children

MDPI

Article

Cross-Cultural Adaptation and Psychometric Evaluation of the European Portuguese PHQ-9 Modified for Teens in a Large School-Based Adolescent Sample

Leonel Lusquinhos [1,2,3,*], Ana Cristina Maia Rocha [4,5], Lara Guedes de Pinho [6,7,8] and Ana Lúcia João [9]

1 School of Nursing, University of Minho, 4710-057 Braga, Portugal
2 Health Sciences Research Unit: Nursing (UICISA:E), University of Coimbra, 3004-531 Coimbra, Portugal
3 Nursing Research Centre, University of Minho, 4710-057 Braga, Portugal
4 National School of Public Health (NOVA), Public Health Research Center, Comprehensive Health Research Centre (CHRC), LA-REAL, CCAL, NOVA University Lisbon, 1600-560 Lisbon, Portugal; acm.rocha@ensp.unl.pt
5 School of Health Sciences (ESSUA), University of Aveiro, 3810-193 Aveiro, Portugal
6 Escola Superior de Saúde, Universidade Politécnica de Viana do Castelo, 4900-314 Viana do Castelo, Portugal; l.pinho@ess.ipvc.pt
7 Comprehensive Health Research Centre (CHRC), LA-REAL, Universidade de Évora, 7000-811 Évora, Portugal
8 Health Sciences Research Unit: Nursing (UICISA:E), Núcleo da Universidade Politécnica de Viana do, 4900-314 Viana do Castelo, Portugal
9 School of Health, Santarém Polytechnic Institute, 2005-075 Santarém, Portugal; ana.joao@essaude.ipsantarem.pt
* Correspondence: oliveira@ese.uminho.pt

Highlights

What are the main findings?

- The European Portuguese PHQ-9 Modified for Teens (PHQ-9_T) showed a unidimensional structure and high internal consistency in a large school-based adolescent sample.
- Scores showed theoretically coherent associations with anxiety symptoms and cognitive distortions, with evidence supporting measurement invariance across sex assigned at birth.

What are the implications of the main findings?

- The European Portuguese PHQ-9_T provides a brief measure for structured assessment of depressive symptom severity in adolescent research and health settings.
- PHQ-9_T scores should be interpreted within broader clinical and psychosocial assessment and not as a standalone diagnosis.

Abstract

Background/Objectives: Depressive symptoms are a major public health concern during adolescence, and brief, psychometrically sound, culturally adapted instruments are needed to support early identification in school settings. Although the PHQ-9 has previously been validated in European Portuguese adult samples, adolescent-specific evidence for the PHQ-9 Modified for Teens (PHQ-9_T) remains limited. This study aimed to cross-culturally adapt the PHQ-9_T for European Portuguese adolescents and evaluate its structural validity, internal consistency, construct validity, and measurement invariance across sex in a school-based sample. **Methods:** A cross-sectional psychometric study was conducted among 1089 students aged 14–19 years attending Portuguese public secondary schools during the 2024/2025 academic year. Following a five-stage cross-cultural adaptation process, the sample was randomly divided into exploratory (n = 554) and confirmatory

Check for updates

Academic Editor: Laura Elvira Prino

Received: 16 September 2026
Revised: 18 October 2026
Accepted: 21 October 2026
Published: 22 October 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

(n = 535) subsamples. Exploratory factor analysis used polychoric correlations and unweighted least squares extraction, whereas confirmatory factor analysis used WLSMV. Internal consistency, hypothesis testing for construct validity using the GAD-7 and CD-Quest-T, and measurement invariance across sex assigned at birth were also examined. **Results:** Exploratory factor analysis supported a unidimensional structure explaining 59.30% of the variance (factor loadings 0.717–0.824). Confirmatory factor analysis provided overall support for the one-factor model (CFI = 0.985; TLI = 0.981; RMSEA = 0.082, 90% CI [0.074, 0.090]; SRMR = 0.036). Internal consistency was high (α = 0.911; ω = 0.910). PHQ-9_T scores showed a strong positive correlation with GAD-7 (ρ= 0.787) and a moderate positive correlation with CD-Quest-T (ρ = 0.535) scores. Configural, metric, and scalar measurement invariance across sex was supported. Among participants with complete PHQ-9_T scores (n = 1070), 41.3% scored within the moderate-or-higher symptom-severity range. **Conclusions:** The European Portuguese PHQ-9_T demonstrated evidence supporting structural validity, internal consistency, construct validity, and measurement invariance across sex in this school-based adolescent sample, supporting its use as a brief measure of depressive symptom severity within structured, multiprofessional school and community mental health pathways, rather than as a standalone diagnostic tool.

Keywords: adolescents; depressive symptoms; PHQ-9 Modified for Teens; psychometric properties; structural validity; measurement invariance; school mental health; Portugal

1. Introduction

Adolescence is a developmental period characterized by substantial physical, cognitive, emotional, and social changes. Exposure to adverse experiences may increase vulnerability to mental health problems, many of which begin before adulthood and can affect well-being, educational attainment, social functioning, and health across the life course [1]. Within this developmental context, assessing depressive symptoms is important because their implications may extend beyond current emotional distress. Depression in young people is associated with recurrence and difficulties in interpersonal, educational, and occupational functioning. Assessment should therefore consider both current symptom severity and its consequences for everyday functioning and developmental transitions [2].

Depressive symptoms in adolescence are associated with psychological distress, impaired school performance, absenteeism, social withdrawal, substance use, self-harm, and suicidal ideation [3]. Schools provide an important setting for early identification because they facilitate regular contact with adolescents and coordination among students, families, educators, nurses, and other health professionals [4].

Identifying depression during adolescence can be challenging because irritability may be prominent and anxiety commonly co-occurs with depressive symptoms. Assessment must also distinguish transient low mood from persistent symptoms associated with functional impairment. Considering symptom duration, effects on daily life, and information about the adolescent's family and school context is therefore important when interpreting self-reported symptoms [2].

Recent evidence reinforces the relevance of depressive symptoms as a major public health concern during adolescence and supports the use of brief screening instruments for early identification. In a representative sample of 2235 Spanish adolescents, Fonseca-Pedrero et al. found that the PHQ-9 showed adequate psychometric properties, including support for a unidimensional structure and meaningful associations with related constructs [5].

Despite the need for early identification, many adolescents do not receive timely support because of stigma, limited-service access, insufficient resources, and the normalization of emotional distress as part of adolescent development. Brief, psychometrically sound, and culturally adapted instruments may support structured symptom assessment and referral [6,7].

Validated self-report measures provide a standardized way to quantify adolescents' depressive symptoms and examine their relationships with other psychological constructs in research. In applied settings, they can contribute symptom information to a broader assessment but should not be treated as independent diagnostic instruments [2,5]. Evidence concerning internal structure and consistency should also be distinguished from evidence concerning diagnostic accuracy. A recent systematic review found variation in the sensitivity and specificity of depression screening instruments across studies, reinforcing the importance of evaluating diagnostic performance rather than assuming it from other measurement properties [8].

Portuguese school-based research provides a relevant psychosocial context for assessing adolescent depressive symptoms. An analysis of Portuguese HBSC 2018 data involving 8215 students found that higher perceived quality of life was associated with lower schoolwork pressure, greater perceived school safety, and more positive relationships with peers and teachers. Girls reported greater schoolwork pressure and lower quality of life than boys, illustrating variation in students' experiences within the same educational context [9]. These findings highlight the importance of considering the school environment and interpersonal relationships alongside individual symptoms.

Complementary evidence comes from the Portuguese Psychological Health and Well-being School Observatory. In a study of 3109 students attending grades 7–12, higher levels of competence, confidence, and connection were associated with lower depressive, anxiety, and stress symptoms and greater life satisfaction [10]. Together, these studies situate adolescent mental health assessment in Portugal within a context of school experiences and personal and relational resources. This context supports the relevance of evaluating a culturally adapted measure of depressive symptom severity among public secondary-school students, while recognizing that findings from the present convenience sample cannot be generalized to all adolescents living in Portugal.

The PHQ-9 is widely used to quantify depressive symptom severity, but its internal structure has not been completely uniform across populations. A recent systematic review and meta-analysis identified support for both unidimensional and multidimensional representations, with the one-factor nine-item model showing the strongest overall support for interpretation of an aggregated total score, while a cognitive/affective–somatic two-factor solution also demonstrated substantial psychometric evidence [11]. Population- and developmental-stage-specific examination of structural validity therefore remains important.

Although the European Portuguese PHQ-9 has previously been psychometrically evaluated in adult community samples [12], evidence from adult populations cannot be assumed to generalize to adolescents or to the adolescent-specific PHQ-9 Modified for Teens. Adolescent-specific evidence for the European Portuguese PHQ-9_T, particularly in school-based populations, therefore, remains limited.

This study aimed to cross-culturally adapt the PHQ-9 Modified for Teens for European Portuguese adolescents and evaluate its measurement properties in a school-based sample. Specifically, we examined structural validity, internal consistency, hypothesis testing for construct validity, and measurement invariance across sex assigned at birth. As a secondary descriptive objective, we characterized the distribution of PHQ-9_T symptom-severity scores.

2. Materials and Methods

2.1. Design and Participants

A cross-sectional psychometric study was conducted using a non-probabilistic convenience sample of students attending participating Portuguese public secondary schools during the 2024/2025 academic year. The study was designed to permit independent exploratory and confirmatory evaluation of the internal structure through random division of the total sample into EFA and CFA subsamples. Eligible participants were students aged 14–19 years enrolled in participating public secondary schools and able to complete the questionnaires in written Portuguese. Students were excluded when evident cognitive difficulties or substantial comprehension barriers prevented them from understanding and completing the questionnaire independently.

Participation was voluntary, and informed consent and assent procedures followed the ethics-approved protocol.

2.2. Measures

A sociodemographic questionnaire assessed age, nationality, sex assigned at birth, gender identity, school grade, academic performance, parental education, school social support, and participation in extracurricular activities.

The source instrument was the PHQ-9 Modified for Teens (PHQ-9_T), an adolescent-specific adaptation included in the Guidelines for Adolescent Depression in Primary Care (GLAD-PC) materials. The form contains nine core depressive-symptom items referring to the previous two weeks, rated from 0 ('not at all') to 3 ('nearly every day'), yielding a total symptom-severity score ranging from 0 to 27. The source form also contains a functional-impairment item and supplementary questions concerning persistent depressed mood and suicidal thoughts and behavior. In accordance with the scoring instructions, only the nine core symptom items contribute to the 0–27 total score and were included in the factor-analytic, internal-consistency, and construct-validity analyses reported in this study. Higher PHQ-9_T scores indicate greater depressive symptom severity [6].

The CD-Quest-T is an eight-item self-report measure of cognitive distortions, with items scored from 0 to 5. The European Portuguese version demonstrated a unidimensional structure, good internal consistency, and convergent validity with anxiety and depressive symptoms in adolescents [13].

The European Portuguese version of the GAD-7 was administered to assess anxiety symptoms during the previous two weeks. The instrument comprises seven items scored from 0 to 3, yielding a total score ranging from 0 to 21. Higher scores indicate greater anxiety symptom severity [14].

2.3. Translation and Cross-Cultural Adaptation

The translation and cross-cultural adaptation of the PHQ-9 Modified for Teens (PHQ-9_T) into European Portuguese followed the cross-cultural adaptation framework proposed by Beaton et al. and Gagnier et al. [15,16] with the aim of preserving semantic, idiomatic, experiential, and conceptual equivalence between the source and target versions. The source instrument was the PHQ-9 Modified for Teens included in the Guidelines for Adolescent Depression in Primary Care (GLAD-PC), Version 3, published by The REACH Institute.

The adaptation process comprised five sequential stages. First, two bilingual translators independently translated the source instrument from English into European Portuguese. Second, the two forward translations were compared and synthesized into a single consensual version. Third, this synthesized version was independently back-translated into English by two translators who had not been involved in the initial translation and had no previous contact with the source version. Fourth, the translated and back-translated

versions were reviewed by an expert panel with expertise in methodology, mental health, and cross-cultural adaptation. The panel examined the semantic, idiomatic, experiential, and conceptual equivalence of the translated content and resolved discrepancies through consensus.

Finally, the pre-final European Portuguese version underwent cognitive pretesting with 30 secondary-school adolescents. This stage assessed the clarity, comprehensibility, and linguistic appropriateness of the translated content for the target population. Wording identified as potentially ambiguous or difficult to understand was reviewed and refined following discussion of the adolescents' feedback. The final European Portuguese version was established after completion of this process, with the aim of retaining the conceptual meaning of the source PHQ-9_T while ensuring linguistic and cultural appropriateness for Portuguese adolescents.

2.4. Procedure

Data were collected anonymously during scheduled classroom sessions using an online questionnaire accessed via participants' smartphones or school computers. Before completion, participants received written information about the study objectives, voluntary participation, confidentiality, the right to withdraw, and available school, community, and emergency mental health services. Questionnaire completion took approximately 15–20 min and included the supplementary PHQ-9_T questions, which do not contribute to the 0–27 symptom-severity score and were excluded from the psychometric analyses reported here. Because responses could not be linked to individual participants, adolescents endorsing thoughts of death or self-harm could not be identified or contacted through the research database. The provision of information about support services therefore did not constitute individualized follow-up of questionnaire responses.

2.5. Statistical Analysis

Analyses were conducted using IBM SPSS Statistics version 30 and JASP. Descriptive statistics were calculated for participant characteristics and item distributions. The total sample was randomly divided into an exploratory subsample (n = 554) and a confirmatory subsample (n = 535).

EFA was conducted using a polychoric correlation matrix and unweighted least squares extraction. Factor retention was determined through parallel analysis based on 1000 random matrices, inspection of the scree plot, factor loadings, and theoretical interpretability. Sampling adequacy was evaluated using the KMO index and Bartlett's test of sphericity.

Confirmatory factor analysis (CFA) was performed in the confirmatory subsample to evaluate the one-factor structure identified through EFA. Given the ordinal nature of the item responses, the model was estimated using the weighted least squares mean and variance adjusted (WLSMV) estimator. Model fit was assessed using the chi-square statistic and degrees of freedom, the comparative fit index (CFI), the Tucker–Lewis index (TLI), the root mean square error of approximation (RMSEA) with its 90% confidence interval, and the standardized root mean square residual (SRMR). Model fit was evaluated by considering the overall pattern of indices, standardized factor loadings, theoretical coherence, and parsimony rather than relying on a single universal cut-off. Given previous evidence supporting both unidimensional and two-factor representations of the PHQ-9, the findings were interpreted as evidence concerning the plausibility of the unidimensional structure rather than as demonstrating its superiority over alternative factorial models.

Evidence based on theoretically expected associations was examined using Spearman's rank-order correlations between PHQ-9_T scores and scores on the Generalized Anxiety

Disorder-7 (GAD-7) and the Cognitive Distortions Questionnaire for Adolescents (CD-Quest-T). We hypothesized positive associations between depressive symptoms and both anxiety symptoms and cognitive distortions. These analyses were intended to examine associations consistent with theoretical expectations and were not designed to establish discriminant validity or provide a comprehensive assessment of construct validity.

Measurement invariance across female and male participants, based on sex assigned at birth, was examined using multigroup CFA for ordinal indicators. Configural, metric, and scalar models were estimated sequentially. Scalar invariance was evaluated by imposing equality constraints on factor loadings and item thresholds across groups, with additional identification constraints determined by the categorical-data parameterization implemented in the software. Model comparisons were based primarily on changes in CFI and RMSEA, considered alongside the overall pattern of model fit.

Missing data were examined before the psychometric analyses. Of the 1089 participants, 1070 (98.3%) provided complete PHQ-9_T responses, whereas 19 (1.7%) had missing item responses. No data imputation was performed. EFA and CFA were conducted using participants with valid responses to the relevant PHQ-9_T items. For correlations involving the PHQ-9_T, GAD-7, and CD-Quest-T, pairwise deletion was used so that each correlation was estimated using all available observations for the corresponding pair of variables.

3. Results

3.1. Sample Characterization

The sample consisted of 1089 adolescents aged 14–19 years. Most participants were aged 15–17 years, with the largest proportion aged 16 years (34.4%), followed by 17 years (29.3%) and 15 years (23.8%). Most participants had Portuguese nationality (90.4%), and Brazilian nationality was the second most frequent nationality (6.7%). Most participants were assigned female at birth (56.9%), followed by male (41.1%); 0.5% were intersex and 1.5% preferred not to answer. Most participants identified as cisgender (93.4%), while smaller proportions identified as trans (1.2%) or non-binary (0.7%), and 4.7% preferred not to answer. Participants were enrolled in the 10th (45.4%), 11th (32.5%), and 12th grades (22.0%).

Regarding socioeconomic indicators, 64.5% did not report receiving school social support, whereas 7.0%, 14.4%, and 14.1% were classified in categories A, B, and C, respectively. Maternal education was mainly secondary (31.8%) or higher education (32.0%), and paternal education was mainly secondary (29.5%), higher education (21.8%), or third cycle/basic education (21.1%). Most participants reported participating in extracurricular activities (62.0%). After characterizing the total sample, it was randomly divided into two subsamples of approximately equivalent size, with the objective of validating the factorial structure of the PHQ-9_T scale.

3.2. PHQ-9_T Item Distributions

Item means ranged from 0.53 for thoughts of death or self-harm to 1.50 for fatigue/low energy. The highest item means were observed for fatigue/low energy, concentration difficulties, and sleep problems. Although the thoughts of death or self-harm item showed the lowest mean, endorsement of this item remains clinically relevant and requires careful interpretation in applied settings (Table 1).

Table 1. Descriptive Statistics of the PHQ-9_T Items.

	PHQ-9_T Items	Minimum	Maximum	Mean	SD
1	"I had little interest or pleasure in doing things?"	0	3	1.18	0.99
2	"I felt down, depressed, angry, or hopeless?"	0	3	1.13	1.04
3	"I had trouble falling asleep, staying asleep, or oversleeping?"	0	3	1.20	1.13
4	"I felt tired or low on energy?"	0	3	1.50	1.06
5	"I had no appetite, weight loss, or overeating?"	0	3	0.91	1.09
6	"I felt bad about myself, or that I am a failure, or that I have disappointed myself or my family?"	0	3	1.05	1.10
7	"I had difficulty concentrating on schoolwork, reading, or watching television?"	0	3	1.26	1.07
8	"Moving or speaking so slowly that other people can notice? Or the opposite: being so restless and agitated that I walk back and forth more than usual?"	0	3	0.75	0.95
9	"I felt that I would be better off dead or willing to hurt myself in some way?"	0	3	0.53	0.92

Participants were classified according to the following symptom-severity ranges: minimal depressive symptoms, 0–4; mild depressive symptoms, 5–9; moderate depressive symptoms, 10–14; moderately severe depressive symptoms, 15–19; and severe depressive symptoms, 20–27.

Among the 1089 participants, 1070 (98.3%) had complete PHQ-9_T total scores, whereas 19 (1.7%) had missing data. Among participants with complete scores, 29.1% were classified in the minimal symptom range, 29.6% in the mild range, 17.0% in the moderate range, 13.0% in the moderately severe range, and 11.3% in the severe range (Table 2). Overall, 58.7% of participants with complete PHQ-9_T scores fell within the minimal or mild symptom-severity ranges, whereas 41.3% scored within the moderate-or-higher ranges. Because the sample was non-probabilistic and no structured diagnostic interview was administered, these findings should be interpreted as the distribution of self-reported depressive symptom severity within the study sample rather than as an estimate of the prevalence of major depressive disorder among Portuguese adolescents.

Table 2. Distribution of PHQ-9_T Symptom-Severity Categories.

Symptom-Severity Category	n	%
Minimal depressive symptoms	311	29.1
Mild depressive symptoms	317	29.6
Moderate depressive symptoms	182	17.0
Moderately severe depressive symptoms	139	13.0
Severe depressive symptoms	121	11.3
Total	1070	100.0

3.3. Adequacy of the Sample to Factor Analysis

The EFA was conducted with the objective of exploring the latent structure of the scale items.

The first stage of the analysis involved verifying the adequacy of the sample for a factor analysis. The Kaiser–Meyer–Olkin index was 0.931, indicating excellent sampling adequacy for factor analysis. In addition, Bartlett's sphericity test showed a highly significant chi-square value ($\chi^2 = 2664.568$; df = 36; $p < 0.001$), confirming that the correlations between the items are sufficiently large to justify an exploratory factor analysis.

The exploratory factor analysis supported a unidimensional structure for the PHQ-9_T. The first factor showed an eigenvalue of 5.34 and explained 59.30% of the total variance,

supporting the adequacy of a 1-factor solution in the present sample. Factor loadings ranged from 0.717 to 0.824, and communalities ranged from 0.514 to 0.678, indicating that all items contributed meaningfully to the latent construct. A summary of these results is presented in Table 3.

Table 3. Eigenvalues and explained variance for the PHQ-9_T exploratory factor analysis.

	Eigenvalue	% of Variance	% Cumulative Variance
1	5.34	59.30	59.30
2	0.77	8.51	67.82
3	0.56	6.23	74.05
4	0.47	5.17	79.22
5	0.44	4.85	84.07
6	0.42	4.62	88.69
7	0.40	4.49	93.18
8	0.35	3.93	97.11
9	0.26	2.89	100.00

3.4. Correlations Between PHQ-9_T Items

Spearman's correlations were calculated to examine the relationships among PHQ-9_T items and between individual items and the total score. All item–total correlations were positive and statistically significant ($p < 0.001$), ranging from 0.613 to 0.816. The strongest correlations were observed for item 2, assessing depressed or irritable mood and hopelessness (ρ = 0.816), item 6, assessing negative self-evaluation or perceived failure (ρ = 0.785), and item 4, assessing tiredness or low energy (ρ = 0.780). Item 9, assessing thoughts of death or self-harm, showed the weakest item–total correlation (ρ = 0.613). Inter-item correlations were also positive and statistically significant ($p < 0.001$), ranging from 0.362 to 0.677, with the strongest association between items 2 and 6 (ρ = 0.677) and the weakest between item 3, assessing sleep difficulties or excessive sleep, and item 9 (ρ = 0.362). Overall, these findings indicate positive associations among the PHQ-9_T items and between individual items and the total score (Table 4).

Table 4. Correlations between the Items and Total PHQ-9_T.

PHQ-9_T Scale		1	2	3	4	5	6	7	8	9
PHQ-9_T Total	ρ	0.753	0.816	0.738	0.780	0.747	0.785	0.766	0.689	0.613
	p	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
1	ρ	1.000	0.612	0.503	0.562	0.496	0.537	0.552	0.469	0.436
	p	.	<0.001	<0.001	<0.001	<0.001	<.0.001	<0.001	<0.001	<0.001
2	ρ	0.612	1.000	0.522	0.590	0.555	0.677	0.541	0.527	0.510
	p	<0.001	.	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
3	ρ	0.503	0.522	1.000	0.548	0.539	0.470	0.522	0.464	0.362
	p	<0.001	<0.001	.	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
4	ρ	0.562	0.590	0.548	1.000	0.516	0.535	0.566	0.441	0.364
	p	<0.001	<0.001	<0.001	.	<0.001	<0.001	<0.001	<0.001	<0.001
5	ρ	0.496	0.555	0.539	0.516	1.000	0.537	0.513	0.505	0.478
	p	<0.001	<0.001	<0.001	<0.001	.	<0.001	<0.001	<0.001	<0.001
6	ρ	0.537	0.677	0.470	0.535	0.537	1.000	0.552	0.526	0.552
	p	<0.001	<0.001	<0.001	<0.001	<0.001	.	<0.001	<0.001	<0.001
7	ρ	0.552	0.541	0.522	0.566	0.513	0.552	1.000	0.509	0.380
	p	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	.	<0.001	<0.001
8	ρ	0.469	0.527	0.464	0.441	0.505	0.526	0.509	1.000	0.554
	p	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	.	<0.001

Table 4. *Cont.*

PHQ-9_T Scale		1	2	3	4	5	6	7	8	9
9	ρ	0.436	0.510	0.362	0.364	0.478	0.552	0.380	0.554	1.000
	p	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	.

Legend: 1: "I had little interest or pleasure in doing things?"; 2: "I felt down, depressed, angry, or hopeless?"; 3: "I had trouble falling asleep, staying asleep, or oversleeping?"; 4: "I felt tired or low on energy?"; 5: "I had no appetite, weight loss, or overeating?"; 6: "I felt bad about myself, or that I am a failure, or that I have disappointed myself or my family?"; 7: "I had difficulty concentrating on schoolwork, reading, or watching television?"; 8: "Moving or speaking so slowly that other people can notice? Or the opposite: being so restless and agitated that I walk back and forth more than usual?"; 9: "I felt that I would be better off dead or willing to hurt myself in some way?".

3.5. Hypothesis Testing for Construct Validity

Hypothesis testing for construct validity was examined through associations with anxiety symptoms and cognitive distortions. PHQ-9_T total scores showed a strong positive correlation with GAD-7 scores ($\rho = 0.787$, $p < 0.001$) and a moderate positive correlation with CD-Quest-T scores ($\rho = 0.535$, $p < 0.001$). A moderate positive association was also observed between GAD-7 and CD-Quest-T scores ($\rho = 0.577$, $p < 0.001$) (Table 5). At the item level, correlations with GAD-7 ranged from 0.441 to 0.731, whereas correlations with CD-Quest-T ranged from 0.244 to 0.498; detailed item-level results are provided in Supplementary Table S1.

Table 5. Spearman correlations between PHQ-9_T, GAD-7, and CD-Quest-T total scores.

Measure	PHQ-9_T	GAD-7	CD-Quest-T
PHQ-9_T	1.000	0.787 ***	0.535 ***
GAD-7	0.787 ***	1.000	0.577 ***
CD-Quest-T	0.535 ***	0.577 ***	1.000

Note: *** $p < 0.001$.

The strong correlation with the GAD-7, while supporting convergent validity, should also be interpreted in the context of the conceptual and clinical overlap between anxiety and the construct assessed by the scale, particularly among adolescents.

3.6. Confirmatory Factor Analysis

Confirmatory factor analysis (CFA) was conducted using the weighted least squares mean and variance adjusted (WLSMV) estimator, which is appropriate for ordinal indicators and Likert-type items. Model fit was evaluated using the chi-square statistic, the comparative fit index (CFI), the Tucker–Lewis index (TLI), the root mean square error of approximation (RMSEA), and the standardized root mean square residual (SRMR).

The one-factor model yielded favorable incremental and residual fit indices (CFI = 0.985, TLI = 0.981, and SRMR = 0.036). However, the RMSEA was 0.082 (90% CI [0.074, 0.090]), indicating a less favorable degree of approximate fit. The robust chi-square test was statistically significant, $\chi^2(27) = 350.673$, $p < 0.001$. Overall, these findings support the plausibility of a unidimensional representation of the PHQ-9_T, although model fit was not uniformly satisfactory across all indices. Alternative factorial structures therefore remain plausible and warrant further investigation (Table 6).

Table 6. Results of the confirmatory factor analysis of PHQ-9_T.

Fit Index	Value
χ^2 (27)	350.673, $p < 0.001$
CFI	0.985
TLI	0.981
RMSEA	0.082
90% CI for RMSEA	[0.074, 0.090]
SRMR	0.036

3.7. Measurement Invariance Across Sex Assigned at Birth

Measurement invariance was examined across participants assigned female or male at birth. The configural model showed adequate fit (CFI = 0.986, TLI = 0.978, RMSEA = 0.083, and SRMR = 0.035). Constraining factor loadings across groups did not worsen model fit, as indicated by changes in CFI (ΔCFI = +0.003) and RMSEA (ΔRMSEA = −0.016), supporting metric invariance (Table 7).

Table 7. Measurement invariance of the PHQ-9_T across sex assigned at birth: configural, metric, and scalar.

Model	χ^2	df	CFI	TLI	RMSEA	SRMR	ΔCFI	ΔRMSEA
Configural	322.84	46	0.986	0.978	0.083	0.035	—	—
Metric	264.69	54	0.989	0.986	0.067	0.039	+0.003	−0.016
Scalar	261.73	53	0.990	0.986	0.067	0.039	+0.001	0.000

Note: Models were estimated for ordered categorical indicators using WLSMV. Degrees of freedom reflect the identification and equality constraints imposed under the categorical-data parameterization.

The scalar model, which constrained factor loadings and item thresholds across groups in accordance with the identification requirements for ordinal indicators, also showed adequate fit (CFI = 0.990, TLI = 0.986, RMSEA = 0.067, and SRMR = 0.039). Changes in fit relative to the metric model were negligible (ΔCFI = +0.001; ΔRMSEA = 0.000), providing further evidence in support of scalar invariance. These findings suggest that the PHQ-9_T operates comparably across the two groups examined, supporting cautious comparisons of scores between them (Table 7).

3.8. Internal Consistency of the PHQ-9_T Scale

Internal consistency was high, with Cronbach's $\alpha = 0.911$ and McDonald's $\omega = 0.910$. Corrected item–total correlations ranged from 0.621 to 0.765, indicating that all nine items contributed meaningfully to the total score. Inter-item correlations were positive and ranged from 0.362 to 0.677, suggesting related but non-redundant item content.

4. Discussion

This study evaluated the psychometric properties of the European Portuguese PHQ-9 Modified for Teens (PHQ-9_T) in a large school-based adolescent sample. The findings provide evidence supporting its internal consistency, the plausibility of a unidimensional structure, theoretically expected associations with related psychological constructs, and measurement invariance across sex assigned at birth. Nevertheless, the structural validity findings require cautious interpretation, and further research is needed to strengthen the evidence supporting the interpretation and use of PHQ-9_T scores.

4.1. Structural Validity and Factorial Structure

Exploratory factor analysis (EFA) supported a unidimensional structure, with a single factor explaining 59.30% of the total variance and factor loadings ranging from 0.717 to 0.824. CFA provided further support for this representation, with favorable CFI, TLI, and SRMR values. However, the RMSEA indicated a less satisfactory degree of approximate fit. Thus, although the overall pattern of results supports the plausibility of a one-factor model, fit was not uniformly satisfactory across all indices.

These findings are broadly consistent with those reported by Fonseca-Pedrero et al. [5] in a representative sample of Spanish adolescents, where factor loadings ranged from 0.62 to 0.82 and the CFI was at least 0.97, and by Rakshasa-Loots et al. [17] in a cross-cultural study of African adolescents, where factor loadings ranged from 0.60 to 0.80. This convergence provides additional support for interpreting the PHQ-9_T total score as an indicator of overall depressive symptom severity in adolescent populations.

However, the findings should be considered in the context of the broader psychometric literature, which has identified both unidimensional and multidimensional representations of the PHQ-9. A systematic review and meta-analysis of 98 psychometric studies across 40 countries found that the nine-item unidimensional model had the strongest overall support for structural validity, internal consistency, and measurement invariance. Nevertheless, the review also identified substantial evidence supporting a two-factor cognitive/affective–somatic model, comprising a somatic factor defined by items 3–5 and a cognitive/affective factor defined by items 1, 2, and 6–9 [11]. Studies involving adolescents have likewise reported adequate fit for both one- and two-factor solutions [16,18].

Research involving Portuguese adults has also supported a two-factor representation of the European Portuguese PHQ-9 [12]. Differences between these findings and those of the present study may reflect variation in developmental stage, instrument version, and sample characteristics rather than a direct contradiction between the proposed models. Importantly, the present CFA evaluated only the unidimensional model identified during the exploratory analysis. Consequently, the study does not establish that this model is superior to alternative factorial representations. Future studies should directly compare theoretically supported one- and two-factor models in independent samples of Portuguese adolescents.

4.2. Internal Consistency and Item-Level Findings

The PHQ-9_T demonstrated high internal consistency (Cronbach's $\alpha = 0.911$; McDonald's $\omega = 0.910$), and corrected item–total correlations indicated that all nine items contributed meaningfully to the total score. The positive inter-item correlations further suggest that the items assess related aspects of depressive symptomatology while retaining some distinctiveness.

Items assessing depressed or irritable mood, negative self-evaluation, and low energy showed particularly strong associations with the total score or with other items, consistent with the central features of depressive symptomatology in adolescence [17]. In contrast, item 9, which assesses thoughts of death or self-harm, showed the lowest endorsement and the weakest item–total association. This pattern may reflect the comparatively lower frequency and distinct clinical significance of these experiences rather than inadequate psychometric performance.

Nevertheless, item 9 should not be interpreted as a standalone suicide-risk assessment measure. In clinical and school health settings, endorsement should prompt individualized clinical and safety assessment, with follow-up according to established procedures and, where indicated, the use of validated suicide-risk assessment tools [19]. In the present study, anonymous data collection prevented responses from being linked to individual

participants for targeted follow-up, highlighting the importance of establishing appropriate safety and referral procedures when implementing similar assessments.

4.3. Evidence Based on Associations with Related Constructs

The PHQ-9_T showed a strong positive correlation with GAD-7 scores ($\rho = 0.787$, $p < 0.001$) and a moderate positive correlation with CD-Quest-T scores ($\rho = 0.535$, $p < 0.001$). Both associations were in the theoretically expected direction and provide evidence relevant to the construct validity of the instrument.

The strong association with anxiety symptoms is consistent with the conceptual and clinical overlap between depression and anxiety during adolescence. However, the magnitude of this correlation indicates substantial shared variance and should not be interpreted as evidence of discriminant validity or as demonstrating that the PHQ-9_T distinguishes depressive symptoms from anxiety [5]. Rather, this finding supports the expected relationship between two closely related dimensions of psychological distress.

The moderate association with CD-Quest-T scores provides complementary evidence of a relationship between depressive symptom severity and cognitive distortions, consistent with the associations reported in the European Portuguese validation of the CD-Quest-T [13]. Taken together, these findings support the theoretically expected relationships between PHQ-9_T scores and related psychological constructs. However, they do not establish the distinctiveness of the instrument from other measures of psychological distress. Further research should examine a broader range of theoretically relevant constructs and employ complementary methods to strengthen the evidence supporting score interpretation.

4.4. Implications of Measurement Invariance Across Sex Assigned at Birth

Measurement invariance was supported at the configural, metric, and scalar levels across participants assigned female or male at birth. These findings suggest that the PHQ-9_T assesses depressive symptoms comparably across the two groups examined and provide a basis for cautious comparisons of scores between them. However, these results should not be generalized to sex or gender groups that were not included in the invariance analysis. Further studies should examine measurement invariance across additional sociodemographic and developmental characteristics, particularly age and socioeconomic background, before assuming that scores are directly comparable across these populations.

4.5. Implications for Research and Practice

The European Portuguese PHQ-9_T may support the structured assessment of depressive symptom severity in adolescent research, school health, primary care, and community settings. Its brevity and favorable internal consistency may facilitate the systematic identification of adolescents who require further assessment.

Nevertheless, the present study did not establish diagnostic accuracy or screening performance. No structured diagnostic interview was administered, and sensitivity, specificity, predictive values, and adolescent-specific cut-off scores were not evaluated. The PHQ-9_T should therefore not be used as a standalone diagnostic tool, nor should symptom-severity categories be interpreted as equivalent to a clinical diagnosis. Instead, scores should be interpreted alongside functional impairment, co-occurring symptoms, family and school context, and individual safety concerns. Their use should be embedded within structured pathways for further assessment, appropriate referral, and follow-up. School and community nurses may contribute to this process in collaboration with multidisciplinary adolescent-health teams [6,20].

The distribution of symptom-severity scores in the present sample also warrants careful interpretation. Among participants with complete PHQ-9_T scores, 41.3% scored within

the moderate-or-higher symptom-severity ranges. This finding describes the self-reported symptom burden among participating students; it does not estimate the prevalence of major depressive disorder among Portuguese adolescents. The non-probabilistic sampling design, reliance on self-report, and absence of diagnostic interviews preclude such an inference. Nonetheless, the findings highlight the potential value of accessible assessment and referral pathways in school settings.

4.6. Limitations and Future Research

Several limitations should be considered when interpreting these findings. First, the inclusion of a small number of participants aged 18–19 years extended the age range slightly beyond that specified for the source PHQ-9 Modified for Teens. Second, the school-based convenience sample limits representativeness and generalizability to adolescents who are not attending school or who are receiving clinical care. Third, self-reported data collected in classroom settings may have been affected by response and privacy-related biases.

Fourth, the absence of a structured diagnostic interview precluded the assessment of criterion validity, diagnostic accuracy, and population-specific cut-off scores. Fifth, the cross-sectional design did not permit evaluation of test–retest reliability, measurement error over time, predictive validity, or responsiveness to change. In addition, the CFA did not compare the unidimensional model with a theoretically supported two-factor solution. Evidence based on associations with related constructs was also limited to anxiety symptoms and cognitive distortions; discriminant validity and a broader range of theoretically relevant constructs were not assessed. Finally, anonymous data collection prevented individualized follow-up of participants who endorsed thoughts of death or self-harm.

Future research should address these limitations by directly comparing theoretically supported one- and two-factor models in independent samples of Portuguese adolescents. Studies incorporating structured diagnostic interviews should evaluate criterion validity and diagnostic accuracy, including sensitivity, specificity, predictive values, and candidate decision thresholds. Longitudinal designs should investigate test–retest reliability, measurement error, and responsiveness to change.

Further validation in clinical and more diverse community samples is also needed to establish whether the findings generalize beyond participating secondary-school students. Measurement invariance across age and socioeconomic groups should be examined before assuming score comparability across these populations. Finally, future data-collection protocols should establish clear procedures for responding to disclosures of self-harm or suicidal thoughts, particularly when anonymity limits the possibility of individual follow-up.

5. Conclusions

The European Portuguese PHQ-9 Modified for Teens demonstrated evidence supporting structural validity, internal consistency, hypothesis testing for construct validity, and measurement invariance across sex assigned at birth in this school-based sample. These findings support its use as a brief measure of depressive symptom severity in adolescent research and structured health assessment. However, the present study did not establish diagnostic accuracy or Portuguese adolescent-specific decision thresholds, and PHQ-9_T scores should not be interpreted as a standalone diagnosis. Further validation is needed to establish diagnostic accuracy, longitudinal measurement properties, and applicability across clinical and more diverse adolescent populations.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/children13111448/s1, Table S1: Item-level Spearman correlations between PHQ-9_T items and PHQ-9_T, GAD-7, and CD-Quest-T total scores.

Author Contributions: Conceptualization, L.L., A.C.M.R. and A.L.J.; methodology, L.L., A.L.J. and L.G.d.P.; software, A.L.J.; validation, L.L., A.L.J. and L.G.d.P.; formal analysis, A.L.J.; investigation, L.L. and A.L.J.; resources, L.L., A.C.M.R. and A.L.J.; data curation, A.L.J.; writing—original draft preparation, L.L.; writing—review and editing, L.L., A.L.J. and L.G.d.P.; visualization, L.L. and A.L.J.; supervision, L.G.d.P.; project administration, L.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki and approved by the Ethics Committee of the University of Évora (Opinion No. 24064).

Informed Consent Statement: Written informed consent was obtained from adult participants. For participants under 18 years of age, written informed consent was obtained from their legal guardians and assent was obtained from the adolescents prior to participation.

Data Availability Statement: The data that support the findings of this study are available from the corresponding author upon reasonable request. The data are not publicly available due to ethical and privacy considerations related to the participation of adolescents and the sensitive nature of the mental health information collected.

Acknowledgments: The authors would like to thank the participating schools, adolescents, and educational professionals who contributed to the implementation of this study.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. World Health Organization Mental Health of Adolescents. Available online: https://www.who.int/news-room/fact-sheets/detail/adolescent-mental-health (accessed on 9 July 2025).
2. Thapar, A.; Eyre, O.; Patel, V.; Brent, D. Depression in Young People. *Lancet* **2022**, *400*, 617–631. [CrossRef] [PubMed]
3. Looijmans, M.; von Spreckelsen, P.; Berkelmans, G.; Popma, A.; van Bergen, D.; Gilissen, R.; Mérelle, S. The Prevalence, Course, and Risk Factors of Suicidal Ideation and Suicide Attempts among Students in Vocational Education. *Child Adolesc. Psychiatry Ment. Health* **2024**, *18*, 132. [CrossRef] [PubMed]
4. Karcher, N.R.; Hicks, R.; Schiffman, J.; Asarnow, J.R.; Calkins, M.E.; Dauberman, J.L.; Garrett, C.D.; Koli, R.L.; Larrauri, C.A.; Loewy, R.L.; et al. Youth Mental Health Screening and Linkage to Care. *Psychiatr. Serv.* **2023**, *74*, 727–736. [CrossRef] [PubMed]
5. Fonseca-Pedrero, E.; Díez-Gómez, A.; Pérez-Albéniz, A.; Al-Halabí, S.; Lucas-Molina, B.; Debbané, M. Youth Screening Depression: Validation of the Patient Health Questionnaire-9 (PHQ-9) in a Representative Sample of Adolescents. *Psychiatry Res.* **2023**, *328*, 115486. [CrossRef] [PubMed]
6. Zuckerbrot, R.A.; Cheung, A.; Jensen, P.S.; Stein, R.E.K.; Laraque, D.; Levitt, A.; Birmaher, B.; Campo, J.; Clarke, G.; Emslie, G.; et al. Guidelines for Adolescent Depression in Primary Care (GLAD-PC): Part I. Practice Preparation, Identification, Assessment, and Initial Management. *Pediatrics* **2018**, *141*, e20174081. [CrossRef] [PubMed]
7. Carroll, H.A.; Hook, K.; Perez, O.F.R.; Denckla, C.; Vince, C.C.; Ghebrehiwet, S.; Ando, K.; Touma, M.; Borba, C.P.C.; Fricchione, G.L.; et al. Establishing Reliability and Validity for Mental Health Screening Instruments in Resource-Constrained Settings: Systematic Review of the PHQ-9 and Key Recommendations. *Psychiatry Res.* **2020**, *291*, 113236. [CrossRef] [PubMed]
8. Viswanathan, M.; Wallace, I.F.; Cook Middleton, J.; Kennedy, S.M.; McKeeman, J.; Hudson, K.; Rains, C.; Vander Schaaf, E.B.; Kahwati, L. Screening for Depression and Suicide Risk in Children and Adolescents. *JAMA* **2022**, *328*, 1543–1556. [CrossRef] [PubMed]
9. Guedes, F.B.; Cerqueira, A.; Gaspar, S.; Gaspar, T.; Moreno, C.; de Matos, M.G. Quality of Life and Well-Being of Adolescents in Portuguese Schools. *Child Indic. Res.* **2023**, *16*, 1381–1394. [CrossRef] [PubMed]
10. Carvalho, M.; Branquinho, C.; Moraes, B.; Cerqueira, A.; Tomé, G.; Noronha, C.; Gaspar, T.; Rodrigues, N.; Matos, M.G. de Positive Youth Development, Mental Stress and Life Satisfaction in Middle School and High School Students in Portugal: Outcomes on Stress, Anxiety and Depression. *Children* **2024**, *11*, 681. [CrossRef] [PubMed]
11. Chae, D.; Lee, J.; Lee, E.H. Internal Structure of the Patient Health Questionnaire-9: A Systematic Review and Meta-Analysis. *Asian Nurs. Res. (Korean. Soc. Nurs. Sci.)* **2025**, *19*, 1–12. [CrossRef] [PubMed]
12. Lamela, D.; Soreira, C.; Matos, P.; Morais, A. Systematic Review of the Factor Structure and Measurement Invariance of the Patient Health Questionnaire-9 (PHQ-9) and Validation of the Portuguese Version in Community Settings. *J. Affect. Disord.* **2020**, *276*, 220–233. [CrossRef] [PubMed]

13. Rocha, A.C.M.; Lusquinhos, L.; João, A.L.d.S.; de Oliveira, I.R.; de Pinho, L.M.G. Cross-Cultural Adaptation and Psychometric Validation of the Cognitive Distortions Questionnaire for Adolescents (CD-Quest-T) in European Portuguese. *Arch. Psychiatr. Nurs.* **2026**, *63*, 152160. [CrossRef] [PubMed]
14. Gonçalves, M.; Santiago, L. Ansiedade Em Crianças e Adolescentes-Validação Da Escala GAD-7. Master's Thesis, Universidade de Coimbra, Coimbra, Portugal, 2019.
15. Beaton, D.E.; Bombardier, C.; Guillemin, F.; Ferraz, M.B. Guidelines for the Process of Cross-Cultural Adaptation of Self-Report Measures. *Spine* **2000**, *25*, 3186–3191. [CrossRef] [PubMed]
16. Gagnier, J.J.; de Arruda, G.T.; Terwee, C.B.; Mokkink, L.B.; Consensus Group; Ahmed, S.; Beaton, D.; Sawatzky, R.; Apfelbacher, C.; Ruseckaite, R.; et al. COSMIN Reporting Guideline for Studies on Measurement Properties of Patient-reported Outcome Measures: Version 2.0. *Qual. Life Res.* **2025**, *34*, 1901–1911. [CrossRef] [PubMed]
17. Rakshasa-Loots, A.M.; Hamana, T.; Fanqa, B.; Lindani, F.; van Wyhe, K.; Kruger, S.; Laughton, B. IsiXhosa Translation of the Patient Health Questionnaire (PHQ-9) Shows Satisfactory Psychometric Properties for the Measurement of Depressive Symptoms [Stage 2]. *Brain Neurosci. Adv.* **2023**, *7*, 23982128231194452. [CrossRef] [PubMed]
18. Díaz de León Castañeda, C.; Díaz-Castro, L. Psychometric Properties of the Spanish Version of the Patient Health Questionnaire in Mexican Adolescents. *Front. Psychol.* **2025**, *16*, 1634804. [CrossRef] [PubMed]
19. Hua, L.L.; Lee, J.; Rahmandar, M.H.; Sigel, E.J. Suicide and Suicide Risk in Adolescents. *Pediatrics* **2024**, *153*, e2023064800. [CrossRef] [PubMed]
20. National Association of School Nurses. Elevating the Role of School Nurses in School-Based Mental and Behavioral Health: A Consensus Document. *NASN Sch. Nurse* **2024**, *39*, 301–304. [CrossRef] [PubMed]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.