprehensive Multidimensional Scales:** Measures like the Reynolds Adolescent Depression Scale, Second Edition (RADS-2); [15], the Children's Depression Inventory 2 (CDI-2); [16]), and the Beck Depression Inventory for Youth (BDI-Y); [17] capture broader developmental psychopathology (e.g., negative self-evaluation, cognitive distortions). Nonetheless, their length (typically 20–30 items) frequently induces respondent burden in busy primary care settings.

3. **Dynamic Screening Tools:** High-sensitivity measures like the Mood and Feelings Questionnaire (MFQ; 33 items, or the 13-item SMFQ) and epidemiological instruments like the CES-DC; [19] capture rapid affective fluctuations and population distress, yet they often exhibit elevated false-positive rates when applied across heterogeneous clinical samples.

A central limitation of the existing literature is that candidate screening instruments are predominantly validated against asymptomatic community controls. While such designs demonstrate proof-of-concept, they tend to yield artificially inflated classification accuracy (e.g., area under the ROC curve). In routine help-seeking practice, the primary clinical challenge is not merely separating depressed adolescents from thriving peers, but differentiating depressive pathology from distress arising from other psychiatric, behavioral, and neurodevelopmental conditions.

Integrating Factor Analytic and Psychometric Network Evidence

Classical scale development relies heavily on Classical Test Theory (CTT) and factor analysis to establish structural validity, internal consistency, and domain representation [20]. While factor analysis excels at identifying shared variance across dimensions, selecting items solely on

the basis of the highest factor loadings carries substantial methodological and clinical risks: it systematically promotes item collinearity (redundancy) and tends to discard low-frequency or neurovegetative complaints that, despite lower inter-item correlations, are clinically indispensable for medical triage and risk detection.

Psychometric network analysis offers a complementary descriptive architecture that directly addresses these structural limitations [2, 6]. Rather than conceptualizing depression merely as the passive reflective output of an underlying latent construct, the Network Theory of Psychopathology frames clinical disorders as dynamic, mutually reinforcing systems of directly interacting symptoms [21]. Emerging developmental evidence indicates that adolescent symptom networks exhibit distinctive topological fragility and elevated interconnectivity [22] within such tightly coupled architectures, initial neurovegetative disruptions – such as severe insomnia or fatigue – can directly trigger downstream cognitive dysfunction, emotional distress, and interpersonal withdrawal. By estimating regularized partial correlation networks, psychometric modeling reveals conditional dependencies and bridge indices that link core depressive affect with comorbid somatic and cognitive distress, pinpointing key transition nodes across symptom clusters.

Crucially, examining item-level network topology alongside response distributions informs the practical calibration of an instrument intended for primary care and outpatient triage. Conventional psychiatric inventories frequently retain indicators of extreme, decompensated pathology (such as profound self-care or personal hygiene neglect); while diagnostically informative in specialized inpatient settings, such items exhibit severe floor effects in non-specialized or ambulatory youth populations. A network-informed perspective, combined with severity gradient analysis, facilitates the deliberate filtering of redundant items and the exclusion

of excessively severe indicators that lack discriminative sensitivity for community-level or mild-to-moderate presentations. This structural refinement ensures that the resulting screening tool retains optimal prognostic precision across early clinical stages without inflating scale length.

Nevertheless, network metrics must be applied with methodological rigor. Centrality indices in cross-sectional psychopathology networks do not establish causal symptom-to-symptom mechanisms, and their stability is susceptible to sampling variability and the composition of the conditioning item pool [5, 23, 24]. Therefore, network indicators should not serve as automated, stand-alone selection algorithms. Instead, integrating factor-analytic structural evidence with network bridge topology, severity gradients, and independent clinical judgment provides a robust, multi-method framework for constructing a parsimonious, clinically responsive adolescent screening tool.

The Present Study

The objective of this study was to develop and conduct an initial evaluation of a novel, ICD-11-anchored adolescent depression screening tool, comprising a comprehensive full form and an abbreviated short form. Specifically, the study addressed three primary research questions:

1. **Structural and Psychometric Evaluation of the Full Item Pool:** To examine the item response distributions, multidimensional factor structure, and network topology of the full screening item set across a diverse sample of adolescents.

2. **Integrative Item Selection for the Abbreviated Form:** To determine whether a multi-method item-selection strategy – integrating factor loadings, network bridge/centrality indicators, clinical content validity, and severity gradients – yields an abbreviated form that preserves core diagnostic content while minimizing collinear redundancy.

3. **Diagnostic Utility and Differential Discrimination:** To evaluate the discriminatory performance of the developed full and short forms against both a community comparison group and an active clinical control group (adolescents seeking help for non-depressive conditions), explicitly accounting for classification accuracy alongside an *uncertain interval* (gray zone) of triage.

Methods

Phase I Development of the Initial Adolescent Depression Item Pool

Domain Specification and Item Generation

The depression screening measure was developed as part of a broader computerized adolescent mental health assessment battery within the National Research Programme project *"Development of a new computer-based assessment method for measuring and monitoring adolescent mental health outcomes in post-COVID pandemic conditions"* (Grant No. 1-PB-2/8/2024). The instrument was specifically designed for adolescents aged 13–19 years presenting to primary care, school counseling, or psychological outpatient services.

Initial content specifications were grounded in the International Classification of Diseases, 11th Revision (ICD-11) diagnostic requirements for depressive episodes [10] and related clinical descriptions. The research team identified 16 distinct clinical domains: depressed mood, anhedonia, hopelessness, negative self-evaluation, guilt, reduced energy/fatigue, impaired concentration, psychomotor retardation, sleep disturbance, appetite alterations, thoughts of death, suicidal ideation, rumination, reduced motivation, crying spells, and impaired personal hygiene. These 16 domains served as theoretical content-generation categories to ensure broad construct

coverage, rather than as an a priori multidimensional scoring model. For each domain, three to six candidate items were drafted, resulting in an initial pool of 69 items (Version 1).

Suicide-related content was defined in alignment with ICD-11 category MB26.A (*Suicidal ideation*) and subsumed two of the 16 domains: (a) passive death wishes (wishing not to wake up or wishing to be dead) and (b) active suicidal thoughts (thinking about ending one's life). The initial suicide-related pool comprised nine items, including a binary (yes/no) screener assessing death wishes over the preceding two weeks. These items were designed to capture risk breadth and safety considerations rather than to form independent psychometric subscales.

Expert Content Review and CVI Analysis

Content validity was evaluated by a multidisciplinary panel of six clinical experts (two certified clinical/health psychologists and four child and adolescent psychiatrists), each possessing at least five years of specialized adolescent clinical experience. Experts independently rated each of the 69 items on a 4-point relevance scale (1 = *not relevant*, 2 = *somewhat relevant*, 3 = *quite relevant*, 4 = *highly relevant*) and provided qualitative feedback regarding developmental appropriateness, readability, and diagnostic clarity.

Item-level content validity indices (I-CVI) were computed in accordance with Yusoff [25]: ratings of 3 and 4 were dichotomized as relevant, whereas ratings of 1 and 2 were coded as non-relevant. For a panel of six experts, the conservative threshold for excellent item-level content validity was set at I-CVI $\geq$ .83 (agreement among at least 5 out of 6 experts). The review followed an iterative consensus procedure, with select candidate items undergoing up to three rounds of editorial revision and re-evaluation.

During the initial review stage, experts suggested modifications for 47 items; 17 items were comprehensively revised and subsequently re-evaluated.

Overall:

- 12 items exhibited insufficient content validity (I-CVI < .83) and were removed;
- 25 items met the quantitative threshold (I-CVI $\geq$ .83) but were systematically excluded due to qualitative concerns regarding construct redundancy, adult-centric phrasing, or excessive severity (e.g., severe hygiene neglect) that would induce floor effects in ambulatory care;
- 32 items demonstrated strong quantitative support (including perfect consensus, I-CVI = 1.00) and sound clinical utility, and were retained for empirical pretesting (Version 2).

This retained 32-item battery comprised 23 general depression items, seven graded suicidal-ideation items, one binary control screener for passive death wishes (SI_K), and one lifetime suicide attempt query (*never*, *once*, *more than once*).

Adolescent Pretesting and Cognitive Debriefing

Version 2 was subsequently pretested through individual cognitive interviews with 15 native Latvian-speaking adolescents (7 boys, 8 girls; aged 13–17 years, $M = 15.4, SD = 1.3$). Semi-structured cognitive debriefing sessions were conducted remotely via Zoom. Participants reviewed item readability, emotional tone, clarity of response anchors, and subjective difficulty when reporting sensitive symptoms.

Response formats were operationalized using a 2-week recall window:

- Depressive symptom items utilized a 5-point frequency scale: 0 = *never or not at all*, 1 = *1–2 days*, 2 = *several days*, 3 = *most days*, and 4 = *nearly every day*.

- Suicidal ideation items were rated on a parallel 5-point graded frequency scale: 0 = *never or not at all*, 1 = *rarely*, 2 = *sometimes*, 3 = *often*, and 4 = *every day or almost every day*.
- The binary control item (SI_K) and the lifetime suicide-attempt question were recorded as independent clinical-risk indices and were excluded from cumulative depression severity sums.

Adolescents confirmed that the items were generally clear, engaging, and developmentally accessible. Based on their qualitative debriefing, minor wording simplifications were made to two items. Furthermore, to reduce respondent fatigue and eliminate conceptual overlap, several redundant items were consolidated into broader composite indicators: multiple loss-of-interest indicators were synthesized into a unified anhedonia item (Item D3), and overlapping ruminative statements were merged into a comprehensive item operationalizing the negative cognitive triad - capturing distressing, pervasive negative thoughts directed toward oneself, others, and life in general (Item D17).

Following cognitive pretesting and final content consolidation, an optimized 26-item protocol was advanced to large-scale empirical evaluation. This protocol consisted of:

1. 19 core depressive symptom items;
2. 5 graded suicidal ideation items;
3. 2 independent clinical safety items (the binary screener and the lifetime attempt indicator).

Phase II: Large-Sample Psychometric Evaluation and Instrument Refinement

Participants

The total study sample comprised 828 adolescents and young adults aged 13 to 19 years ($M = 15.64$, $SD = 1{,}65$). Overall, 511 participants (61.7%) identified as female, 301 (36.4%) as male, and 16 (1.9%) as non-binary. Across age bands, 252 participants (30.4%) were aged 13–14 years, 298 (36.0%) were aged 15–16 years, and 278 (33.6%) were aged 17–19 years.

The sample was stratified into three primary comparison groups:

1. **Community comparison group** ($n = 661, 79.8\%$): adolescents recruited from general secondary education cohorts without documented specialized mental health intake;
2. **Depression-symptom group** ($n = 80, 9.7\%$): adolescents receiving clinical mental health services who presented with formally diagnosed depressive disorders or prominent depressive symptomatology;
3. **Other clinical comparison group** ($n = 87, 10.5\%$): adolescents receiving clinical mental health care for primary psychiatric, neurodevelopmental, or behavioral concerns other than depression.

Within the combined clinical cohort ($n = 167$), participants were recruited across three distinct healthcare levels: inpatient child and adolescent psychiatric care ($n = 42, 25.1\%$), specialized outpatient psychiatric care ($n = 53, 31.7\%$), and community-based multidisciplinary youth mental health services ($n = 67, 40.1\%$; care setting was unrecorded for 5 participants, 3.0%). Detailed demographic and clinical characteristics across all groups are summarized in Table 1.

Table 1

Demographic and clinical characteristics of the study sample across groups

Characteristic	Total sample (N = 828)	Community group (n = 661)	Depression-symptom group (n = 80)	Other clinical group (n = 87)
Age (years), *M*(*SD*)	15.6 (1.65)	15.6 (1.66)	16.2 (1.55)	15.7 (1.58)
Age cohorts, *n*(%)				
13-14 years	252 (30.4)	218 (33.0)	11 (13.8)	23 (26.4)
15–16 years	298 (36.0)	236 (35.7)	34 (42.5)	28 (32.2)
17–19 years	278 (33.6)	207 (31.3)	35 (43.8)	36 (41.4)
Gender identity, *n*(%)				
Female	511 (61.7)	379 (57.3)	65 (81.3)	67 (77.0)
Male	301 (36.4)	266 (40.2)	15 (18.8)	20 (23.0)
Non-binary	16 (1.9)	16 (2.4)	0 (0.0)	0 (0.0)
Recruitment care setting, *n*(%)				
Inpatient psychiatric care	42 (5.1)	—	26 (32.5)	16 (18.4)
Specialist outpatient psychiatric care	53 (6.4)	—	30 (37.5)	23 (26.4)
Community-based youth mental health care	67 (8.1)	—	24 (30.0)	43 (49.4)
Setting unrecorded	5 (0.6)	—	0 (0.0)	5 (5.7)

Note. Percentages are calculated relative to the total sample size of each respective column. In the clinical cohort ($n = 167$), community-based mental health services correspond to the Adolescent Resource Centre (ARC).

Clinical Classification and Diagnostic Procedures

Because participants in the clinical cohort ($n = 167$) were recruited across tertiary inpatient, secondary outpatient, and community-based services, diagnostic allocation relied on multi-tier clinical verification available within each respective recruitment setting:

1. **Tertiary and secondary psychiatric settings (Inpatient and Outpatient clinics):** Clinical status was established via documented, physician-adjudicated psychiatric diagnoses according to the International Classification of Diseases, 10th Revision (ICD-10) extracted directly from official medical health records.
2. **Community-based youth mental health care (Adolescent Resource Centre [ARC]):** Group allocation was based on comprehensive, semi-structured intake evaluations conducted by certified multidisciplinary mental health practitioners (psychologists, psychotherapy specialists, psychotherapists), confirming primary and secondary problem areas upon entry.

Consequently, the depression-symptom cohort was operationalized using a clinically meaningful symptom-presentation criterion rather than restricting inclusion exclusively to unipolar major depressive disorder.

Depression-Symptom Group

Adolescents were allocated to the depression-symptom group ($n = 80$) if they presented with documented syndromal depressive disorders or clinically verified primary depressive complaints, confirmed via:

- **ICD-10 diagnostic codes from medical records:**
    - Major depressive episodes and recurrent depressive disorder (F32.x, F33.x);
    - Bipolar affective disorder, current episode depressive or mixed (F31.3–F31.6);
    - Persistent and other mood disorders, including dysthymia (F34.1, F38.x);
    - Mixed syndromes where depressive pathology represented a primary clinical focus, including mixed anxiety and depressive disorder (F41.2), adjustment disorder with depressive or mixed emotional/conduct reactions (F43.20–F43.22, F43.25), depressive conduct disorder (F92.0), or other mixed disorders of conduct and emotions (F92.8);
    - Secondary or post-psychotic depressive syndromes (F06.32, F20.4, F25.1).

Structured intake assessment (ARC): Multi-method evaluation confirming moderate-to-severe depressive symptomatology as the principal referral reason or as a dominant co-occurring presentation upon intake.

Other Clinical Comparison Group

The active clinical comparison group ($n = 87$) comprised adolescents seeking or receiving mental health care whose primary clinical presentation involved conditions other than mood or depressive disorders. Verified through medical documentation and structured intake records, this group presented with:

- **Eating disorders and related metabolic concerns:** Anorexia nervosa (F50.0), atypical and unspecified eating disorders (F50.1, F50.8, F50.9, R63.0), and associated somatic nutritional complications (e.g., E43, E44, N91.0);
- **Non-affective anxiety, somatoform, and dissociative disorders:** Generalized anxiety disorder (F41.1), other anxiety disorders (F41.3), social phobia (F40.1), dissociative and conversion disorders (F44.x), and somatoform presentations (F45.x);
- **Neurodevelopmental and disruptive behavioral disorders:** Attention-deficit hyperactivity disorder (F90.0, F90.x), unsocialized conduct disorder (F91.1), mixed emotional/conduct difficulties without primary mood disturbance (F92.8), emotional disorders specific to childhood (F93.8), and specific developmental disorders of scholastic skills or speech (F80–F81);
- **Substance misuse and behavioral addictions:** Harmful use of alcohol or psychoactive substances (F10.1, F19.1, F55.9) and behavioral/process addictions (e.g., excessive screen use, gambling);
- **Acute situational crises and other psychiatric presentations:** Relational trauma, exposure to domestic violence, severe school bullying, acute academic crisis, organic personality alterations (F07.8), and schizotypal disorder (F21) without active depressive decompensation.

Measures

Initial Adolescent Depression Item Pool (version 3)

Participants completed version 3 of the questionnaire developed in the preceding content-development phase. This version comprised 19 depression-symptom items (D1–D19) and five items assessing thoughts of death and suicide (SI1–SI5), together with two separately recorded

criterion-related questions. The 24 graded symptom items covered affective, cognitive, interpersonal, and somatic complaints, as well as passive wishes for death and explicit suicidal thoughts. Both symptom sections referred to the preceding two weeks.

For the depression items, participants indicated how often each problem had bothered them using a five-category frequency response scale: 0 = never or not at all, 1 = one or two days, 2 = some days, 3 = most days, and 4 = every or almost every day. The five suicide-related items used the same 0–4 scoring range with different frequency labels: 0 = never or not at all, 1 = rarely, 2 = sometimes, 3 = often, and 4 = every day or very often. Higher item scores indicated more frequent symptom experiences.

Initial unweighted scores were calculated by summing the graded items included in each form. For comparability across forms with different item counts, scores were expressed on a 0–100 scale as 100 × sum of item scores / (4 × number of included items). This transformation preserves the ordering of respondents within a form. The administered 24-item symptom set was subsequently reduced to the 23-item full form by excluding D16; the item evaluation and exclusion are reported in the Results. Neither of the two criterion-related questions contributed to the symptom total.

Separate questions about death wishes and suicide attempts

A yes/no question (SI_K) asked whether participants had experienced thoughts of not wanting to live or wishing to be dead during the preceding two weeks. A second question asked whether they had ever attempted suicide, with responses coded 0 = never, 1 = once, and 2 = more than once. These questions were recorded separately as supplementary criterion-related indicators, but were not used in this analysis.

WHO-5 Well-Being Index

The WHO-5 Well-Being Index [26, 27] is a five-item self-report measure of positive psychological well-being during the preceding two weeks. The Latvian adaptation was developed by Boļšakova and Perepjolkina [28]. Each positively worded item is rated on a six-category frequency scale from 0 (at no time) to 5 (all of the time). Item scores are summed to yield a raw score from 0 to 25 and multiplied by four to obtain a 0–100 score, with higher values indicating better well-being [10, 28].

The Latvian adaptation study reported Cronbach's α = .89 in an adult sample of 123 participants aged 18–69 years [28]. Internal consistency in the present adolescent sample was α = α = .89. Previous research supports the WHO-5 as a measure of well-being and as a screening instrument for depression, including evidence from pediatric settings [27, 29]. In the present study, WHO-5 scores were used as a continuous external measure to evaluate convergent validity; higher depression scores were expected to be associated with lower well-being. WHO-5 scores did not serve as the diagnostic reference standard for group assignment.

Demographic information

Participants reported their age and gender. These variables were used to describe the sample and, where relevant, to define the demographic comparisons reported in the study.

Procedure

Ethics approval and consent to participate

The study protocol, encompassing both the qualitative cognitive debriefing pretest and the large-scale test development and validation study, was reviewed and approved by the Research Ethics Committee of Rīga Stradiņš University (Approval No. 2-PĒK-4/209/2024). Participation across all settings was entirely voluntary and confidential. In the community cohort, information letters detailing the study objectives, data protection procedures, and passive parental/guardian consent forms were distributed electronically through the official nationwide school management platform (*E-klase*); adolescents provided electronic informed assent before completing the survey. In the clinical cohort, formal informed consent was obtained in writing from each participant and, where required by national regulations and institutional ethics requirements for minors, signed consent was additionally obtained from their parent or legal guardian.

Community sample administration

Recruitment of the community comparison sample was coordinated across general secondary education institutions in Riga and various regional urban centers, in close professional collaboration with the Latvian Association of School Psychologists. Following formal municipal and institutional approvals, certified school psychologists facilitated the on-site recruitment and data collection. The digitized assessment battery – comprising the 26-item protocol, the WHO-5 Well-Being Index, and a brief sociodemographic questionnaire – was administered via Microsoft Forms. Adolescents completed the self-report measures independently on classroom computers or digital devices during scheduled school hours under the direct supervision of school psychologists. The survey protocol required approximately 10–15 minutes to complete.

Clinical sample recruitment and administration

Participants within the clinical groups were recruited across three specialized healthcare settings: the Children's Clinical University Hospital (inpatient psychiatric units), the Child and Youth Mental Health Centre (specialized outpatient services), and regional branches of the Adolescent Resource Centre (ARC; community-based multidisciplinary outpatient services). Data collection was coordinated directly by certified psychologists practicing within these respective institutions during diagnostic intake evaluations, outpatient consultations, or stabilized inpatient stays.

To obtain objective clinical criteria without compromising adolescent confidentiality, a dual-source data collection protocol was implemented:

1. **Adolescent self-report:** Clinical participants independently completed the digitized self-report battery on a clinic tablet or computer.
2. **Clinician-reported record form:** Treating psychologists independently completed a brief standardized clinical form for each enrolled patient. In hospital and specialized outpatient settings, clinicians recorded primary and secondary ICD-10 psychiatric diagnoses and patient treatment status (inpatient versus outpatient). At the Adolescent Resource Centre, where formal medical diagnostic coding is not utilized, multidisciplinary staff completed an intake record documenting reasons for referral and observed behavioral, emotional, and psychosocial difficulties.

To link adolescent self-reports with clinician records while preserving complete anonymity, each clinical participant was assigned a unique pseudonymous alphanumeric identifier (ID) at enrolment. This identifier was entered into both the self-report survey and the clinician evaluation form, enabling deterministic linkage across files while preventing the inclusion of direct personal identifiers in the analytical dataset.

Data analysis methods

Analyses were exploratory and conducted in the development sample using JASP, jamovi (including the jYS module; [30], and R. Detailed computational configurations, software versions, candidate-form compositions, and threshold comparisons are detailed in the Supplementary Methods; the item-level evidence matrix is provided as a separate supplementary file.

Item structure and symptom network

Exploratory factor analyses (EFA) were conducted in JASP using polychoric correlations, principal axis factoring, and promax rotation. Factor retention was evaluated jointly via factor-based parallel analysis, model fit indices, interpretability, and loading patterns. Item-response distributions were examined by study group, with between-group comparisons evaluated using two-sided Mann–Whitney tests and rank-biserial correlations. Group-specific endorsement proportions and two-by-two overlap tables at response levels 1–4 and 3–4 informed item redundancy and distinct symptom coverage. An exploratory somatic-predominant profile was examined descriptively; these criteria did not serve as diagnostic thresholds.

A regularized symptom network was estimated in R using EBICglasso with polychoric correlations via *bootnet* and visualized with *qgraph*. One-step bridge expected influence (*networktools*) was calculated using bridge communities aligned with the empirical EFA solution. Edge-weight accuracy and centrality stability were evaluated using nonparametric and case-dropping bootstraps [5, 6]. Network parameters were interpreted as conditional associations rather than causal pathways.

Item evidence reviews, short-form selection, and scoring

The 23 retained items were evaluated using a structured evidence matrix integrating primary and general-factor loadings, response distributions, between-group differentiation, network centrality, clinical content, and cognitive debriefing observations. Alignment with established depressive and suicide-risk screening domains [10, 31, 32] further guided evaluation. Factor loadings below .40 prompted substantive review rather than automatic exclusion. Short-form selection prioritized comprehensive symptom coverage while minimizing redundancy, without relying on mechanical numerical rankings.

Unweighted scale scores were calculated by summing raw responses (0–4). For cross-form comparisons, scores were transformed to a 0–100 metric (100 × raw total / theoretical maximum), while raw sums were retained for operational use. Provisional weights were assigned through qualitative clinical synthesis and observed group contrasts rather than mathematical optimization; full-form weights were held constant in exploratory shortened forms.

Scoring configurations were compared in R against the other clinical and community groups. Receiver operating characteristic areas under the curve (AUCs), accounting for ties, and paired differences were estimated using 2,000 stratified paired bootstrap resamples. For the selected 13-item configuration, unweighted and weighted scores were further compared via ridge logistic regression (*glmnet*) using 20 repetitions of nested stratified five-fold cross-validation (with inner regularization tuned via binomial deviance and the one-standard-error rule). This evaluation was conditional on the selected items and did not validate the preceding selection process.

Exploratory score boundaries and form comparisons

Uncertain-interval analyses were conducted for each clinical contrast in jamovi using the jYS module, following Landsheer [33, 34]. Estimated score boundaries, interval proportions, and classification metrics outside the interval were evaluated jointly. The uncertain interval represents an inconclusive score range rather than a statistical confidence interval.

Operational raw-score boundaries were explored in R. Initial lower boundaries targeted at least 90% sensitivity (with a 95% target also examined) and were adjusted in one-point increments to examine sensitivity–specificity trade-offs. Upper candidates were evaluated via classification counts and subgroup analyses across gender and care settings. Wilson 95% confidence intervals were calculated for full-sample sensitivity and specificity, without adjusting for threshold-selection optimism.

Unweighted short and full forms were compared using paired AUC differences (2,000 stratified paired bootstrap resamples). Three-band agreement was assessed using cross-tabulations, exact agreement rates, and directional disagreement. Lower-boundary sensitivity was also examined descriptively across age and gender subgroups within the depression-symptom cohort.

Internal consistency and additional validity evidence

Item distributions, corrected item–rest correlations, and McDonald's omega (overall and following item omission) were calculated in jamovi across the pooled sample and within each study group. Short-form dimensionality was re-evaluated in JASP using one- and two-factor EFA solutions alongside parallel analysis.

Subgroup differences across gender (Mann–Whitney tests), age, and care settings (Kruskal–Wallis tests) were analyzed in R with Holm corrections applied within comparison families; significant omnibus tests were followed by pairwise comparisons. Concurrent validity correlations with the WHO-5 and between form variants were estimated in jamovi with 95% confidence intervals based on available paired observations without imputation. Distribution plots were generated using *ggplot2*.

Results

Exploratory factor structure and initial item retention

To examine the underlying latent dimensionality of the initial 24-item pool and identify poorly performing items for preliminary reduction, we conducted exploratory factor analyses using polychoric correlations.

The initial empirical pool comprised 24 polytomous items scored on a 5-point frequency scale: 19 depression-symptom items and five suicidal-ideation items. In the initial four-factor analysis, D16 (severe hygiene neglect) did not reach a loading of .30 on any factor and was excluded from subsequent solutions. The remaining 23 items were then examined in fixed one-, two-, three-, and four-factor solutions in the full development sample (N = 828). These analyses characterized the initial item pool and informed subsequent selection; they did not establish the final screening form.

All five response categories were used for each of the 23 items retained after the initial item review, with no missing item responses in the development sample. The correlation matrix was suitable for factor analysis (overall KMO = .964; Bartlett's test of sphericity, $\chi^2(253) = 19{,}450.24$, $p < .001$). Factor-analysis-based parallel analysis supported four factors: the observed

factor eigenvalues exceeded the corresponding simulated means for the first four factors but not the fifth.

Although the one-factor solution captured substantial shared variation, its fit was poor (Table S2). In the two-factor solution, suicidal-ideation items formed a distinct factor, whereas the remaining symptoms largely formed a broad depression factor. D8 and several negative self-evaluation items loaded on both factors (D8 = .520, D9 = .421, D10 = .362, D11 = .323). The three-factor solution additionally separated somatic and functional symptoms from the broader emotional–cognitive symptom factor. The four-factor solution further distinguished negative self-evaluation and related cognitive symptoms. Fit improved across the one- through four-factor solutions, with the four-factor solution showing the best fit among those examined (RMSEA = .085, 90% CI [.080, .090], CFI = .948, TLI = .921, SRMR = .021). Nevertheless, the RMSEA indicated some remaining misfit, and the four factors were substantially correlated (*r*s = .596–.761).

In the four-factor solution (Table S3), the suicidal-ideation factor showed its strongest primary loadings for SI2 (.918), SI3 (.914), SI1 (.875), and SI5 (.870). The depressed mood and anhedonia factor was marked by D3 (.879), D7 (.759), and D6 (.641). The negative self-evaluation factor was marked particularly by D10 (.741) and D9 (.648). The somatic and functional symptom factor showed its strongest loading for sleep problems (D14 = .756), followed by appetite change (D15 = .576), concentration difficulties (D12 = .567), and fatigue (D2 = .555). D8, SI4, D13, and D19 showed cross-loadings. The four-factor solution was therefore retained as a working description of symptom organization to inform subsequent item selection, rather than as the established structure of the eventual short form.

Item response patterns and clinical differentiation

Following the dimensional evaluation, we examined item-level response distributions and between-group contrasts to determine how effectively individual symptoms differentiated adolescents with depression from those with other clinical conditions or community peers.

Response distributions across all three study groups are detailed in Table S4, with descriptive frequencies of categorized responses (levels 1–4 and 3–4) displayed in Table S5 and Figures S1–S2. Across items, the depression-symptoms group demonstrated an overall shift toward higher response categories. However, the extent of category overlap with the other clinical cohort varied substantially across symptoms. For instance, while 51.2% of adolescents in the depression-symptoms group and 8.0% in the other clinical group endorsed severe levels (categories 3–4) for D8, the corresponding proportions for fatigue (D2) were 67.5% and 33.3%, indicating that certain somatic complaints were prevalent across broader clinical presentations.

Mann–Whitney comparisons between the two clinical groups confirmed significantly higher scores in the depression-symptoms group for all 23 items (all ps < .001; Table S6), with rank-biserial correlations ranging from .340 to .642. The most pronounced differentiations were observed for suicidal ideation (SI1: $r$rb = .642; SI3: $r$rb = .611; SI2: $r$rb = .609), D8 ($r$rb = .640), and concentration difficulties (D12: $r$rb = .589). The smallest effects emerged for guilt (D11: $r$rb = .340) and tearfulness (D19: $r$rb = .362). Conversely, severe endorsement of D13 and SI5 occurred in only 30.0% of the depression-symptoms cohort, despite significant differentiation across the full ordinal distribution.

Symptom overlap and content coverage

Item-overlap analyses at response levels 3–4 were examined to balance symptom coverage and redundancy (Table S7 and electronic supplementary matrix). Discordant endorsement patterns confirmed that narrower manifestations (e.g., endorsing D6 or D7 without D3, D4 without D8, or SI1 without SI2/SI3) reflected clinically relevant distinctions warranting retention for subsequent review, rather than immediate exclusion.

To determine whether prominent somatic or functional impairment could emerge without explicit mood disturbance, an exploratory somatic-predominant profile was evaluated (defined as endorsing levels 3–4 on at least two somatic/functional items [D2, D12–D15] alongside levels below 3 across all five primary mood/anhedonia items [D1, D3, D5–D7]). This profile occurred with comparable overall frequency in the community (12.6%, 83/661) and depression-symptoms groups (12.5%, 10/80; Table S7). However, among participants lacking frequent endorsement of those specific mood symptoms, the conditional proportion exhibiting the somatic profile was markedly higher in the depression-symptoms group (45.5%, 10/22) than in the community group (19.8%, 83/419). Although exploratory, these patterns demonstrate that screening based solely on core affective items risks misclassifying adolescents presenting primarily with somatic complaints, supporting the provisional retention of all 23 items at this stage.

Exploratory symptom network

Because latent factor models capture shared variance but do not reveal direct conditional dependencies among specific symptoms, we estimated a regularized symptom network to evaluate symptom-to-symptom interactions and identify bridge items spanning clinical domains.

The initial regularized network inspected in JASP yielded 134 non-zero edges among 253 possible item pairs (sparsity = .470). The network subsequently estimated in R demonstrated an identical

edge count (134 non-zero edges; Figure S3). The strongest edge weights emerged between suicidal-ideation items (SI1–SI3: .355; SI2–SI3: .350), negative self-evaluation items (D9–D10: .337), and cognitive complaints (D12–D13: .321).

Case-dropping bootstrap analyses yielded correlation stability (CS) coefficients of .361 for strength, .284 for closeness, and .128 for betweenness. In line with methodological recommendations [5], betweenness was not used to prioritize items, and strength differences were interpreted with caution. Four symptom communities were designated based on the empirical EFA solution: Suicidal Ideation; Emotional and Cognitive Symptoms; Negative Self-Evaluation and Interpersonal Distress; and Somatic and Activity-Related Symptoms. Across these communities, the highest one-step bridge expected influence values were observed for D17 (.587), D15 (.586), D8 (.584), D1 (.557), and D9 (.536; Table S8). Bridge expected influence exhibited excellent stability, maintaining a correlation of at least .70 with original estimates at the maximum case-dropping proportion of .75 (CS-coefficient ≥ .75; Figure S4).

Nonparametric bootstrapping evaluated edge-weight precision. Because the regularized network was relatively dense and minimal regularization was selected across several bootstrap draws, the presence and exact rank order of weaker edges were interpreted conservatively. Network metrics were evaluated in conjunction with factor loadings, diagnostic differentiation, and substantive clinical content to guide subsequent item selection.

Construction of short-form candidates

Rather than relying on a single item-reduction criterion, we explored alternative item-selection strategies-ranging from strictly data-driven criteria (factor loadings, network metrics) to hybrid approaches that balanced empirical performance with qualitative diagnostic coverage. These

candidate short forms were designed to identify an optimal balance between brief administration and comprehensive clinical symptom representation.

Development of candidate short forms proceeded iteratively from initial factor-derived configurations to variants incorporating clinical content coverage and network bridge centrality. The baseline candidate, EFA12A_Loading, comprised the three items demonstrating the highest primary loadings on each of the four EFA factors. EFA12B_Content replaced D3 with D5 to ensure direct representation of depressed mood and reduce content redundancy among positive affect and interest items. EFA12C_SI5 further substituted SI1 with SI5 to include explicit assessment of active suicidal ideation. These step-wise modifications permitted systematic evaluation of content substitutions while holding total test length constant at 12 items.

The Prior12 candidate represented a 12-item configuration established during an earlier phase of this research [11]. The current development sample expanded that initial cohort by including 19-year-olds, individuals identifying as nonbinary, and broader diagnostic presentations within the depression-symptoms group. Consequently, this evaluation constitutes a re-assessment within an expanded, overlapping sample rather than an independent validation.

Subsequent candidates integrated factor structure, between-group contrasts, network connectivity, and clinical utility. An expanded 14-item integrative variant (Integrative14_D18) added D18 to evaluate the diagnostic utility of assessing social withdrawal. A distinct 13-item configuration (NetworkContent13) prioritized items demonstrating strong bridge centrality across symptom domains while retaining core depressive symptoms (depressed mood, anhedonia, hopelessness, cognitive/vegetative features, and suicidal thoughts). A 14-item derivative (NetworkContent14_D6) added D6 to evaluate the incremental contribution of explicit interest loss. The full composition and selection rationale for each candidate short form are summarized in

Table 2. In clinical practice, responses to suicide-related items necessitate immediate, independent clinical evaluation; their inclusion within total screening scores does not substitute for dedicated suicide risk assessment protocols.

Table 2

Candidate short forms and selection rationale

Candidate short form	***k***	**Item codes**	**Selection rationale**
Full23	23	D1, D2, D3, D4, D5, D6, D7, D8, D9, D10, D11, D12, D13, D14, D15, D17, D18, D19, SI1, SI2, SI3, SI4, SI5	Full-length reference scale.
EFA12A_Loading	12	D3, D6, D7, D9, D10, D11, D12, D14, D15, SI1, SI2, SI3	Three highest primary factor loadings per EFA factor.
EFA12B_Content	12	D5, D6, D7, D9, D10, D11, D12, D14, D15, SI1, SI2, SI3	EFA12A with D3 replaced by D5 to prioritize depressed mood and reduce affect redundancy.
EFA12C_SI5	12	D5, D6, D7, D9, D10, D11, D12, D14, D15, SI2, SI3, SI5	EFA12B with SI1 replaced by SI5 to incorporate explicit suicidal thought assessment.
Prior12	12	D1, D2, D3, D4, D5, D8, D9, D12, D18, D19, SI3, SI5	Form derived during preliminary research phase in an overlapping sample.

Integrative13	13	D2, D4, D5, D6, D7, D8, D9, D10, D12, D14, D15, SI3, SI5	Synthesizes broad symptom coverage, clinical contrasts, EFA, and clinical utility.
Integrative14_D18	14	Integrative13 items plus D18	Evaluates incremental contribution of social withdrawal.
NetworkContent13	13	D1, D4, D7, D8, D9, D10, D12, D14, D15, D17, SI3, SI4, SI5	Balances core symptom coverage with high cross-domain network connectivity.
NetworkContent14_D6	14	NetworkContent13 items plus D6	Evaluates incremental contribution of explicit loss of interest.

Note. k = number of items. Each scale score was computed as an unweighted item sum transformed to a 0–100 metric ($100 \times$ raw sum$/[4 \times k]$). Item codes refer to the standardized inventory item pool. Inclusion of Prior12 illustrates model refinement and does not represent independent external validation. Endorsement of suicidal-ideation (SI) items mandates independent clinical risk review regardless of the composite score. Candidate selection rationales reflect construction criteria and do not imply established incremental predictive superiority.

Comparison of short-form candidates

To identify the most robust short-form configuration for operational use, we compared candidate forms against the full 23-item pool in terms of diagnostic discrimination (AUC) and classification indeterminacy via uncertain-interval analyses.

Across candidates, AUCs ranged from .856 to .877 for distinguishing the depression-symptoms group from the other clinical group and from .824 to .838 for distinguishing it from the

community group. The full 23-item form yielded AUCs of .868 and .836, respectively (Tables S9a and S9b). Several shorter forms thus demonstrated comparable empirical discrimination, although their uncertain intervals (UIs) and the allocation of participants outside the UI boundaries differed.

The 13-item network-and-content form yielded AUCs of .876 against the other clinical group and .838 against the community group, compared with .868 and .836 for the full form. In the clinical comparison, 43/167 participants (25.7%) fell within its UI, compared with 37/167 (22.2%) for the full form. Thirteen versus 12 depression-symptoms participants scored below the respective lower boundaries, and five versus seven other clinical participants scored above the respective upper boundaries. In the community comparison, the corresponding UI counts were 105/741 (14.2%) and 104/741 (14.0%). Thirteen versus 14 depression-symptoms participants scored below the lower boundaries, whereas 109 versus 98 community participants scored above the upper boundaries.

Adding D6 to the network-and-content form did not improve discrimination (.866 versus .876 against the other clinical group; .836 versus .838 against the community group). The 12-item EFA variant with SI5 yielded AUCs of .876 and .834 but assigned a substantially larger proportion of participants to the UI (62/167 and 144/741, respectively). The prior 12-item form yielded AUCs of .867 and .837, with 28/167 and 75/741 participants falling within the UIs; 13 other clinical and 130 community participants scored above its upper boundaries.

Based on these findings, the 13-item network-and-content form (NetworkContent13; comprising D1, D4, D7, D8, D9, D10, D12, D14, D15, D17, SI3, SI4, and SI5) was selected as the optimal configuration. It reduced scale length by nearly half while preserving broad symptom coverage and achieving diagnostic discrimination comparable to the 23-item pool across both contrast groups. This configuration is hereafter referred to as the **short form** (and the 23-item

reference version as the **full form**). Its selection concluded the initial candidate-development stage, establishing the item composition for subsequent evaluations of item weighting, cut-off thresholds, and psychometric properties.

Item evidence review and provisional weighting

Having selected the 13-item configuration, we conducted a comprehensive review across all empirical indices to establish qualitative weighting tiers reflecting clinical priority and diagnostic differentiation. The structured item review supported retaining broad content coverage spanning mood, cognitive, negative self-evaluation, interpersonal, somatic, functional, and suicidal-ideation domains. Overlap analyses and bridge expected influence estimates informed qualitative weighting tiers rather than serving as independent diagnostic criteria.

As detailed in Table 3, provisional item weights ranged from 2.0 to 4.0 across five discrete tiers. The highest weight (4.0) was assigned to items demonstrating both strong diagnostic contrast and high network bridge centrality (D8, D9) or direct safety relevance (SI3, SI5). Intermediate weights (2.5–3.5) reflected core depressive features, prominent factor loadings, or clinically vital somatic complaints characterized by higher base-rate endorsement. The lowest weight (2.0) was reserved for symptoms exhibiting comparatively weaker between-group separation (D11, D13, D19). All item-level indices and weighting rationales are documented in the electronic supplementary evidence matrix.

Table 3

Provisional item weights and tier rationales for the full-form pool

Weight tier	Items	Selection and tier rationale
4	D8, D9, SI3, SI5	Distinct, clinically critical content; D8 and D9 combined marked diagnostic differentiation with high bridge centrality, whereas SI3 and SI5 represent direct safety indicators.
3.5	D1, D3, D4, D5, D7, D10, D12, SI1, SI2, SI4	Strong clinical or factor evidence, accompanied by minor qualifications (e.g., overlapping affective content, compound phrasing, or non-negligible background endorsement).
3	D6, D17, D18	Distinct and clinically informative content; D17 demonstrated high contrast and bridge influence balanced against broad wording and modest primary factor loading.
2.5	D2, D14, D15	Core somatic domains with appreciable endorsement in non-depressive clinical groups or limited unique binary coverage.
2	D11, D13, D19	Retained distinct content demonstrating comparatively weaker between-group differentiation or limited unique severe endorsement.

Note. Tiers 1.0 and 1.5 were available in the weighting scheme but were not assigned to any retained item. Affirmatory responses to suicidal-ideation items mandate independent clinical risk

evaluation regardless of the composite score. The full item-level evidence matrix is provided in the supplementary material.

Comparison of scoring methods

Following the assignment of provisional weights, we evaluated whether weighted totals or regularized algorithmic scoring (ridge logistic regression) yielded measurable diagnostic gains over simple unweighted summation across validation resamples (Table S10).

In the full form (23 items), unweighted and provisionally weighted scores were nearly collinear (r = .9984). Apparent AUCs against the other clinical group were .8459 (95% paired-bootstrap percentile interval [.7786, .9048]) for the unweighted sum and .8473 ([.7812, .9067]) for the weighted sum (paired difference = .0014, 95% CI [−.0029, .0056]). Against the community group, corresponding AUCs were .8236 ([.7686, .8748]) and .8270 ([.7722, .8779]), yielding a negligible difference of .0035 ([.0006, .0066]). Exploratory backward step-wise omission of lowest-performing items (sequentially removing D11, D13, and D6) yielded 22-, 21-, and 20-item scores; while the 20-item version achieved marginally higher observed AUCs (.8504 unweighted and .8507 weighted against other clinical; .8261 and .8289 against community), these minor increments did not justify purely data-driven item elimination.

For the selected 13-item short form, the correlation between unweighted and weighted sums remained .9986. Paired AUC differences between weighted and unweighted scores were minimal: −.0013 (95% CI [−.0065, .0034]) against the other clinical cohort and .0032 ([.0008, .0059]) against the community cohort (Table S10).

In repeated nested five-fold cross-validation across 20 repetitions, ridge logistic regression demonstrated no mean discriminative advantage over the simple unweighted sum in the clinical contrast (mean AUC difference = −.0006; Table 4). In the community contrast, ridge regression

showed a modest but consistent advantage (mean AUC = .8340 vs. .8255; difference = .0085), which remained positive across all 20 random partition splits (Table S11). However, given the absence of empirical benefit in distinguishing depressive from other clinical presentations, and to prioritize operational simplicity and direct interpretability in routine screening, the unweighted item sum was retained as the definitive scoring metric. Exploratory interval and threshold outcomes for all alternative scoring variants are detailed in Supplementary Table S12.

Table 4

Discrimination of alternative scoring configurations for the 13-item short form in repeated nested cross-validation

Scoring method	**Other clinical contrast (mean AUC)**	**Community contrast (mean AUC)**
Unweighted sum	0.8486	0.8255
Provisionally weighted sum	0.8472	0.8285
Ridge logistic regression	0.848	0.834
Ridge minus unweighted sum	−.0006	0.0085

Note. AUC = area under the receiver operating characteristic curve. Both diagnostic contrasts designate the depression-symptoms cohort as the target positive condition.

Values reflect grand means across 20 independent repetitions of nested stratified five-fold cross-validation; within each repetition, test-fold AUCs were averaged using positive–negative pair counts as weights. Ridge models were tuned separately by contrast. Item selection preceded cross-validation and was not re-executed within resampling folds.

Exploratory development of common score thresholds

Whereas the preceding uncertain-interval analyses estimated contrast-specific score regions, routine clinical screening necessitates a single, unified scoring rule applicable across diverse presentation contexts. We therefore explored common operational raw-score boundaries designed to function effectively against both clinical and community comparators, while evaluating their classification stability across demographic subgroups (gender, age) and clinical care settings.

Lower boundaries were initially selected as the highest integer scores retaining at least 90% sensitivity against depression-symptoms group membership, yielding thresholds of 13 for the short form and 21 for the full form. However, diagnostic specificity at these baseline cut-offs was limited: 55.2% and 54.0%, respectively, against the other clinical group, and 51.7% and 49.5% against the community group. Candidate lower boundaries were consequently incremented in one-point steps to systematically evaluate the trade-off between sensitivity and specificity across scores of 13–20 for the short form and 21–35 for the full form (Supplementary Tables S13 and S14).

For the short form, establishing a lower boundary at 19 preserved 83.8% sensitivity while increasing specificity to 72.4% against the other clinical group and 72.5% against the community group. Elevating the threshold from 18 to 19 enhanced specificity without sacrificing sensitivity. A further increase to 20 reclassified three additional depression-symptoms respondents below the cut-off, reducing sensitivity to 80.0% for modest specificity gains (4.6 and 2.4 percentage points, respectively). For the full form, sensitivity remained stable at 85.0% across cut-offs of 28–32, accompanied by progressive gains in specificity. At a threshold of 32, specificity reached 71.3% against other clinical cases and 70.0% against community controls; increasing the boundary to 33 reduced sensitivity to 82.5% for comparatively marginal gains in specificity. Boundaries of 19

(short form) and 32 (full form) were therefore retained as optimal empirical compromises (Table 5).

The previously evaluated upper boundaries were retained. For the short form, shifting the upper boundary from 23 to 25 reduced high-band classifications from 57 to 53 depression-symptoms respondents, disproportionately affecting male participants (decreasing from 9/15 to 6/15 among boys, compared with 48/65 to 47/65 among girls). These subgroup observations, evaluated alongside comparison-group classifications, supported retaining 23 as the upper threshold. For the full form, cut-offs of 44 and 45 classified the identical set of 53 depression-symptoms respondents as high-risk, while 45 yielded fewer false positives in the comparison cohorts (Table S15).

The resulting operational three-tier bands were designated as low (0–18), intermediate (19–22), and high (23–52) for the short form, and low (0–31), intermediate (32–44), and high (45–92) for the full form. In the community group, the short form categorized 72.5% of adolescents into the low-risk band, 8.9% into the intermediate band, and 18.6% into the high-risk band; corresponding full-form proportions were 70.0%, 15.1%, and 14.8%, respectively (Table S16).

Twelve depression-symptoms participants scored below the lower boundary across both scales; one additional respondent scored below the short-form lower threshold but fell within the full-form intermediate band. Exploratory subgroup evaluations revealed lower sensitivity among boys and particularly among younger adolescents aged 13–14 years across both forms. In this youngest subgroup, short-form sensitivity was already constrained at the initial threshold of 13 and deteriorated further at cut-offs exceeding 16 (Table S17). Four depression-symptoms respondents exhibited marked non-endorsement on both instruments, recording short-form raw

totals of 0–4 and full-form totals of 3–13. All operational boundaries were selected and evaluated within the development sample and remain provisional pending independent external validation.

Table 5

Apparent clasification performance of provisional raw-score thresholds for the unweighted forms

Form	Score threshold	Sensitivity [95% CI]	Specificity: other clinical [95% CI]	Specificity: community [95% CI]
Short form	≥19: intermediate or high	.838 [.742, .903]	.724 [.622, .807]	.725 [.689, .757]
Short form	≥23: high	.713 [.605, .800]	.828 [.735, .893]	.814 [.782, .842]
Full form	≥32: intermediate or high	.850 [.756, .912]	.713 [.610, .797]	.700 [.664, .734]
Full form	≥45: high	.663 [.554, .757]	.874 [.788, .928]	.852 [.823, .877]

Note. CI = confidence interval. The target condition comprised 80 adolescents in the depression-symptoms group; comparison cohorts comprised 87 adolescents with other clinical presentations and 661 community peers. All respondents contributed to each diagnostic estimate. Confidence intervals were calculated using the Wilson score method and do not adjust for threshold-selection optimism. Intermediate plus high tiers represent screening-positive cases at the lower threshold; only high-tier classifications are considered positive at the upper threshold. Operational score tiers:

short form (low = 0–18, intermediate = 19–22, high = 23–52); full form (low = 0–31, intermediate = 32–44, high = 45–92). These operational bands are distinct from algorithm-estimated uncertain intervals. Elevated screening scores indicate heightened risk and do not constitute an independent psychiatric diagnosis.

Discrimination and agreement between forms

To evaluate whether reducing scale length compromised diagnostic utility, we directly compared the diagnostic discrimination and classification concordance of the 13-item short form against the 23-item full form using paired bootstrap resampling and cross-classification across the provisional score bands.

The short and full forms demonstrated virtually equivalent empirical discrimination. Against the other clinical group, AUCs were .848 (95% CI [.785, .905]) for the short form and .846 ([.781, .903]) for the full form, yielding a negligible paired difference of .003 (95% CI [−.009, .014]). Against the community group, corresponding AUCs were .825 ([.770, .876]) and .824 ([.769, .874]), with a difference of .002 (95% CI [−.007, .011]). Both bootstrap difference intervals spanned zero (Table 6, Panel A). While these findings confirm strong comparative consistency in the development cohort, they reflect conditional performance in an overlapping sample rather than formal statistical equivalence.

Applying the provisional operational score tiers (short form: 0–18, 19–22, 23–52; full form: 0–31, 32–44, 45–92), exact tier agreement reached 90.2% across the total sample (n = 747/828) and ranged from 85.1% to 93.8% across individual study groups (Table 6, Panel B; Supplementary Table S18). All observed discrepancies were confined to adjacent categories; notably, no participant was classified into the low-risk band by one scale and the high-risk band by the other.

Within the target depression-symptoms group, complete concordance was observed for severe presentations: all 53 participants designated as high-risk by the full form were identically classified by the short form. An additional four participants were categorized as intermediate by the full form but high-risk by the short form. Twelve participants were categorized into the low-risk band by both instruments, and one participant scored below the threshold solely on the short form.

Table 6

Paired diagnostic discrimination and classification agreement between the full and short forms

Panel A. Paired discrimination

Comparator	*n* depression / comparator	Short AUC [95% CI]	Full AUC [95% CI]	Short − full [95% CI]
Other clinical	80 / 87	.848 [.785, .905]	.846 [.781, .903]	.003 [−.009, .014]
Community	80 / 661	.825 [.770, .876]	.824 [.769, .874]	.002 [−.007, .011]

Panel B. Classification agreement across provisional three-tier score bands

Sample	*n*	Same band, *n* (%)	Short higher, *n* (%)	Short lower, *n* (%)
Other clinical	87	74 (85.1)	8 (9.2)	5 (5.7)
Depression symptoms	80	75 (93.8)	4 (5.0)	1 (1.3)
Community	661	598 (90.5)	36 (5.4)	27 (4.1)

Sample	*n*	Same band, *n* (%)	Short higher, *n* (%)	Short lower, *n* (%)
All respondents	828	747 (90.2)	48 (5.8)	33 (4.0)

Note. AUC = area under the receiver operating characteristic curve; CI = confidence interval. Panel A: Evaluated across all eligible participants including intermediate score bands. Confidence intervals reflect stratified paired-bootstrap percentile intervals based on 2,000 resamples and do not adjust for item-selection optimism. Panel B: Operational score bands: short form (low = 0–18, intermediate = 19–22, high = 23–52); full form (low = 0–31, intermediate = 32–44, high = 45–92). "Short form higher" and "Short form lower" designate discrepancies between ordinal risk tiers, not differences between raw numerical scores. No data were missing, and no classification discrepancies spanned non-adjacent tiers (low vs. high). High concordance indicates empirical consistency across scale lengths but does not establish absolute clinical interchangeability.

Item performance and internal consistency

To evaluate the reliability and item-level contribution of retained questions, we examined corrected item–rest correlations and internal consistency across the total development cohort and individual study groups.

Corrected item–rest correlations were positive for all retained items, ranging from .525 to .808 in the full form and from .505 to .805 in the short form (Table S19). Across both versions, D8 demonstrated the highest and D14 the lowest item–rest correlation. Hypothetical omission of D14 increased McDonald's omega by only .001 in each form; these marginal diagnostic variations did not warrant removing any selected items.

McDonald's omega was .960 for the full form and .935 for the short form in the total development sample (N = 828). Within-group reliability estimates remained high, ranging from

.933 to .965 for the full form and from .885 to .943 for the short form, with the lowest values observed in the other clinical group (Table S20). Both instruments thus demonstrated strong internal consistency across clinical and community cohorts.

Exploratory structure of the short form

To verify whether the 13-item short form retained an essentially unidimensional scale structure suitable for composite sum scoring, we evaluated fixed one- and two-factor exploratory solutions alongside parallel analysis.

In the one-factor solution, all 13 items loaded positively, with factor loadings ranging from .538 (D14) to .941 (D8) and accounting for 67.3% of the shared variance. While the standardized root mean square residual was acceptable (SRMR = .051), other indices indicated substantial residual misfit: RMSEA = .194 (90% CI [.179, .212]), CFI = .828, and TLI = .793 (Tables S21–S22).

A manually specified two-factor solution modestly improved model fit, though considerable misfit persisted (SRMR = .039; RMSEA = .183, 90% CI [.165, .202]; CFI = .876; TLI = .816). The two latent dimensions were strongly correlated (r = .802). Suicidal-ideation items loaded predominantly on the first factor, whereas vegetative complaints (sleep and appetite disturbances) loaded primarily on the second. Several core affective and cognitive self-evaluation items demonstrated substantial cross-loadings across both dimensions, precluding the formation of distinct, interpretable subscales (Table S21).

Factor-based parallel analysis indicated a prominent general factor; the second observed eigenvalue only marginally exceeded simulated threshold means across estimation runs, and the fourth eigenvalue exceeded the threshold in only a single run (Table S22). Taken together, these findings demonstrate a robust dominant common dimension characterized by residual

multidimensionality, supporting a single summary screening score rather than separate subscale metrics.

Score distributions and exploratory subgroup comparisons

To establish normative baseline patterns and assess whether screening scores varied systematically by demographic characteristics or clinical setting, we compared score distributions across gender, age bands, and healthcare facilities.

Mean short-form raw totals were 30.36 (*SD* = 13.50) in the depression-symptoms group, 12.79 (*SD* = 9.44) in the other clinical group, and 13.95 (*SD* = 10.82) in the community group. Corresponding full-form means were 53.24 (*SD* = 23.27), 22.95 (*SD* = 16.71), and 24.92 (*SD* = 18.61), respectively. Detailed distributional parameters across raw and standardized 0–100 metrics are reported in Supplementary Table S23; overlapping density distributions and provisional risk boundaries are illustrated in Figures S5 and S6.

Female adolescents exhibited significantly higher short-form rank scores than male peers within both the depression-symptoms group (rank-biserial $r = .389$, Holm-adjusted $p = .040$) and the community group ($r = .361$, Holm-adjusted $p < .001$). Gender differences were not statistically significant in the other clinical cohort ($r = .144$, Holm-adjusted $p = .332$; Table S24). Community participants identifying as nonbinary were described separately ($n = 16$; $M = 27.75$, $SD = 12.85$) without inferential hypothesis testing due to subsample size.

Statistically significant age differences emerged within the community group ($H(2) = 11.30$, Holm-adjusted $p = .011$), but not within either clinical group (Table S25). Post-hoc pairwise comparisons revealed that community adolescents aged 15–16 and 17–19 years had higher score ranks than those aged 13–14 years, with small effect sizes ($r = .136$, Holm-adjusted $p = .025$; and

r = .180, Holm-adjusted p = .004, respectively); scores did not differ between the two older age bands (Holm-adjusted p = .468).

Differences across clinical care settings (inpatient psychiatric units, specialized outpatient clinics, and community outpatient services) were not statistically significant after familywise correction (other clinical: $H(2) = 4.85$, Holm-adjusted p = .088; depression symptoms: $H(2) = 6.31$, Holm-adjusted p = .085; Table S26). Five clinical participants with unrecorded care settings were excluded from these comparisons.

Associations with well-being

To establish concurrent criterion validity against an independent measure of positive psychological functioning, we examined bivariate correlations between scale totals and WHO-5 Well-Being Index scores.

Among 766 participants providing complete paired responses, screening scores correlated inversely and strongly with the WHO-5: $r = -.631$ (95% CI [−.672, −.586]) for the short form and $r = -.649$ (95% CI [−.688, −.606]) for the full form (both ps < .001; Table S27). Total scores on the short and full forms were exceptionally correlated (r = .987, 95% CI [.985, .989]; N = 828), confirming that scale reduction retained the overarching construct variance. Concurrent validity analyses encompassed 92.5% of the development sample; missing WHO-5 observations were not imputed.

Discussion

Principal findings

This study developed and psychometrically evaluated a novel, ICD-11-anchored adolescent depression screening tool, culminating in a 23-item comprehensive inventory (The Adolescent Depressive Symptoms Screener; ADSS-23) and a 13-item abbreviated screener

(ADSS-13). Addressing the three core research questions outlined in the introduction, the findings demonstrate that: (1) adolescent depressive symptomatology exhibits an overarching general distress dimension underpinned by correlated cognitive, affective, vegetative, and suicidal domains; (2) an integrative item-selection strategy balancing empirical loading patterns, network bridge topology, and clinical content coverage successfully reduced test length by nearly half while preserving full-scale diagnostic discrimination; and (3) both the ADSS-13 and ADSS-23 robustly differentiated depressive presentations not only from community controls, but importantly, from an active clinical control cohort presenting with diverse non-depressive psychiatric complaints. Furthermore, unit scoring (unweighted sum) demonstrated parity with complex algorithmic and qualitative weighting, supporting operational simplicity for routine screening.

Structural dimensionality and clinical construct representation

The exploratory structural analyses of the full item pool revealed a multidimensional yet strongly interdependent symptom architecture. Although four distinct symptom domains emerged (suicidal ideation, emotional/anhedonic symptoms, negative self-evaluations, and somatic/functional disruptions), high inter-factor correlations (*r*s = .60–.76) and parallel analysis indicated a dominant common factor. When evaluating the 13-item short form, neither strict unidimensional nor two-factor models demonstrated adequate goodness-of-fit, despite a single dominant eigenvalue and strong item loadings (.54–.94).

In adolescent psychopathology, this structural pattern is consistent with an overarching "p-factor" or broad internalizing distress dimension characterized by substantive residual multidimensionality. Attempting to force these interdependent features into separate diagnostic subscales risks over-interpreting shared symptom variance. Consequently, the total score on both the ADSS-13 and ADSS-23 is conceptualized as an aggregate indicator of depressive symptom

burden rather than a reflection of a homogeneous, singular latent entity. The robust inverse correlation with the WHO-5 Well-Being Index further substantiates that elevated aggregate scores align with marked decrements in positive emotional and functional well-being.

Methodological insights: Integrative selection and robust scoring

A central contribution of this investigation is the empirical evaluation of alternative item-reduction strategies. Traditional psychometric reduction often relies strictly on maximizing statistical metrics (e.g., top factor loadings or item-total correlations), which frequently leads to "bloated specific" scales that over-sample narrowly worded sadness items while discarding peripheral vegetative or interpersonal symptoms. By contrasting data-driven candidates against integrative configurations, our findings demonstrate that network bridge expected influence provided critical insights for identifying symptoms (such as sleep disturbance, appetite changes, and cognitive fatigue) that actively bridge distinct clinical domains.

Importantly, our evaluation of alternative scoring frameworks confirms a classic psychometric principle – often termed the Dawes or Wilks phenomenon: [12, 13] equal unit weighting performs virtually as well as, or better than, empirically optimized weights in real-world cross-validation. Neither qualitative clinical weighting nor regularized ridge logistic regression provided a meaningful or consistent diagnostic advantage over simple unweighted summation. For primary care and educational settings, retaining a simple arithmetic sum dramatically lowers barriers to implementation, minimizes clinician calculation errors, and preserves direct interpretability.

Diagnostic utility, clinical differentiation, and the three-tier triage model

A critical strength of this evaluation is the inclusion of an active clinical comparison group. Many existing adolescent screeners (such as the PHQ-9 modified for adolescents) have primarily

established discrimination against unselected community samples, where separation is artificially pronounced due to substantial baseline contrast. In real-world screening contexts, however, the primary diagnostic challenge lies in distinguishing adolescent depression from transdiagnostic distress, anxiety, trauma, and neurodevelopmental conditions. Both the ADSS-13 and ADSS-23 achieved high discrimination against other clinical presentations (AUCs of .848 and .846), maintaining performance comparable to community separation (.825 and .824).

Rather than forcing an artificial binary classification ("depressed" versus "non-depressed"), we implemented a three-tier decision model (low, intermediate, and high risk) grounded in uncertain-interval principles [14, 15]. Binary cut-offs often conceal diagnostic indeterminacy, resulting in high rates of false negatives or overwhelming specialist referral pathways. The intermediate score band (19–22 on the ADSS-13; 32–44 on the ADSS-23) designates an explicit "triage zone" where self-reported distress is elevated but diagnostically ambiguous. In practical screening workflows, this tier signals the need for structured clinical inquiry into symptom duration, situational triggers, and functional impairment, reserving immediate intensive psychiatric evaluation for the high-tier band (≥23 on ADSS-13; ≥45 on ADSS-23).

Subgroup variations and clinical safety considerations

Exploratory subgroup analyses highlighted crucial developmental nuances. Although overall classification agreement between scale lengths reached 90.2%, diagnostic sensitivity was noticeably lower among boys and adolescents aged 13–14 years. This pattern reflects well-documented developmental challenges in adolescent psychopathology: younger adolescents and males may experience less emotional differentiation, exhibit higher rates of somatic or externalizing manifestations (e.g., irritability, withdrawal) that are less readily captured by standardized depression inventories, or demonstrate reluctance to endorse explicit internalizing

items due to gender-role socialization. Consequently, low screening scores in early adolescence or among boys must be interpreted with clinical caution and contextualized alongside parent, teacher, or observational reports.

Crucially, suicide-related items (SI3, SI4, SI5 in the short form) must be decoupled from aggregate score interpretation. Suicidal ideation in adolescence can occur in the absence of severe syndromal depression, and composite sum scores risk diluting emergent safety threats. In accordance with clinical safety guidelines [16], any positive endorsement of suicide-related items necessitates immediate, independent clinical risk assessment, regardless of whether the total screening score falls into the low or intermediate band.

Limitations and future directions

Several methodological constraints warrant explicit consideration. First, all structural evaluations, item reductions, scoring comparisons, and threshold selections were conducted within a single developmental sample. Although internal validity was guarded through nested cross-validation and bootstrap resampling, apparent test performance remains susceptible to development optimism. Independent replication across external clinical and community cohorts is imperative before clinical implementation.

Second, the sample size within specific diagnostic and demographic subgroups (such as nonbinary youth, clinical males, and specific care settings) precluded formal tests of measurement invariance and differential item functioning (DIF; [17]. Establishing whether item parameters function equivalently across gender and developmental stages remains an essential next step. Third, criterion diagnoses in the clinical cohort relied on clinical chart evaluations and intake assessments rather than uniform, blinded structured clinical interviews (e.g., K-SADS). Finally,

test–retest reliability, sensitivity to treatment-induced change, and the prospective longitudinal impact of screening on clinical outcomes require dedicated empirical evaluation.

Conclusions

The Adolescent Depressive Symptoms Screener offers a psychometrically grounded, ICD-11-aligned instrument available in both an abbreviated 13-item screening format (ADSS-13) and a comprehensive 23-item profile (ADSS-23). By coupling robust unit scoring with an explicit three-tier triage decision framework, the ADSS provides a practical, evidence-based solution that mitigates diagnostic indeterminacy and optimizes resource allocation in adolescent mental healthcare.

The ADSS-13 serves as an efficient, low-burden instrument for rapid primary care and school-based screening, while the ADSS-23 delivers a granular symptom profile for specialized clinical evaluation. Crucially, by integrating network-informed bridge items with mandatory decoupled suicidal-risk protocols, the screener balances pragmatic brevity with uncompromised clinical safety. Early conceptual frameworks and preliminary methodological validations supporting this architecture were presented at the 31st International Congress of Applied Psychology [36, 37]. Upon independent external validation across prospective longitudinal cohorts, the ADSS framework holds strong potential to enhance early detection, streamline clinical referral pathways, and improve diagnostic precision for adolescents presenting with depressive spectrum disorders.

Declarations

Ethics approval and consent to participate

Ethical approval was granted by the Ethics Committee of Rīga Stradiņš University (2-PĒK-4/209/2024). Written informed consent was obtained from all participants and/or their legal guardians prior to enrollment.

Consent for publication

Not applicable.

Availability of data and materials

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Competing interests

The authors declare that they have no competing interests.

Funding

This research was supported by the National Research Program project "Development of a new computer-based assessment method for measuring and monitoring adolescent mental health outcomes in post-COVID pandemic conditions" (No. 1-PB-2/8/2024).

Authors' contributions

ZG, VP, JK, IK, NB, and AS contributed to the study conception and design. Material preparation, data collection, and analysis were performed by ZG, VP, and JK. The first draft of the manuscript was written by ZG, and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Acknowledgements

Not applicable.

References

1. World Health Organization. Mental health of adolescents. 2025. https://www.who.int/news-room/fact-sheets/detail/adolescent-mental-health. Accessed 10 Feb 2026.
2. Christensen AP, Golino H, Silvia PJ. A psychometric network perspective on the validity and validation of personality trait questionnaires. Eur J Pers. 2020;34(6):1095–108. https://doi.org/10.1002/per.2265
3. Collins GS, Moons KGM, Dhiman P, Riley RD, Beam AL, Van Calster B, et al. TRIPOD+AI statement: Updated guidance for reporting clinical prediction models that use regression or machine learning methods. BMJ. 2024;385:e078378. https://doi.org/10.1136/bmj-2023-078378
4. Epskamp S, Cramer AOJ, Waldorp LJ, Schmittmann VD, Borsboom D. qgraph: Network visualizations of relationships in psychometric data. J Stat Softw. 2012;48(4):1–18. https://doi.org/10.18637/jss.v048.i04
5. Epskamp S, Borsboom D, Fried EI. Estimating psychological networks and their accuracy: A tutorial paper. Behav Res Methods. 2018;50(1):195–212. https://doi.org/10.3758/s13428-017-0862-1
6. Jones PJ, Ma R, McNally RJ. Bridge centrality: A network approach to understanding comorbidity. Multivar Behav Res. 2021;56(2):353–67. https://doi.org/10.1080/00273171.2019.1614898
7. National Institute for Health and Care Excellence. Depression in children and young people: Identification and management (NICE Guideline NG134). 2019. https://www.nice.org.uk/guidance/ng134. Accessed 10 Feb 2026.

8. American Educational Research Association, American Psychological Association, National Council on Measurement in Education. Standards for educational and psychological testing. Washington, DC: American Educational Research Association; 2014.
9. Ziegler M. Psychological test adaptation and development—How papers are structured and why. Psychol Test Adapt Dev. 2020;1(1):3–11. https://doi.org/10.1027/2698-1866/a000002
10. World Health Organization. Clinical descriptions and diagnostic requirements for ICD-11 mental, behavioural and neurodevelopmental disorders. 2024. https://www.who.int/publications/i/item/9789240077263. Accessed 10 Feb 2026.
11. Patton GC, Coffey C, Posterino M, Carlin JB, Wolfe R. Adolescent depressive disorder: A population-based study of ICD-10 symptoms. Aust N Z J Psychiatry. 2000;34(5):741–7. https://doi.org/10.1080/j.1440-1614.2000.00825.x
12. Rice F, Riglin L, Lomax T, Souter E, Potter R, Smith DJ, et al. Adolescent and adult differences in major depression symptom profiles. J Affect Disord. 2019;243:175–81. https://doi.org/10.1016/j.jad.2018.09.015
13. Miller AB, Esposito-Smythers C, Leichtweis RN. A short-term, prospective test of the interpersonal–psychological theory of suicidal ideation in an adolescent clinical sample. Suicide Life Threat Behav. 2016;46(3):337–51. https://doi.org/10.1111/sltb.12204
14. Johnson JG, Harris ES, Spitzer RL, Williams JB. The Patient Health Questionnaire for Adolescents: Validation of an instrument for the assessment of mental disorders among adolescent primary care patients. J Adolesc Health. 2002;30(3):196–204. https://doi.org/10.1016/S1054-139X(01)00333-0

15. Reynolds WM. Reynolds Adolescent Depression Scale: Professional manual. 2nd ed. Lutz: Psychological Assessment Resources; 2002.
16. Kovacs M. Children's Depression Inventory 2nd Edition (CDI 2): Technical manual. North Tonawanda: Multi-Health Systems; 2011.
17. Beck JS, Beck AT, Jolly JB, Steer RA. Beck Youth Inventories: Second edition for children and adolescents (BYI-II). San Antonio: Pearson; 2005.
18. Angold A, Costello EJ, Messer SC, Pickles A. Development of a short questionnaire for use in epidemiological studies of depression in children and adolescents. Int J Methods Psychiatr Res. 1995;5(4):237–49.
19. Radloff LS. The use of the Center for Epidemiologic Studies Depression Scale in adolescents and young adults. J Youth Adolesc. 1991;20(2):149–66. https://doi.org/10.1007/BF01537606
20. Boateng GO, Neilands TB, Frongillo EA, Melgar-Quiñonez HR, Young SL. Best practices for developing and validating scales for health, social, and behavioral research: A primer. Front Public Health. 2018;6:149. https://doi.org/10.3389/fpubh.2018.00149
21. Borsboom D. A network approach to psychopathology. World Psychiatry. 2017;16(1):5–13. https://doi.org/10.1002/wps.20375
22. Landsheer JA. Interval of uncertainty: An alternative approach for the determination of decision thresholds, with an illustrative application for the prediction of prostate cancer. PLoS One. 2016;11(11):e0166007. https://doi.org/10.1371/journal.pone.0166007
23. Bringmann LF, Elmer T, Epskamp S, Krause RW, Schoch D, Wichers M, et al. What do centrality measures measure in psychological networks? J Abnorm Psychol. 2019;128(8):892–903. https://doi.org/10.1037/abn0000446

24. Hallquist MN, Wright AGC, Molenaar PCM. Problems with centrality measures in psychopathology symptom networks: Why network psychometrics cannot escape psychometric theory. Multivar Behav Res. 2021;56(2):199–223. https://doi.org/10.1080/00273171.2019.1640103
25. Yusoff MSB. ABC of content validation and content validity index calculation. Educ Med J. 2019;11(2):49–54. https://doi.org/10.21315/eimj2019.11.2.6
26. World Health Organization, Regional Office for Europe. Wellbeing measures in primary health care: The DepCare project. Stockholm: WHO Regional Office for Europe; 1998.
27. Topp CW, Østergaard SD, Søndergaard S, Bech P. The WHO-5 Well-Being Index: A systematic review of the literature. Psychother Psychosom. 2015;84(3):167–76. https://doi.org/10.1159/000376585
28. Boļšakova V, Perepjolkina V. PVO–5 labbūtības indekss (1998. gada versija). Rīga: Rīgas Stradiņa universitātes Psiholoģijas laboratorija; 2023. https://www.rsu.lv/sites/default/files/imce/Dokumenti/testi_aptaujas/apraksts_pvo5_rev.pdf. Accessed 10 Feb 2026.
29. Boļšakova V. PVO-5 labbūtības indeksa adaptācija [Bachelor's thesis]. Rīga: Rīgas Stradiņa universitāte; 2024. https://dspace.rsu.lv/items/60c83f3c-7c46-4491-a16d-6121ba3f4434. Accessed 10 Feb 2026.
30. Allgaier A-K, Pietsch K, Frühe B, Prast E, Sigl-Glöckner J, Schulte-Körne G. Depression in pediatric care: Is the WHO-Five Well-Being Index a valid screening instrument for children and adolescents? Gen Hosp Psychiatry. 2012;34(3):234–41. https://doi.org/10.1016/j.genhosppsych.2012.01.007

31. Smolyakov YN. jYS: Jamovi Extended Functions [jamovi module]. 2025. https://doi.org/10.5281/zenodo.15597090
32. Horowitz LM, Bridge JA, Teach SJ, Ballard E, Klima J, Rosenstein DL, et al. Ask Suicide-Screening Questions (ASQ): A brief instrument for the pediatric emergency department. Arch Pediatr Adolesc Med. 2012;166(12):1170–6. https://doi.org/10.1001/archpediatrics.2012.1276
33. Landsheer JA. The clinical relevance of methods for handling inconclusive medical test results: Quantification of uncertainty in medical decision-making and screening. Diagnostics. 2018;8(2):32. https://doi.org/10.3390/diagnostics8020032
34. Perepjolkina V, Krone I, Koļesņikova J, Bezborodovs N, Gulbe Z, Stepens A. Developing optimized screening and comprehensive diagnostic tools for adolescent depression: A multi-methodological approach. Poster presentation at the 31st International Congress of Applied Psychology; 2026 July 21–25; Florence, Italy.
35. Putnick DL, Bornstein MH. Measurement invariance conventions and reporting: The state of the art and future directions. Int J Behav Dev. 2016;40(1):71–90. https://doi.org/10.1177/0165025415573349
36. Perepjolkina V, Krone I, Koļesņikova J, Bezborodovs N, Gulbe Z, Stepens A. Developing optimized screening and comprehensive diagnostic tools for adolescent depression: A multi-methodological approach [poster presentation]. In: 31st International Congress of Applied Psychology (ICAP 2026); 2026 Jul 21–25; Florence, Italy.
37. Krone I, Perepjolkina V, Koļesņikova J, Bezborodovs N, Gulbe Z, Stepens A. Development and psychometric validation of a multidimensional adolescent depression

scale: A content-driven and empirical approach [poster presentation]. In: 31st International Congress of Applied Psychology (ICAP 2026); 2026 Jul 21–25; Florence, Italy.

Supplementary Files

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

- SupplementaryMaterial1.docx