Measurement Properties, Longitudinal Measurement Invariance, and Short-Term Stability of the RCADS Major Depressive Disorder Subscale in Vietnamese Lower Secondary School Students

Duc Minh Dao, Hang Thi Cong, Hong Phuong Ta
Hanoi National University of Education, Hanoi, Vietnam.

This study evaluated the suitability of the RCADS-MDD for assessing depressive symptoms among 805 Vietnamese lower secondary school students by examining its internal structure, reliability, measurement invariance, associations with related constructs, short-term temporal stability, and measurement error. Confirmatory factor analyses supported a one-factor structure at both assessment waves, with good internal consistency, although average variance extracted remained below the conventional threshold. Longitudinal configural, metric, and threshold invariance were supported, indicating that scores could be meaningfully compared across the two assessment occasions. Measurement invariance was also supported across gender, grade level, and school at both waves, supporting the comparability of scores across key student subgroups. Correlations with related psychological constructs provided additional evidence for the interpretability of RCADS-MDD scores. Scores demonstrated moderate rank-order stability over the short retest interval. However, indices of absolute agreement, standard error of measurement, minimal detectable change, and Bland-Altman analysis indicated that small changes in individual scores may reflect measurement variability rather than meaningful change. Overall, the findings support the RCADS-MDD as a suitable measure for assessing depressive symptom severity and for group-level comparisons among Vietnamese lower secondary school students, while caution is warranted when interpreting small score changes at the individual level.

Author info: Correspondence should be sent to: Duc Dao Minh PhD,
Email: ducmd@hnue.edu.vn

North American Journal of Psychology, 2026, Vol. 28, No. 4, 2624-2655
© NAJP https://doi.org/10.65696/001c.173467

INTRODUCTION

Depression Among Lower Secondary School Students and the Need for Early Screening

Depression during adolescence can substantially affect emotional functioning, concentration, academic motivation, social relationships, and everyday activities. Early identification is challenging because manifestations such as fatigue, loss of interest, sleep difficulties, reduced energy, and poor concentration may be interpreted as ordinary changes associated with puberty or temporary reactions to life stressors. Consequently, students with clinically meaningful symptoms may remain undetected until their difficulties have substantially affected their academic functioning and social adjustment.

Lower secondary school students constitute a particularly vulnerable group because they undergo rapid physical, cognitive, emotional, and social changes simultaneously. In addition to the developmental tasks associated with puberty, students must adapt to increasing academic demands, frequent evaluation, achievement expectations, and changes in their relationships with parents, teachers, and peers. These conditions do not necessarily cause depression directly, but they may increase vulnerability or make internalizing symptoms more apparent.

In Vietnam, depressive symptoms among adolescents have been associated with academic pressure, family and school relationships, interpersonal violence, life satisfaction, and psychological resilience (Ho et al., 2022; Tran et al., 2025). In particular, pressure related to grades, academic workload, parental and teacher expectations, and examinations may coincide with sleep deprivation, fatigue, diminished interest, concentration difficulties, and negative self-evaluation (Ho et al., 2022; Tran et al., 2025). Meanwhile, Vietnamese adolescents' access to mental health services remains limited relative to actual need. Schools therefore represent an important setting for initial screening and referral of at-risk students for more comprehensive assessment.

Screening, however, is not equivalent to diagnosis. A screening instrument is useful only when its scores adequately represent the intended construct, demonstrate acceptable precision, and are interpreted appropriately for the population, purpose, and context of use (American Educational Research Association et al., 2014; Prinsen et al., 2018). Evidence supporting a scale in another country or age group does not guarantee that it will function equivalently among Vietnamese lower secondary school students (Mokkink et al., 2010a; Stevanovic et al., 2017). Its measurement properties must therefore be directly evaluated before its scores are used to identify risk, compare groups, or monitor change.

The RCADS-MDD and Gaps in the Measurement Evidence

The Revised Child Anxiety and Depression Scale was developed to assess anxiety- and depression-related symptom domains in children and adolescents. The full version includes subscales measuring separation anxiety, social anxiety, generalized anxiety, panic, obsessive-compulsive symptoms, and major depressive disorder. The 10-item Major Depressive Disorder subscale assesses emotional, cognitive, motivational, and somatic manifestations of depression, including sadness or emptiness, loss of interest, fatigue, reduced energy, sleep difficulties, changes in appetite, difficulty thinking clearly, and feelings of worthlessness (Chorpita et al., 2000, 2005).

The RCADS-MDD has several advantages for school-based screening because it is relatively brief, uses a self-report format, and captures internal experiences that parents or teachers may not always observe. The MDD subscale is also embedded within the same measurement system as the anxiety subscales, which is consistent with the frequent co-occurrence of anxiety and depression during adolescence (Chorpita et al., 2000, 2005; Piqueras et al., 2017). Nevertheless, the RCADS-MDD reflects symptom severity or preliminary risk only. A high score does not, by itself, establish a diagnosis and should be considered alongside functional impairment, professional interviews, and other sources of information (Chorpita et al., 2005; Radez et al., 2021).

From a psychometric perspective, the quality of an instrument cannot be established solely on the basis of Cronbach's alpha or an adequately fitting CFA model. The validity of its scores depends on a broader body of evidence supporting their interpretation and use within a particular context. Important properties include structural validity, internal consistency, reliability, measurement error, relationships with other variables, and measurement invariance (American Educational Research Association et al., 2014; Mokkink et al., 2010b, 2018; Prinsen et al., 2018). These properties are not interchangeable: a high alpha does not establish unidimensionality, and a high test-retest correlation does not guarantee agreement in absolute scores across measurement occasions (de Vet et al., 2006; Mokkink et al., 2010a).

International studies have generally supported the structure and reliability of the RCADS-47, RCADS-25, and shortened versions across multiple languages and populations (Giannopoulou et al., 2022; Grothus et al., 2023, 2025; Lisøy et al., 2022; Petrič et al., 2024). The review by Piqueras et al. (2017) also indicated that the RCADS has a relatively strong reliability foundation. At the same time, multinational studies and studies across gender and age groups have emphasized the need to establish the equivalence of the measurement structure before comparing scores across groups (Donnelly et al., 2019; Stevanovic et al., 2017).

Several gaps nevertheless remain in the existing evidence. First, most studies have evaluated the full RCADS or its shortened versions, whereas relatively few have focused specifically on the 10-item MDD subscale. Strong measurement properties of the full instrument do not automatically demonstrate that the MDD total score can be interpreted as a sufficiently precise unidimensional indicator of depressive symptoms (American Educational Research Association et al., 2014; Mokkink et al., 2010b).

Second, most studies have used cross-sectional designs. CFA conducted at a single time point cannot determine whether an instrument retains the same measurement meaning over time. When scores are used to assess change, configural, metric, and threshold invariance should be

examined. Threshold invariance is particularly important because it provides a basis for interpreting mean differences as changes in the underlying construct rather than changes in item functioning (Chen, 2007; Putnick & Bornstein, 2016). A recent review showed that invariance across gender, age, and time is still frequently neglected in depression measures for children and adolescents (Schlechter et al., 2024). Studies evaluating the RCADS across cultural, gender, and age groups have similarly demonstrated the need to establish invariance before comparing scores or means across groups (Donnelly et al., 2019; Stevanovic et al., 2017).

Third, test-retest studies often report only a correlation coefficient or a single ICC without clearly distinguishing mean-level change, rank-order stability, absolute agreement, and measurement error. Pearson's correlation and consistency ICC primarily reflect relative stability, whereas an absolute-agreement ICC also accounts for systematic differences. The SEM and MDC95 quantify score uncertainty, while Bland-Altman analysis evaluates bias and the limits of agreement between two measurement occasions (de Vet et al., 2006; Koo & Li, 2016). A single test-retest coefficient is therefore insufficient to determine whether an instrument is suitable for monitoring change at the individual-student level (de Vet et al., 2006; Mokkink et al., 2010a).

Fourth, very few studies have examined the RCADS-MDD under two school conditions that differ in psychological relevance. During studies, manifestations such as fatigue, sleep difficulties, loss of interest, concentration problems, or negative self-evaluation may become more pronounced and may subsequently decline after the immediate pressure has ended. However, score differences can be interpreted validly only when the scale retains the same structure and measurement meaning across the two occasions (Chen, 2007; Putnick & Bornstein, 2016). Finally, studies in Vietnam have mainly focused on symptom prevalence and associated factors, whereas evidence concerning the structural validity, longitudinal invariance, absolute agreement, and measurement error of screening instruments remains limited (Ho et al., 2022; Tran et al., 2025).

The Present Study

The present study assessed the same 805 Vietnamese lower secondary school students at two time points. The study extends previous research in five ways: (1) it focuses specifically on the 10-item RCADS-MDD rather than drawing inferences from the full RCADS or its shortened versions; (2) it evaluates factor structure, item properties, and internal consistency at both assessment waves; (3) it tests longitudinal measurement invariance at the configural, metric, and threshold levels; (4) it examines measurement invariance across gender, grade, and school; and (5) it distinguishes mean-level change, rank-order stability, absolute agreement, and measurement error.

Based on previous evidence, a one-factor model and good internal consistency were expected at both time points (Chorpita et al., 2005; Piqueras et al., 2017). The RCADS-MDD was also expected to demonstrate longitudinal invariance. Mean symptom levels were expected to be lower after the studies and examinations, while the relative ranking of students was expected to remain acceptably stable.

METHODS

Participants

The study was conducted among students in Grades 6-9 from three lower secondary schools located in the inner-city, suburban, and peri-urban areas of Hanoi, Vietnam. The schools were selected using a random sampling procedure across these three geographical settings to capture variation in students' educational and sociocultural environments. The inclusion of schools from different settings was intended to account for contextual diversity; however, it does not establish national representativeness. The mean age of the participants was 11.99 years (SD = 1.24). At Wave 1, 853 students who met the inclusion criteria completed the questionnaire and were assigned anonymous identification codes to facilitate longitudinal data matching. Fourteen days later, all Wave 1 participants were invited to complete the questionnaire again under identical conditions, and 853 students participated at Wave 2. The two waves of data were subsequently matched using the identification codes.

Of these, 48 cases were excluded because cross-wave matching could not be verified with sufficient certainty. Consequently, the final test-retest sample comprised 805 students with valid, matched data at both waves.

As part of the broader assessment, 11 students were identified as showing elevated symptom levels and potential clinical risk indicators based on a combined review of scores from multiple instruments, including the 3 scales: DSM-5 Level 2-Anxiety-Child Age 11-17 measure, the RCADS anxiety and major depressive disorder subscales (*T*-score>65), and indicators of school connectedness and perceived social support (*T*-score<40). These findings suggested that the students might require further, in-depth clinical assessment to determine whether their symptoms met the criteria for a mental disorder. However, no independent clinical diagnostic interviews were conducted; therefore, these students were considered potentially at risk rather than classified as having confirmed mental disorders.

Table 1 Characteristics of Participants (n=805)

Variable	**Category**	n	%
Gender	Male	331	41.1
	Female	474	58.9
Grade	Grade 6	271	33.7
	Grade 7	185	23.0
	Grade 8	206	25.6
	Grade 9	143	17.7
Potential clinical risk indicators	Non-clinical	794	98.6
	Clinical	11	1.4
Residential area	Urban	323	40.1
	Peri-urban	218	27.1
	Suburban	264	32.8

Note. M = mean; SD = standard deviation.

Research Design and Procedure

Research Design

The study employed a two-wave test-retest design involving lower secondary school students recruited from three schools representing three

distinct geographical areas of Hanoi. The study examined the factor structure, score reliability, measurement invariance across time, grade level, gender, and school, as well as score stability and changes between the two assessment occasions for the RCADS-MDD subscale. Together, these analyses were intended to provide a comprehensive evaluation of the subscale's short-term temporal stability and measurement properties among Vietnamese lower secondary school students.

Procedure

The study was conducted using a standardized sequence of procedures. First, based on the objective of evaluating the measurement properties and short-term stability of the RCADS-MDD, the research team developed a two-wave design. Next, schools that met the eligibility criteria were invited to participate. After approval was obtained, the research team collaborated with school administrators to determine the scope of the survey, identify the homeroom teachers involved, establish the assessment schedule, and arrange the conditions for classroom administration. School staff who assisted with data collection were trained in the standardized administration procedures, including providing instructions to students, managing the assessment time, monitoring questionnaire completion, distributing the questionnaires, and collecting them after completion. Both Wave 1 and Wave 2 were administered using paper-and-pencil self-report questionnaires during regular class periods.

Data Analysis

Data were analyzed using SPSS, AMOS version 31.0, and jamovi version 2.7.38. Descriptive statistics were used to examine item distributions, floor and ceiling effects, corrected item-total correlations, and inter-item correlations. Internal consistency was evaluated using Cronbach's alpha, McDonald's omega, and corrected item-total correlations. Composite reliability (CR), average variance extracted (AVE), and mean inter-item correlations were used to assess reliability and convergent validity.

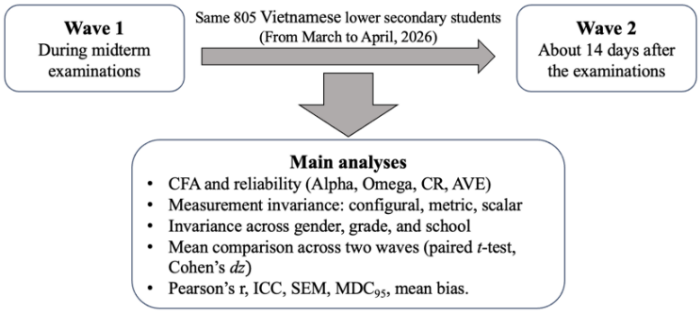
Confirmatory factor analyses (CFA) were conducted separately at each assessment wave using AMOS version 31.0 with maximum

likelihood (ML) estimation to evaluate the one-factor structure of the RCADS-MDD. The CFA results were also cross-checked using jamovi. Model fit was evaluated using the chi-square-to-degrees-of-freedom ratio (χ^2/df), comparative fit index (CFI), Tucker-Lewis index (TLI), root mean square error of approximation (RMSEA) with its 90% confidence interval, and standardized root mean square residual (SRMR).

Measurement invariance was examined using jamovi version 2.7.38, with the four-category RCADS-MDD items specified as ordinal indicators. Configural, metric, and threshold invariance were evaluated across the two assessment waves and across gender, grade level, and school. At the configural level, the same factor structure was specified across occasions or groups. Metric invariance was examined by constraining corresponding factor loadings to equality, followed by threshold invariance, in which corresponding item thresholds were constrained to equality while retaining the preceding constraints. Changes in CFI (ΔCFI $\leq$ .010) and RMSEA (ΔRMSEA $\leq$ .015) were used as approximate criteria for evaluating measurement invariance (Chen, 2007; Putnick & Bornstein, 2016).

Test-retest reliability and measurement stability were examined using paired-samples *t* tests, Cohen's *d*_z, Pearson's *r*, the intraclass correlation coefficient (ICC), the standard error of measurement (SEM), the minimal detectable change at the 95% confidence level (MDC95), and Bland-Altman analyses (de Vet et al., 2006; Koo & Li, 2016). Pearson correlations between the RCADS-MDD, other RCADS subscales, and selected external subscales were also examined to evaluate theoretically expected relationships. Statistical significance was set at $p < .05$.

Figure 1. Research Participants and Psychometric Evaluation of the RCADS-MDD Subscale Across Waves



RESULTS

The results are presented in terms of item characteristics, reliability, factorial validity, measurement invariance, and short-term stability of the RCADS-MDD across the two waves.

Item Properties and Reliability

Descriptive statistics were examined to evaluate the distributional characteristics and item functioning of the RCADS-MDD across the two waves.

Table 2 Item Descriptive Statistics of the RCADS-MDD Across Waves

Item	Item content	M	SD	Skew	Kurtosis	Floor %	Ceiling %
2	I feel sad or empty	2.23	.86	.314	-.501	19.6	8.1
6	Nothing is much fun anymore	1.94	.97	.738	-.474	40.2	9.3
11	I have trouble sleeping	1.90	.97	.806	-.413	43.4	9.3
15	I have problems with my appetite	1.65	.86	1.194	.599	55.2	4.8
19	I have no energy for things	1.75	.79	.898	.420	42.7	3.5
21	I am tired a lot	2.27	.90	.325	-.641	20.1	10.8
25	I cannot think clearly	2.00	.84	.593	-.142	28.9	5.8
29	I feel worthless	1.79	.91	.988	.115	46.8	7.0

40	I feel like I don't want to move	1.66	.86	1.173	.541	54.9	4.7
47	I feel restless	1.70	.83	1.086	.548	49.6	4.6
	Total score	***18.90***	***5.75***	***.636***	***.016***	***3.5***	***.1***
2	I feel sad or empty	2.10	.84	.485	-.250	24.1	6.5
6	Nothing is much fun anymore	1.86	.91	.795	-.297	43.2	6.5
11	I have trouble sleeping	1.67	.89	1.221	.585	55.5	6.1
15	I have problems with my appetite	1.51	.78	1.526	1.706	63.5	3.4
19	I have no energy for things	1.70	.81	.994	.347	48.8	3.5
21	I am tired a lot	2.17	.91	.358	-.686	25.7	8.6
25	I cannot think clearly	1.87	.84	.712	-.125	37.6	4.6
29	I feel worthless	1.74	.88	1.030	.189	50.2	5.6
40	I feel like I don't want to move	1.58	.82	1.342	1.051	59.5	3.9
47	I feel restless	1.59	.79	1.258	.976	56.6	3.2
	Total score	***17.78***	***5.82***	***.798***	***.334***	***6.7***	***.1***

Note. *M = Mean; SD = Standard deviation.*

Table 3. Item Analysis and Inter-Item Correlations of the RCADS-MDD Across Waves

Wave	**Item**	**CITC**	**Correlation**								
			2	6	11	15	19	21	25	29	40
Wave 1	2	.609	-								
	6	.623	.549	-							
	11	.405	.259	.289	-						
	15	.401	.253	.265	.307	-					
	19	.635	.433	.460	.268	.305	-				
	21	.678	.535	.510	.347	.331	.562	-			
	25	.560	.403	.408	.283	.234	.452	.423	-		
	29	.562	.439	.424	.209	.212	.468	.421	.404	-	
	40	.468	.285	.330	.198	.250	.327	.364	.328	.340	-
	47	.618	.428	.442	.306	.279	.462	.477	.401	.443	.408
	2	.660	-								

Wave 2	6	.669	.569	-							
	11	.471	.302	.365	-						
	15	.433	.265	.301	.409	-					
	19	.669	.524	.550	.328	.300	-				
	21	.697	.546	.541	.405	.337	.587	-			
	25	.595	.468	.467	.279	.251	.448	.472	-		
	29	.578	.474	.485	.238	.239	.453	.451	.443	-	
	40	.545	.405	.385	.285	.312	.423	.408	.399	.343	-
	47	.666	.503	.448	.386	.339	.493	.523	.469	.478	.467

Note. CITC: Corrected Item-Total Correlation; p<0.01

Tables 2 and 3 provide favorable initial evidence regarding the distributional characteristics and item functioning of the RCADS-MDD in the lower secondary school student sample. All 10 items of the RCADS-MDD subscale showed variability, with low ceiling effects, and the total score maintained relatively stable dispersion across the two waves. However, several items, particularly Items 11, 15, 40, and 47, exhibited noticeable floor effects. At Wave 2, the floor effect for Item 15 reached 63.5%, suggesting more limited sensitivity at the lower end of the symptom continuum. This may limit the item's ability to distinguish between students with low symptom levels and detect small changes in symptoms. Nevertheless, total-score floor effects remained low, at 3.5% at Wave 1 and 6.7% at Wave 2. Corrected item-total correlations ranged from .401 to .678 at Wave 1 and from .433 to .697 at Wave 2, indicating acceptable discrimination for all items. No item showed evidence that its removal would meaningfully improve Cronbach's alpha, while the inter-item correlations were all positive and did not indicate substantial redundancy. Overall, these findings support retaining all 10 items of the RCADS-MDD. Subsequent CFA and measurement invariance analyses will further examine the psychometric validity of this subscale.

The item-level response distributions (Appendix 1) showed a general concentration at the lower response categories, which became more pronounced at Wave 2. Floor effects were particularly evident for Items 15, 40, 47, and 11, whereas ceiling effects were minimal across all items. The overall shift toward lower response categories at Wave 2 was consistent with the observed decrease in RCADS-MDD scores, although

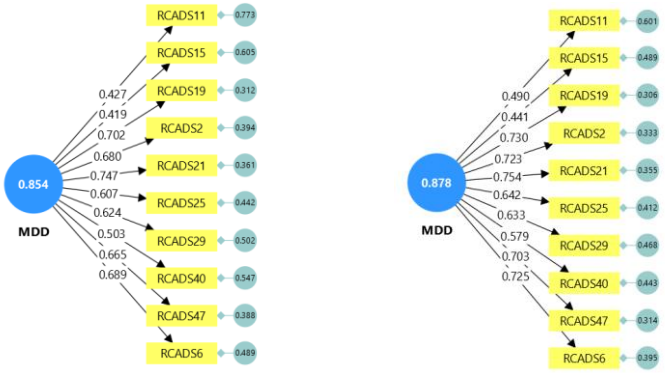
the marked floor effect for some items suggests reduced sensitivity at the lower end of the symptom continuum.

Table 4. Reliability and Convergent Validity of the RCADS-MDD Across Waves

Index	Wave 1	Wave 2
Cronbach α	.851	.875
McDonald ω	.853	.877
Composite Reliability	.854	.878
AVE	.380	.423
Mean Inter-item Correlation	.367	.413

Note. *AVE = Average Variance Extracted*

Figure 2. Composite Reliability of the RCADS-MDD Subscale Across Waves



Wave 1 *Wave 2*

The results presented in Table 4 indicate that the RCADS-MDD demonstrated good internal consistency reliability at both waves. Cronbach's α, McDonald's ω, and composite reliability ranged from .851 to .854 at Wave 1 and from .875 to .878 at Wave 2, indicating consistent reliability across different estimation methods and a slight improvement at Wave 2. The mean inter-item correlations were .367 and .413, respectively, reflecting an appropriate degree of internal relatedness among the items. However, the AVE values were only .380 at Wave 1 and .423 at Wave 2, both below the conventional .50 threshold, indicating that the MDD factor accounted for less than half of the average variance in the items. These findings suggest that good internal consistency does not

necessarily indicate adequate convergent validity. In particular, the relatively low AVE values, together with the comparatively weaker factor loadings of Items 11 and 15, warrant caution in interpreting the extent to which the items converge on the common depression factor. Overall, the findings support the reliability of the RCADS-MDD total score, although the convergence of the items on the common factor was not optimal and should be further evaluated in conjunction with factor loadings, CFA model fit, and measurement invariance.

Factorial Validity

Table 5 Standardized Factor Loadings Across Waves

Item	Item	Wave 1 λ	Wave 2 λ	Difference
2	I feel sad or empty	.680	.723	.043
6	Nothing is much fun anymore	.689	.725	.036
11	I have trouble sleeping	.427	.490	.063
15	I have problems with my appetite	.419	.441	.022
19	I have no energy for things	.702	.730	.028
21	I am tired a lot	.747	.754	.007
25	I cannot think clearly	.607	.642	.035
29	I feel worthless	.624	.633	.009
40	I feel like I don't want to move	.503	.579	.076
47	I feel restless	.665	.703	.038

Note. *λ = standardized factor loading.*

Table 6 Confirmatory Factor Analysis Model Across Waves

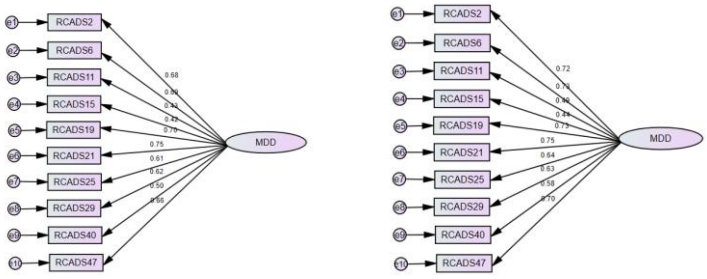
Wave	χ^2(df)	χ^2/df	CFI	TLI	RMSEA [90 % CI]	SRMR
Wave 1	118.077 (35)	3.374	.966	.956	.054 [.044, .065]	.030
Wave 2	148.335 (35)	4.238	.961	.950	.063 [.053, .074]	.035

Note. *CFI = comparative fit index; TLI = Tucker-Lewis index; RMSEA = root mean square error of approximation; CI = confidence interval; SRMR = standardized root mean square residual.*

Tables 5 and 6 provide generally favorable evidence for the unidimensional structure of the RCADS-MDD at both waves. All standardized factor loadings were positive, ranging from .419 to .747 at

Wave 1 and from .441 to .754 at Wave 2. Most items showed moderate to strong loadings, indicating that they reflected the common MDD factor with reasonable consistency. Item 21 was the strongest indicator, whereas Items 11 and 15 had the lowest loadings at both time points, suggesting that difficulties with sleep and changes in appetite were explained to a lesser extent by the common factor. The model demonstrated good fit at Wave 1, CFI = .966, TLI = .956, RMSEA = .054, and SRMR = .030, and acceptable fit at Wave 2, CFI = .961, TLI = .950, RMSEA = .063, and SRMR = .035. Overall, the findings support the use of a unidimensional total score, however, the relatively low loadings of Items 11 and 15 suggest that the factor structure was not entirely homogeneous and should be considered alongside evidence of content validity, item-level analyses, and measurement invariance.

Figure 3 Confirmatory Factor Analysis of the RCADS-MDD Subscale Across Waves



Wave 1 *Wave 2*

Measurement Invariance

Table 7 Longitudinal Measurement Invariance of the RCADS-MDD Across Waves

Model	χ^2 (df)	CFI	RMSEA [95% CI]	SRMR	ΔCFI	ΔRMSEA
Configural	288 (70)	.982	.062 [.055, .070]	.032	-	-
Metric	216 (79)	.989	.046 [.039, .054]	.034	+.007	-.016

Threshold	259 (98)	.987	.045 [.039, .052]	.038	-.002	-.001

Note. *CFI = comparative fit index; TLI = Tucker-Lewis index; RMSEA = root mean square error of approximation; CI = confidence interval; SRMR = standardized root mean square residual; ΔCFI = change in CFI; ΔRMSEA = change in RMSEA.*

The results of the longitudinal measurement invariance analysis are presented in Table 7. The configural model showed an acceptable-to-good fit to the data across the two assessment occasions (scaled χ^2 (70) = 288, CFI = .982, RMSEA = .062, 95% CI [.055, .070], SRMR = .032). Constraining the factor loadings to equality across occasions yielded improved approximate fit, with CFI increasing to .989 and RMSEA decreasing to .046 (95% CI [.039, .054]; SRMR = .034; scaled χ^2(79) = 216). The threshold invariance model, which retained the loading constraints and additionally constrained corresponding item thresholds to equality, also showed good fit (scaled χ^2(98) = 259, CFI = .987, RMSEA = .045, 95% CI [.039, .052], SRMR = .038). Relative to the metric model, CFI decreased by .002 and RMSEA decreased by .001, indicating minimal changes in approximate model fit. Overall, the changes in CFI and RMSEA were within the prespecified criteria for evaluating measurement invariance, providing support for longitudinal invariance across the two assessment occasions.

Table 8. Measurement Invariance of the RCADS-MDD Across Gender, Grade, and School

Group	Wave	Model	χ^2 (df)	CFI	RMSEA [95% CI]	SRMR	ΔCFI	ΔRMSEA
Gender	W1	*Configural*	171 (70)	.983	.060 [.049, .071]	.037	-	-
		Metric	156 (79)	.987	.049 [.038, .061]	.041	+.004	-.011
		Threshold	180 (98)	.986	.046 [.035, .056]	.050	-.001	-.003
	W2	*Configural*	224 (70)	.979	.074 [.063, .085]	.040	-	-

		Metric	182 (79)	.986	.057 [.046, .068]	.043	+.007	-.017
		Threshold	239 (98)	.981	.060 [.050, .070]	.061	-.005	+.003
Grade	W1	*Configural*	276 (140)	.977	.070 [.058, .082]	.046	-	-
		Metric	296 (167)	.979	.062 [.050, .074]	.057	+.002	-.008
		Threshold	363 (224)	.977	.056 [.045, .066]	.079	-.002	-.006
	W2	*Configural*	283 (140)	.980	.071 [.059, .083]	.045	-	-
		Metric	305 (167)	.981	.064 [.053, .076]	.057	+.001	-.007
		Threshold	367 (224)	.980	.057 [.046, .067]	.085	-.001	-.007
School	W1	*Configural*	207 (105)	.982	.060 [.048, .072]	.039	-	-
		Metric	210 (123)	.985	.051 [.039, .063]	.047	+.003	-.009
		Threshold	252 (161)	.984	.046 [.035, .057]	.063	-.001	-.005
	W2	*Configural*	251 (105)	.980	.072 [.061, .084]	.041	-	-
		Metric	223 (123)	.986	.055 [.043, .067]	.047	+.006	-.017
		Threshold	259 (161)	.987	.048 [.037, .058]	.058	+.001	-.007

Note. *CFI = comparative fit index; TLI = Tucker-Lewis index; RMSEA = root mean square error of approximation; CI = confidence interval; SRMR = standardized root mean square residual; ΔCFI = change in CFI; ΔRMSEA = change in RMSEA.*

The measurement invariance results were generally consistent across gender, grade level, and school at both waves. The configural models showed acceptable fit in all comparisons, suggesting that the same basic factor structure was maintained across groups. Metric invariance was also supported, as constraining factor loadings did not meaningfully worsen model fit.

Threshold invariance was likewise supported across gender, grade level, and school at both waves. Changes in CFI ranged from -.005 to +.001, while changes in RMSEA ranged from -.007 to +.003, indicating little deterioration in model fit when item thresholds were constrained to equality. These findings suggest that the relevant measurement parameters were sufficiently comparable across the examined groups.

One result warrants some caution. For grade level at Wave 2, the Threshold model had an SRMR of .085, which was higher than that of the other models. However, the corresponding CFI (.980), RMSEA (.057), ΔCFI (-.001), and ΔRMSEA (-.007) did not indicate a meaningful deterioration in overall model fit. Taken together, the results support configural, metric, and threshold invariance across gender, grade level, and school at both waves, although the fit of the Wave 2 grade-level threshold model was slightly less favorable according to SRMR.

Mean-Level Change, Short-Term Stability, Agreement, and Measurement Error

Table 9. Mean-Level Change, Temporal Stability, Absolute Agreement, and Measurement Error Across Two Waves

Domain	Statistic	Estimate	CI / Additional information
Descriptive statistics	Wave 1, M (SD)	18.9 (5.76)	N = 805
	Wave 2, M (SD)	17.8 (5.80)	N = 805
Mean-level change	Mean difference (W1 - W2)	1.14	[0.835, 1.44]
	Paired *t*	7.41	df = 804, $p < .001$
	Cohen's *dz*	.261	[.191, .331]
Temporal stability	Pearson's *r*	.717	[.682, .749], $p < .001$
Temporal consistency	ICC (3,1), consistency	.717	[.688, .744]

Absolute agreement	ICC (2,1), agreement	.704	[.659, .743]
	CCC	.704	[.668, .737]
Bland-Altman agreement	Mean bias (W1 - W2)	1.14	95% CI [0.834, 1.44]
	Lower 95% LoA	-7.39	90% CI [-7.733, -7.05]
	Upper 95% LoA	9.66	90% CI [9.321, 10.00]
Measurement error	SEM	3.07	-
	MDC95	8.52	-

Note. *W1 = Wave 1; W2 = Wave 2; CI = Confidence Interval; ICC = intraclass correlation coefficient; CCC = concordance correlation coefficient; SEM = standard error of measurement; MDC95 = minimal detectable change at 95% confidence; LoA = limits of agreement.*

Table 9 shows that RCADS-MDD scores decreased between the two assessments, with a paired mean difference of 1.14 points (95% CI [0.835, 1.44], $p < .001$). The standardized mean change was small (Cohen's $dz = .261$), indicating that the average reduction between the two time points was modest.

Scores showed a reasonably consistent relative ranking across the two assessments, as reflected by Pearson's $r = .717$ and ICC (3,1) = .717. Absolute agreement was slightly lower, with both ICC (2,1) and the concordance correlation coefficient (CCC) equal to .704. These findings indicate that relative differences between individuals were fairly stable over time, although absolute scores were not fully equivalent across the two occasions.

The Bland-Altman analysis showed a mean bias of 1.14 points, consistent with the lower mean score at Wave 2 than at Wave 1. The 95% limits of agreement ranged from -7.39 to 9.66, indicating considerable variability in individual score differences around the average change. The SEM was 3.07 points and the MDC95 was 8.52 points. Thus, although the mean score decreased significantly at the group level, small changes in an individual student's score may reflect measurement error rather than meaningful change. An individual change exceeding the MDC95 would indicate change beyond expected measurement error at the 95% confidence level; however, this threshold does not establish clinical

significance. Therefore, the findings support short-term score stability and agreement but do not establish clinically meaningful individual change.

Table 10. Correlations of the RCADS-MDD Score With RCADS Subscale, MSPSS and ESSA Scores Across Waves

MDD	RCADS Subscales					MSPSS		ESSA
	SP	**PD**	**SAD**	**GAD**	**OCD**	**MSPSS-Family**	**MSPSS-Friends**	**ESSA-PFS**
Wave 1	.708	.742	.513	.700	.612	-.199	-.346	.545
Wave 2	.672	.728	.707	.537	.665	-.208	-.325	.531

Note. *Values are Pearson's r. RCADS-MDD = Revised Child Anxiety and Depression Scale-Major Depressive Disorder subscale; SP = Social Phobia; PD = Panic Disorder; SAD = Separation Anxiety Disorder; GAD = Generalized Anxiety Disorder; OCD = Obsessive-Compulsive Disorder; MSPSS-Family = Multidimensional Scale of Perceived Social Support-Family (item 3,4,8,11); MSPSS-Friends = Multidimensional Scale of Perceived Social Support-Friends (item 6,7,9,12). ESSA-PFS = Educational Stress Scale for Adolescents- Pressure From Study (item 4,5,6,11). All correlations were statistically significant at* $p < .001$.

RCADS-MDD scores were moderately to strongly positively correlated with the RCADS anxiety subscales across both waves (rs = .513-.742 at Wave 1 and .537-.728 at Wave 2). Negative correlations were observed with perceived social support from family and friends (rs = −.199 to −.346), whereas positive correlations were found with academic pressure (rs = .531-.545). All correlations were statistically significant ($p < .001$). The direction and magnitude of these associations were generally consistent with theoretical expectations across both waves. These findings provide preliminary support for the expected patterns of association of RCADS-MDD scores with related psychological and contextual variables. However, correlations with other RCADS subscales should be interpreted cautiously because they were derived from the same instrument and self-report method. The absence of an independent depression measure or clinician-based assessment also limits the conclusions that can be drawn regarding convergent validity.

DISCUSSION

Overview of the Main Findings

This study was grounded in the contemporary view that the validity of a scale should be established through an integrated body of evidence linked to the target population, context, and intended use, rather than inferred from any single psychometric index (American Educational Research Association et al., 2014; Mokkink et al., 2010b; Prinsen et al., 2018).

The study yielded five main findings. First, the RCADS-MDD retained a one-factor structure at both waves and showed good internal consistency, although AVE remained below .50. Second, longitudinal configural, metric, and threshold invariancewere established across the two assessment occasions. Third, metric and threshold invariance were also supported across gender, grade level, and school at both waves. Fourth, RCADS-MDD scores showed a small mean-level decrease between waves, together with moderate short-term rank-order stability and slightly lower absolute agreement. Fifth, the Bland-Altman limits of agreement, SEM, and MDC95 indicated that variability and measurement error in individual score changes were substantially greater than the average mean-level change observed in the sample.

Internal Structure, Item Performance, and Score Reliability

The findings support the interpretation of the RCADS-MDD total score as a unidimensional indicator of self-reported depressive symptoms among lower secondary school students. The one-factor model demonstrated good fit at Wave 1 and acceptable fit at Wave 2; all factor loadings were positive, and most were moderate to strong. Replication of this structure within the same group of students at two different time points strengthens the evidence for the structural validity of the MDD total score and is consistent with the original theoretical framework and previous studies of the RCADS across diverse cultural contexts (Chorpita et al., 2000, 2005; Donnelly et al., 2019; Giannopoulou et al., 2022; Lisøy et al., 2022; Piqueras et al., 2017; Stevanovic et al., 2017). An additional contribution of the present study is the direct evaluation of the 10-item

MDD subscale as an independently interpretable score, rather than inferring its measurement properties from the full RCADS or from abbreviated versions. From a contemporary measurement perspective, validity evidence must be linked to the specific score, population, and intended use under consideration (American Educational Research Association et al., 2014; Mokkink et al., 2010b; Prinsen et al., 2018).

However, the evidence based on internal structure was not entirely uniform. An AVE below .50 indicates that the common factor did not explain the majority of item variance, although composite reliability and most factor loadings remained favorable. Items 11 and 15 showed the lowest factor loadings at both waves, suggesting that symptoms related to sleep and appetite converged less strongly with the MDD factor than did the remaining items. One plausible explanation is that these vegetative symptoms may be influenced not only by depression but also by developmental characteristics, physical health, or daily routines; however, this interpretation requires confirmation using additional evidence.

In contrast, Cronbach's alpha, McDonald's omega, composite reliability, and corrected item-total correlations consistently indicated good internal consistency, while deleting any single item did not meaningfully improve reliability. Accordingly, the current evidence does not support removing Item 11 or Item 15 solely on the basis of their factor loadings or the AVE results. The principal limitation concerns the uneven degree of item convergence rather than a lack of consistency in the scale as a whole. Any decision to shorten the subscale should therefore be based on cross-validation, DIF or IRT analyses, and additional evidence of content validity, rather than on statistical optimization within a single sample.

Longitudinal and Group Measurement Invariance

The RCADS-MDD showed generally consistent measurement properties over time and across student groups. Longitudinal analyses supported configural, metric, and threshold invariance across the two assessment occasions, indicating that the one-factor structure, factor loadings, and item thresholds remained sufficiently comparable from the

first to second period approximately 14 days. This allows changes between the two waves to be interpreted within a common measurement framework (Chen, 2007; Putnick & Bornstein, 2016). The finding is also noteworthy because longitudinal invariance has been examined relatively infrequently in depression measures for children and adolescents (Schlechter et al., 2024).

A similar pattern was found across gender, grade level, and school. Metric and threshold invariance were supported at both waves, with only small changes in CFI and RMSEA as equality constraints were added. These findings suggest that the RCADS-MDD measured the depressive construct in broadly comparable ways across the groups examined and provide a reasonable basis for group comparisons within this sample. This pattern is consistent with recent studies reporting measurement comparability of RCADS-based measures across demographic and cultural groups, while also reinforcing the point that gender- and age-related invariance should not be assumed for child and adolescent depression measures (Cervin et al., 2022; Schlechter et al., 2024; Lu et al., 2024).

One point warrants some caution. The Wave 2 threshold invariance model across grade levels had an SRMR of .085. Although this value was less favorable than those observed in the other models, the corresponding CFI, RMSEA, and changes in fit did not indicate a meaningful deterioration in model fit. Overall, the findings support longitudinal and group measurement invariance of the RCADS-MDD in the present sample.

Short-Term Stability, Agreement, and Interpretation of Change

The longitudinal findings indicate a small mean-level decrease in RCADS-MDD scores across the two assessment occasions, alongside moderate stability in individual differences. Because longitudinal measurement invariance was established, the observed mean change is less likely to reflect a change in how the scale functioned across waves. The consistency ICC (.717) was also broadly comparable with the test-retest ICC of .70 previously reported for the 10-item RCADS-25 MDD scale

(Klaufus et al., 2020), suggesting that students tended to retain their relative positions over time.

The agreement and measurement-error results qualify this finding. Absolute agreement was slightly lower, while the Bland-Altman limits of agreement indicated considerable variability in within-person score differences (Bland & Altman, 1986). The SEM and MDC95 were also substantially larger than the average change of 1.14 points. Thus, the small mean-level decrease observed at the group level should not be equated with clear change at the individual level. This distinction is important because temporal stability, absolute agreement, and measurement error represent related but different aspects of repeated measurement (Mokkink et al., 2020; Parker et al., 2020).

Overall, the findings indicate that the RCADS-MDD retained its measurement properties and longitudinal measurement equivalence across the two assessment occasions, while showing moderate short-term stability in individual differences. At the same time, the relatively wide limits of agreement and the magnitude of the MDC95 suggest that small changes in individual scores should be interpreted cautiously. Because MDC95 reflects change beyond measurement error rather than clinically or practically meaningful change, further studies using external anchors are needed to establish meaningful change thresholds at the individual level.

Validity Evidence Based on Relations With Other Variables

The pattern of associations observed across both waves provides additional support for the validity of the RCADS-MDD scores. Moderate-to-strong positive correlations with the RCADS anxiety subscales are consistent with the close relationship between depressive and anxiety symptoms in adolescents, thereby supporting convergent validity evidence. At the same time, higher RCADS-MDD scores were associated with lower perceived social support from family and friends and with greater pressure from study, in directions consistent with theoretical expectations regarding psychosocial correlates of depressive symptoms. Importantly, this pattern of associations was broadly replicated across both

assessment waves, suggesting relative consistency in the external correlates of RCADS-MDD scores across differing academic contexts. Overall, these findings provide validity evidence based on relations with other variables, including convergent and nomological evidence, for the interpretation of RCADS-MDD scores among Vietnamese lower secondary school students.

Limitations and Directions for Future Research

Several limitations should be considered. The sample comprised 805 students from three lower secondary schools in Hanoi. Although this sample was adequate for the current analyses, the limited number of participating schools and the restriction to a single city limit the generalizability of the findings. The sample was not nationally representative and may not reflect the diversity of Vietnamese lower secondary school students across different regions and populations. Self-report data may have been influenced by differences in item interpretation and response tendencies. The absence of diagnostic interviews, an independent depression measure, and indicators of functional impairment prevented the evaluation of criterion validity, classification accuracy, sensitivity, and specificity.

The 14-day interval captured only short-term stability. The partial Threshold invariance models were not cross-validated, and floor effects in some items may have limited discrimination among students with low symptom levels. Future studies should include broader geographical and clinical samples, incorporate additional waves of assessment, and apply CFA methods appropriate for ordinal data. Further research should also use item response theory, differential item functioning analyses, and cognitive interviewing, while establishing the minimally important change before individual-level score changes are interpreted.

CONCLUSION

The present findings support the potential utility of the RCADS-MDD for assessing depressive symptoms among Vietnamese lower secondary school students. Across two assessment occasions, the scale showed a predominantly unidimensional structure, good internal

consistency, and acceptable-to-good model fit. Longitudinal configural, metric, and threshold invariance were supported, indicating sufficient comparability of RCADS-MDD scores across the two assessment occasions. Measurement invariance was also supported across gender, grade level, and school at both waves, providing further evidence of score comparability across these subgroups**.** Correlations with theoretically related psychological and contextual variables provided preliminary support for the expected patterns of association, although the use of other RCADS subscales limited the independence of the convergent validity evidence.

Several findings warrant caution in interpreting the scale's measurement properties. Some items showed comparatively weaker associations with the general depression factor, and average variance extracted remained below the conventional .50 criterion. Noticeable floor effects in several items may also limit discrimination among students with low symptom levels**.** Short-term rank-order stability was moderate rather than high. Although mean depressive symptom scores decreased significantly between waves, the relatively wide limits of agreement, together with the standard error of measurement and MDC95, indicate that small individual score changes may reflect measurement error rather than meaningful change. The MDC95 indicates the magnitude of change required to exceed expected measurement error at the 95% confidence level; it does not establish clinical significance.

Taken together, the findings suggest that the RCADS-MDD may be useful for assessing depressive symptom severity and conducting group-level research among Vietnamese lower secondary school students. However, the sample was drawn from only three schools in Hanoi and was not nationally representative. Caution is therefore needed when generalizing the findings to the wider population of Vietnamese students. The interpretation of individual score changes should also remain cautious until further research establishes responsiveness to change and externally anchored thresholds for clinically meaningful change.

DECLARATIONS

Funding Statement

This research received no external funding.

Conflict of Interest Statement

The authors declare that there are no conflicts of interest related to this study.

Consent to Participate

Informed consent was obtained from all participants and their legal guardians prior to participation.

Consent for Publication

Participants and their legal guardians provided consent for the publication of anonymized data.

Availability of Data and Materials

The data that support the findings of this study are not publicly available due to privacy and ethical restrictions but are available from the corresponding author upon reasonable request.

Code Availability

Not applicable.

REFERENCES

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. American Educational Research Association.

Bland, J. M., & Altman, D. G. (1986). Statistical methods for assessing agreement between two methods of clinical measurement. *The Lancet, 1*(8476), 307-310. https://doi.org/10.1016/S0140-6736(86)90837-8

Cervin, M., Veas, A., Piqueras, J. A., & Martínez-González, A. E. (2022). A multi-group confirmatory factor analysis of the Revised Children's Anxiety and Depression Scale (RCADS) in Spain, Chile and Sweden. *Journal of Affective Disorders, 310,* 228-234. https://doi.org/10.1016/j.jad.2022.05.031

Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal, 14*(3), 464-504. https://doi.org/10.1080/10705510701301834

Chorpita, B. F., Moffitt, C. E., & Gray, J. (2005). Psychometric properties of the Revised Child Anxiety and Depression Scale in a clinical sample. *Behaviour Research and Therapy, 43*(3), 309-322. https://doi.org/10.1016/j.brat.2004.02.004

Chorpita, B. F., Yim, L., Moffitt, C., Umemoto, L. A., & Francis, S. E. (2000). Assessment of symptoms of DSM-IV anxiety and depression in children: A Revised Child Anxiety and Depression Scale. *Behaviour Research and Therapy, 38*(8), 835-855. https://doi.org/10.1016/S0005-7967(99)00130-8

de Vet, H. C. W., Terwee, C. B., Knol, D. L., & Bouter, L. M. (2006). When to use agreement versus reliability measures. *Journal of Clinical Epidemiology, 59*(10), 1033-1039. https://doi.org/10.1016/j.jclinepi.2005.10.015

Donnelly, A., Fitzgerald, A., Shevlin, M., & Dooley, B. (2019). Investigating the psychometric properties of the Revised Child Anxiety and Depression Scale in a non-clinical sample of Irish adolescents. *Journal of Mental Health, 28*(4), 345-356. https://doi.org/10.1080/09638237.2018.1437604

Giannopoulou, I., Pasalari, E., Bali, P., Grammatikaki, D., & Ferentinos, P. (2022). Psychometric properties of the Revised Child Anxiety and Depression Scale in Greek adolescents. *Clinical Child Psychology and Psychiatry, 27*(2), 424-438. https://doi.org/10.1177/13591045211056502

Grothus, S., Sommer, A., Claus, B. B., Stahlschmidt, L., Chorpita, B. F., & Wager, J. (2023). The German version of the Revised Children's Anxiety and Depression Scale—Psychometric properties and normative data for German 8- to 17-year-olds. *International Journal of Methods in Psychiatric Research, 32*(4), e1965. https://doi.org/10.1002/mpr.1965

Grothus, S., Sommer, K., Claus, B. B., Stahlschmidt, L., Höfel, L., Chorpita, B. F., & Wager, J. (2025). Short forms of the German Revised Children's Anxiety and Depression Scale: Validation and normative data of the 11- and 25-item versions. *International Journal of Methods in Psychiatric Research, 34*(2), e70022. https://doi.org/10.1002/mpr.70022

Ho, T. T. Q., Nguyen, B. T. N., & Nguyen, N. P. H. (2022). Academic stress and depression among Vietnamese adolescents: A moderated mediation model of life satisfaction and resilience. *Current Psychology*. Advance online publication. https://doi.org/10.1007/s12144-022-03661-3

Klaufus, L., Verlinden, E., van der Wal, M., Kösters, M., Cuijpers, P., & Chinapaw, M. (2020). Psychometric evaluation of two short versions of the Revised Child Anxiety and Depression Scale. *BMC Psychiatry, 20*, Article 47. https://doi.org/10.1186/s12888-020-2444-5

Kline, R. B. (2023). *Principles and practice of structural equation modeling* (5th ed.). The Guilford Press.

Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine, 15*(2), 155-163. https://doi.org/10.1016/j.jcm.2016.02.012

Lisøy, C., Neumer, S.-P., Waaktaar, T., Ingul, J. M., Holen, S., & Martinsen, K. (2022). Making high-quality measures available in diverse contexts: The psychometric properties of the Revised Child Anxiety and Depression Scale in a Norwegian sample. *International Journal of Methods in Psychiatric Research, 31*(4), e1935. https://doi.org/10.1002/mpr.1935

Lu, B., Lin, L., & Su, X. (2024). Global burden of depression or depressive symptoms in children and adolescents: A systematic review and meta-analysis. *Journal of Affective Disorders, 354*, 553-562. https://doi.org/10.1016/j.jad.2024.03.074

Mokkink, L. B., Boers, M., van der Vleuten, C. P. M., Bouter, L. M., Alonso, J., Patrick, D. L., de Vet, H. C. W., & Terwee, C. B. (2020). COSMIN Risk of Bias tool to assess the quality of studies on reliability or measurement error of outcome measurement instruments: A Delphi study. *BMC Medical Research Methodology, 20*, Article 293. https://doi.org/10.1186/s12874-020-01179-5

Mokkink, L. B., Prinsen, C. A. C., Patrick, D. L., Alonso, J., Bouter, L. M., de Vet, H. C. W., & Terwee, C. B. (2018). *COSMIN methodology for systematic reviews of patient-reported outcome measures: User manual*. COSMIN. https://doi.org/10.1007/978-3-031-17299-1_2972

Mokkink, L. B., Terwee, C. B., Patrick, D. L., Alonso, J., Stratford, P. W., Knol, D. L., Bouter, L. M., & de Vet, H. C. W. (2010a). The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: An international Delphi study. *Quality of Life Research, 19*(4), 539-549. https://doi.org/10.1007/s11136-010-9606-8

Mokkink, L. B., Terwee, C. B., Patrick, D. L., Alonso, J., Stratford, P. W., Knol, D. L., Bouter, L. M., & de Vet, H. C. W. (2010b). The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. *Journal of Clinical Epidemiology, 63*(7), 737-745. https://doi.org/10.1016/j.jclinepi.2010.02.006

Parker, R. A., Scott, C., Inácio, V., & Stevens, N. T. (2020). Using multiple agreement methods for continuous repeated measures data: A tutorial for practitioners. *BMC Medical Research Methodology, 20*, Article 154. https://doi.org/10.1186/s12874-020-01022-x

Petrič, M., Kurnik Mesarič, K., Kodrič, J., & Janjušević, P. (2024). Psychometric validation of the Slovenian version of the Revised Child Anxiety and Depression Scale: Child and parent versions. *Slovenian Journal of Public Health, 63*(4), 164-171. https://doi.org/10.2478/sjph-2024-0022

Piqueras, J. A., Martín-Vivar, M., Sandín, B., San Luis, C., & Pineda, D. (2017). The Revised Child Anxiety and Depression Scale: A systematic review and reliability generalization meta-analysis. *Journal of Affective Disorders, 218*, 153-169. https://doi.org/10.1016/j.jad.2017.04.022

Prinsen, C. A. C., Mokkink, L. B., Bouter, L. M., Alonso, J., Patrick, D. L., de Vet, H. C. W., & Terwee, C. B. (2018). COSMIN guideline for systematic reviews of patient-reported outcome measures. *Quality of Life Research, 27*(5), 1147-1157. https://doi.org/10.1007/s11136-018-1798-3

Putnick, D. L., & Bornstein, M. H. (2016). Measurement invariance conventions and reporting: The state of the art and future directions for psychological research. *Developmental Review, 41*, 71-90. https://doi.org/10.1016/j.dr.2016.06.004

Radez, J., Waite, P., Chorpita, B. F., Creswell, C., Orchard, F., Percy, R., Spence, S. H., & Reardon, T. (2021). Using the 11-item version of the Revised Child Anxiety and Depression Scale to identify anxiety and depressive disorders in adolescents. *Research on Child and Adolescent Psychopathology, 49, 1241-1257.* https://doi.org/10.1007/s10802-021-00817-w

Schlechter, P., Hillmann, M., & Neufeld, S. A. S. (2024). Gender, age, and longitudinal measurement invariance of child and adolescent depression scales: A systematic review. *Clinical Psychology Review, 113*, Article 102481. https://doi.org/10.1016/j.cpr.2024.102481

Stevanovic, D., Bagheri, Z., Atilola, O., Vostanis, P., Stupar, D., Moreira, P., Franic, T., Davidovic, N., Knez, R., Nikšić, A., Dodig-Ćurković, K., Avicenna, M., Multazam Noor, I., Nussbaum, L., Deljkovic, A., Aziz Thabet, A., Petrov, P., Ubalde, D., Monteiro, L. A., & Ribas, R. (2017). Cross-cultural measurement invariance of the Revised Child Anxiety and Depression Scale across 11 world-wide societies. *Epidemiology and Psychiatric Sciences, 26*(4), 430-440. https://doi.org/10.1017/S204579601600038X

Tran, L. C. T., et al. (2025). Prevalence of depressive symptoms and their determinants among adolescents in Can Tho City, Vietnam: A cross-sectional study. *Global Mental Health*. Advance online publication. https://doi.org/10.1017/gmh.2025.10096