psychiatry international

MDPI

Article

Assessment of Longitudinal Measurement Invariance of Short Versions of the CES-D in Maternal Caregivers

Luis Villalobos-Gallegos [1,*], Salvador Trejo [1], Diana Mejía-Cruz [2], Aldebarán Toledo-Fernández [3] and Diana Alejandra González García [1]

[1] School of Medicine and Psychology, Autonomous University of Baja California-Tijuana, Tijuana 22390, Mexico; strejo@uabc.edu.mx (S.T.); alejandra.gonzalez89@uabc.edu.mx (D.A.G.G.)
[2] Psychology Department, Sonora Institute of Technology, Ciudad Obregon 85860, Mexico; diana.mejia@itson.edu.mx
[3] School of Psychology, Anahuac University Mexico, Huixquilucan 52786, Mexico; aldebaran.toledofe@anahuac.mx
* Correspondence: villalobos.luis@uabc.edu.mx

Abstract

We tested the longitudinal invariance of seven short versions of the Center for Epidemiological Studies Depression Scale (CES-D) in maternal caregivers, following recent analytic recommendations for ordered categorical responses. Data for this study were drawn from the Longitudinal Studies in Child Abuse and Neglect (LONGSCAN) consortium, based on responses from 427 maternal caregivers across five waves corresponding to their children's ages: 4, 6, 12, 14, and 16 years. We employed a comprehensive approach using differences in two approximate fit indices (CFI and RMSEA), the chi-square difference test (χ^2), and a sensitivity analysis based on predicted response differences. Only one version demonstrated full invariance across all levels, while the others showed only partial evidence for loading or threshold invariance. These findings highlight concerns regarding the use of brief CES-D versions in longitudinal research, particularly over extended time periods. They also underscore the need to reassess whether item content aligns with current definitions of depressive syndrome. Our results suggest that evaluating the longitudinal invariance of short depression measures is essential to ensure the validity of conclusions about changes over time.

Keywords: longitudinal invariance; depression; maternal caregivers; CES-D; clinical assessment

check for updates

Academic Editor: Paul E Rapp

Received: 5 August 2025
Revised: 6 September 2025
Accepted: 14 October 2025
Published: 16 October 2025

Citation: Villalobos-Gallegos, L.; Trejo, S.; Mejía-Cruz, D.; Toledo-Fernández, A.; García, D.A.G. Assessment of Longitudinal Measurement Invariance of Short Versions of the CES-D in Maternal Caregivers. *Psychiatry Int.* **2025**, *6*, 126. https://doi.org/10.3390/psychiatryint6040126

Copyright: © 2025 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 (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

In the study of emotional development, the family is a central factor in understanding complex processes such as emotional regulation and emotional distress. A child's emotional development is closely related to several elements within the family system, including the observation of significant others, parenting practices, and the emotional climate [1]. These determinants, in turn, are strongly associated with the mental health status of the parents—particularly depressive symptoms in the maternal caregiver—which play a decisive role in the child's optimal emotional development [2], regardless of whether the caregiver is the biological mother or a substitute figure [3].

A large body of evidence has linked maternal caregiver depression to several negative outcomes in children, such as blunted neural activity [4] and reduced use of secondary control coping strategies [5]. It is also associated with impaired interactions, exemplified

by lower parent-adolescent relationship quality [6], which correlates with depressive symptoms during adolescence. This is significant, given that the heritability of depression is reported to be less than 50% [7], with the remaining variance explained by environmental factors. Studies analyzing trajectories of maternal caregiver depression have found that chronic patterns over time increase the likelihood of psychopathology in offspring, compared to less chronic trajectories (e.g., minimal or moderate depression) [8].

One of the most common methods for assessing depressive symptoms is the use of self-report scales. This approach offers several advantages: it requires minimal training, is time-efficient, and can be easily implemented in clinical settings. However, one of its most important limitations is the need for equivalence in the factor structure across subgroups of the studied population—known as measurement invariance (MI) [9]. When MI is achieved, we can confidently interpret between-group differences as true variations rather than measurement artifacts [10]. For example, if the factor structure is not invariant between males and females, then observed gender differences may reflect measurement bias rather than "true differences" in the latent construct. Such artifact may result from a lack of equivalence in factor loadings (weak invariance) or other parameters of the factor model (strong or strict invariance) [9].

A second form of invariance pertains to repeated measures. It is assumed that a measurement must retain the same structure within a population over time—this is referred to as longitudinal invariance (LI) [11]. LI is particularly relevant in two scenarios (though not limited to them): (1) when evaluating the efficacy of an intervention, the absence of LI increases the risk of bias in outcome assessment, making it difficult to determine whether changes are due to the treatment or to measurement inconsistencies; and (2) in developmental research, where it is expected that the measurement structure remains stable so that changes in item responses reflect only quantitative changes in the latent trait.

The Center for Epidemiological Studies–Depression Scale (CES-D) [12] is one of the most widely used instruments for measuring depressive symptoms. Several studies have tested LI in the CES-D among adolescents and adults [13–16], finding that a first-order model with four second-order factors remained invariant over time in versions ranging from 18 to 20 items. In addition to full versions, the CES-D has several short forms that have been previously published, validated, and are commonly used to optimize resources. Through a literature review, we identified seven different short versions ranging from 5 to 11 items, developed using various methods: two five-item versions derived from factor analysis and discriminant function analysis, respectively [17]; a nine-item version based on item endorsement probabilities among individuals with major depressive disorder [18]; two 10-item versions developed using item-total correlations [19] and the Rasch measurement model [19]; and both 10- and 11-item versions selected from factor analysis results in the original study [20].

However, a key requirement when shortening a measure is to evaluate its psychometric properties as if it were a new instrument, to ensure that reliability and validity are preserved. This principle also applies to longitudinal invariance. To our knowledge, only one study has assessed Kohout's [20] 11-item version in a Korean sample [21], and there is limited evidence regarding the other short forms. Despite this scarcity, the use of CES-D short versions is common in longitudinal research. Examples include nationally representative studies such as the Canadian Longitudinal Study of Aging [22], the China Health and Retirement Longitudinal Study [23], and other investigations exploring the role of depression in episodic drinking [24] and perfectionism [25]. This also includes studies measuring depression in caregiver samples [26,27], such as the Longitudinal Studies of Child Abuse and Neglect [28], which assessed maternal caregiver depressive symptoms from childhood through adolescence.

In summary, longitudinal research on depressive symptoms in maternal caregivers is essential for understanding the intergenerational transmission of depression and related emotional difficulties. It is also crucial to ensure that measurements in such studies reflect true changes in latent traits rather than biases stemming from inconsistent factor structures. Therefore, we aimed to test the longitudinal invariance of seven short versions of the CES-D using data from the Longitudinal Studies of Child Abuse and Neglect (LONGSCAN). We anticipate that, at least for Kohout's 11-item version, our findings will align with those from the Korean sample, supporting its longitudinal invariance.

2. Materials and Methods

2.1. Participants

In brief, the LONGSCAN study recruited participants across five sites: Baltimore, Chicago, Seattle, San Diego, and North Carolina. Maternal caregivers of children aged four years or younger (depending on the site) were invited to participate, including those with a history of maltreatment or considered at risk for maltreatment (55%), as well as community controls (children not considered at risk). Recruitment criteria for maltreatment and at-risk participants varied by site: in Baltimore, participants were referred from a medical institution based on risk criteria during the first year of life; in Chicago, children with confirmed maltreatment reports were referred for family intervention; in Seattle, children were reported to social services for suspected maltreatment prior to official determination; in San Diego, children were removed from their homes and placed in foster care within the first four years of life due to confirmed maltreatment; and in North Carolina, infants identified as high-risk at birth were recruited through a state public program. At every site except San Diego, dyads with children without prior maltreatment reports were also recruited from the community.

This recruitment process resulted in an initial sample of 1443 caregiver–child dyads, composed primarily of African American families (52%) residing in urban communities (69%). LONGSCAN employed a prospective longitudinal design, beginning with recruitment and continuing through multiple follow-up assessments. At various points in the study—particularly during the early stages—data were also provided by informants other than the children. Participants were contacted annually to assess maltreatment, with additional evaluations conducted every two years, continuing until the children reached age 18 (for more details on measures, procedures, and participant characteristics, see [28]). For the present study, we selected cases in which maternal caregivers completed all five CES-D assessments of depressive symptoms, collected when children were 4, 6, 12, 14, and 16 years old.

2.2. Measure

The CES-D is a widely used scale for assessing the severity of depressive symptoms across diverse populations [29]. It was developed by integrating items from previously validated depression measures to capture depressed mood, feelings of guilt or worthlessness, helplessness or hopelessness, psychomotor retardation, loss of appetite, and sleep disturbance [12]. The CES-D measures depressive symptoms experienced during the past week, using four response options: *Rarely or none of the time* (less than 1 day), *Some or a little of the time* (1–2 days), *Occasionally or a moderate amount of time* (3–4 days), and *Most or all of the time* (5–7 days). In the LONGSCAN study, the full 20-item version of the CES-D was administered; accordingly, we selected the corresponding items for each short version evaluated in this study. A summary of the seven short forms analyzed is presented in Table 1.

Table 1. Summary of the CES-D Brief versions analyzed in this study.

#	Item	A	B	C	D	E	F	G
1	Bothered	0	0	1	0	1	1	0
2	Appetite	0	0	0	0	0	0	1
3	Blues	0	0	1	0	0	1	0
4	As good	0	0	0	0	0	1	0
5	Mind	0	0	1	0	1	1	0
6	Depressed	1	1	1	1	1	0	1
7	Effort	0	0	1	1	1	1	1
8	Hopeful	0	1	0	0	1	1	0
9	Failure	0	0	0	0	0	1	0
10	Fearful	0	0	0	0	1	1	0
11	Sleep	1	0	1	1	1	0	1
12	Happy	0	1	1	1	1	0	1
13	Talked less	0	0	0	0	0	0	0
14	Lonely	1	0	0	1	1	1	1
15	Unfriendly	0	0	0	1	0	1	1
16	Enjoyed	0	0	1	1	0	0	1
17	Crying	1	1	0	0	0	0	0
18	Felt sad	0	0	1	1	0	0	1
19	Disliked	0	1	0	1	0	0	1
20	Get Going	1	0	0	1	1	0	1
Number of items		5	5	9	10	10	10	11

Notes: 0-Item is not in the version; 1-Item is in the version. Versions were developed by (A/B) [17]; (C) [18]; (D/G) [20]; (E) [30]; and (F) [19].

2.3. Statistical Analysis

Descriptive statistics for baseline data (age 4) were presented using means and standard deviations for numerical variables, and frequencies and percentages for categorical variables. The first step was to test the assumption of data being Missing Completely at Random (MCAR), using Little's test. For this purpose, we used the function na. test from the *misty* package (v0.6.2) in R (v4.5.1).

The analysis of measurement invariance was conducted using Mplus 8 [31]. CES-D items were treated as ordered categorical indicators. Given that the most common scoring practice in previous studies involves the use of sum scores, models were specified with a single factor; no modifications or correlated error terms were added. Since items were treated as ordered categorical variables, the analysis was performed using the mean- and variance-adjusted weighted least squares (WLSMV) estimator. Missing data were handled using pairwise deletion. All procedures followed recommendations from the literature [32].

In brief, the first step involved fitting a baseline model that imposed no constraints but included covariance terms between time-dependent factors and error terms. Subsequently, equality constraints were imposed on factor loadings, thresholds, and unique variances to test invariance at each level. For model identification, one indicator was selected as a marker variable—an observable item assumed to be invariant across all time points. To support invariance at a given level (loadings, thresholds, or unique variances), we applied the following criteria: changes in two approximate fit indices (ΔAFIs), specifically a difference in Root Mean Square Error of Approximation (ΔRMSEA) < 0.015 and a difference in Comparative Fit Index (ΔCFI) < 0.01 [33], along with a chi-square difference test ($\Delta\chi^2$) p-value < 0.05 (performed using the DIFFTEST option in Mplus)

If a model failed to meet both ΔAFI and $\Delta\chi^2$ criteria, it was considered non-invariant. However, if only one criterion was not met, we conducted a sensitivity analysis to assess the magnitude of the violation. This procedure involved comparing expected response

probabilities between the non-invariant model and its preceding step. A model was considered non-invariant if any response option at any time point showed a difference in expected probability greater than 0.050 [32]. This procedure was computed using R software (v4.5.1) [34].

2.4. Ethical Considerations

Study procedures and measures were approved by the Institutional Review Boards (IRBs) of each local site, as well as by the study's Coordinating Center. Approval for the Coordinating Center was obtained from the University of North Carolina at Chapel Hill Public Health–Nursing Institutional Review Board (IRB #99-0449). Federal and local regulations regarding human subject protection were followed, particularly those related to consent and confidentiality. Caregivers provided written informed consent to participate and received monetary compensation at each visit. Because LONGSCAN primarily focused on child maltreatment, all sites were legally mandated to report cases to local Child Protection Services agencies. For more details on specific maltreatment reporting procedures, see the article by Knight et al. [35].

3. Results

3.1. Participant Characteristics

A sample of 452 maternal caregivers with complete data was included in this study. The average age at baseline was approximately 32 years. Marital status was predominantly single or married. The sample was also composed mostly of individuals who identified as Black, with the majority being homemakers. Approximately half of the participants reported a yearly family income ranging from USD 5000 to USD 19,000 (see Table 2). The result from Little's test was: $\chi^2 = 44.92$, df = 43, $p = 0.391$. Therefore, we assume that the data meet the MCAR assumption.

Table 2. Maternal caregivers demographics at age 4 (n = 427).

	M (SD) or Freq. (%)
Age	32.8 (9.9)
Completed years of education	11.8 (2.0)
Race	
White	137 (29.0)
Black	244 (51.7)
Hispanic	24 (5.1)
Native American	3 (0.6)
Asian	1 (0.2)
Mixed race/Other	18 (3.8)
Marital status	
Married	139 (29.4)
Single	209 (44.3)
Separated	30 (6.4)
Divorced/Widowed	49 (10.4)
Current employment	
Regular full-time/part-time	138 (29.2)
Retired/Disabled	103 (21.8)
Student	29 (6.1)
Homemaker	147 (31.1)
Other	10 (2.1)

Table 2. *Cont.*

	M (SD) or Freq. (%)
Family income per year	
<USD 5000	55 (11.7)
USD 5000–USD 9999	118 (25.0)
USD 10,000–USD 19,999	125 (26.5)
USD 20,000–USD 29,999	63 (13.3)
USD 30,000–USD 49,999	37 (7.8)
>USD 50,000	24 (5.1)

Note: 25 caregivers did not complete demographics at age 4.

3.2. Longitudinal Invariance

All versions achieved optimal goodness-of-fit (RMSEA < 0.05, CFI > 0.95) in most steps of the analysis, with the exception of version B in the unique factor model. However, only versions A, E, and F met the criteria for loading invariance, showing a non-significant $\Delta\chi^2$ and minimal changes in ΔAFIs. The remaining versions (B, C, D, and G) met the ΔAFI criteria for loading invariance but exhibited a significant $\Delta\chi^2$. Regarding threshold invariance, only version A satisfied both ΔAFI and $\Delta\chi^2$ criteria. The other versions—except for version B (C, D, E, F, and G)—met only the ΔAFI criterion. Version B failed to meet either ΔAFI or $\Delta\chi^2$ thresholds and was therefore considered non-invariant. None of the versions met the $\Delta\chi^2$ criterion for unique variance invariance; however, all showed evidence of invariance based on the ΔAFI criterion. For further details, please refer to Table 3.

Table 3. Assessment of longitudinal invariance in Brief CES-D versions.

	Versions						
	A	**B**	**C**	**D**	**E**	**F**	**G**
Baseline							
χ^2	286.178 *	410.513 *	1271.744 *	1715.894 *	1500.226 *	1748.924 *	1949.255 *
χ^2 df	215	215	845	1065	1065	1065	1310
CFI	0.987	0.951	0.965	0.946	0.960	0.924	0.949
RMSEA	0.027	0.045	0.034	0.037	0.030	0.038	0.033
RMSEA 90% C.I.	0.018–0.035	0.038–0.051	0.030–0.037	0.034–0.040	0.027–0.034	0.035–0.041	0.030–0.036
Loading							
χ^2	297.170 *	426.615 *	1336.005 *	1783.061 *	1531.195 *	1756.686 *	2024.312 *
χ^2 df	231	231	877	1101	1101	1101	1350
CFI	0.988	0.951	0.963	0.943	0.960	0.927	0.946
RMSEA	0.025	0.043	0.034	0.037	0.029	0.037	0.033
RMSEA 90% C.I.	0.016–0.033	0.037–0.049	0.030–0.038	0.034–0.040	0.026–0.033	0.033–0.040	0.030–0.036
ΔCFI	0.001	0.000	−0.001	−0.003	0.000	0.003	−0.003
ΔRMSEA	−0.002	−0.002	0.000	0.000	−0.001	−0.001	0.000
χ^2 difference (df)	13.080 (16)	28.024 (16) *	92.401 (32) *	107.600 (36) *	48.779 (36)	44.245 (36)	110.595 (40) *
Threshold							
χ^2	337.066 *	524.646 *	1481.033 *	1907.638 *	1624.174 *	1815.052 *	2148.992 *
χ^2 df	267	267	945	1177	1177	1177	1434
CFI	0.988	0.935	0.956	0.939	0.959	0.929	0.943
RMSEA	0.024	0.046	0.036	0.037	0.029	0.035	0.033
RMSEA 90% C.I.	0.015–0.032	0.040–0.052	0.032–0.039	0.034–0.040	0.026–0.032	0.032–0.038	0.030–0.036
ΔCFI	0.000	−0.016	−0.007	−0.004	−0.001	0.002	−0.003
ΔRMSEA	−0.001	0.003	0.002	0.000	0.000	−0.002	0.000
χ^2 difference (df)	45.457 (36)	132.440 (36) *	214.221 (68) *	199.328 (76) *	135.432 (76) *	103.430 (76) *	206.985 (84) *
Unique factor							
χ^2	379.502 *	574.682 *	1503.105 *	1897.263 *	1708.459 *	1902.932 *	2150.020 *
χ^2 df	287	287	981	1217	1217	1217	1478
CFI	0.984	0.928	0.957	0.943	0.954	0.924	0.947

Table 3. *Cont.*

	Versions						
	A	**B**	**C**	**D**	**E**	**F**	**G**
RMSEA	0.027	0.047	0.034	0.035	0.030	0.036	0.032
RMSEA 90% C.I.	0.019–0.034	0.041–0.052	0.031–0.038	0.025–0.032	0.027–0.033	0.027–0.033	0.023–0.029
ΔCFI	−0.004	−0.007	0.001	0.004	−0.005	−0.005	0.003
ΔRMSEA	0.003	0.001	−0.002	−0.002	0.001	0.001	−0.001
χ^2 difference (df)	44.946 (20) *	63.392 (20) *	64.485 (36) *	63.644 (40) *	104.473 (40) *	122.101(40) *	71.919 (44) *

Note: χ^2 difference test was performed to compare against the previous step. CFI: Comparative Fit Index; RMSEA: Root Mean Square Error of Approximation; C.I.: Confidence Interval; df: Degrees of Freedom. * $p < 0.05$.

3.3. Sensitivity Analysis

Results indicated that Version A showed no substantial differences in expected response probabilities, with the largest deviation being −0.044 for the item "Sleep" in the response category "Some or a little of the time" at age 16. Therefore, this version was considered invariant.

The remaining versions did not meet the <0.05 threshold for at least one item. In the case of Version G, all items exceeded the criterion at age 16. Table 4 summarizes the main findings, and complete data on differences in predicted probabilities are provided in the Supplementary Materials.

Table 4. Summary of the sensitivity analysis results.

Version	Items with Probability Difference > 0.05	Categories	Wave	Comparison
A	None	None	None	Threshold vs. unique
B	Was not performed.	Not Applicable	Not Applicable	Not Applicable
C	Enjoyed	1 and 2	16	Loading vs. configural
D	Effort	1 and 2	6	Threshold vs. loading
E	Bothered	0 and 1	16	Threshold vs. loading
	Depressed	0 and 1	4	
	Hopeful	2 and 3	4 and 16	
	Hopeful	3	14	
F	As good	2 and 3	6	Loading vs. configural
	Effort	0	4	
	Hopeful	2 and 3	4, 14 and 16	
G	Appetite	3	16	Loading vs. configural
	Depressed	1 and 3	16	
	Effort	1 and 2	6	
	Effort	3	16	
	Sleep	1 and 3	16	
	Happy	2	16	
	Lonely	3	16	
	Unfriendly	0 and 3	16	
	Enjoyed	1 and 2	4 and 6	
	Enjoyed	0 and 2	16	
	Felt sad	1 and 3	16	
	Disliked	0 and 3	16	
	Get Going	1 and 3	16	

Response categories: 0 = rarely or none of the time; 1 = some or a little of the time; 2 = occasionally or a moderate amount of time; 3 = most or all of the time.

4. Discussion

This study tested the longitudinal invariance of short versions of the CES-D in a sample of maternal caregivers. We found that three versions provided sufficient evidence for loading invariance, which is considered the weakest form of invariance [9]. However, only Version A met the statistical criteria for the most robust level of invariance—unique variance.

It is important to note that Version A is one of the shortest versions identified in the literature. Due to its brevity, it may offer advantages in terms of time efficiency. On the other hand, a plausible concern is that with only five items, its coverage of the depressive syndrome may be limited, potentially reducing its clinical utility. Nevertheless, in addition to demonstrating superior model fit in this study, previous research [36] has shown that this version is equally effective for screening and assessing depression when compared to longer forms. This suggests that the items included in Version A may reflect a set of "core characteristics" of depression, such as mood disturbance, loss of interest, and sleep/fatigue [37].

We initially hypothesized that Version G would yield results similar to those reported in a previous study [21], which found evidence of short-term invariance (4 years) based solely on ΔAFI. However, that study reported substantial changes in ΔCFI in the strict invariance model over a longer period (10 years). In our analysis, ΔAFI values were relatively small, but the sensitivity analysis revealed violations of invariance, particularly in the final wave (age 16). Moreover, the previous study did not report $\Delta\chi^2$ or conduct sensitivity analyses. Taken together, these findings suggest that Version G may be suitable for shorter observation periods (e.g., 4 years), but its structural stability may not hold over longer durations (e.g., 10 or more years).

The remaining versions showed varying degrees of longitudinal invariance violations. For example, Version B failed to meet the $\Delta\chi^2$ criterion in the loading model and did not satisfy either ΔAFI or $\Delta\chi^2$ in the threshold model. Versions C, D, E, and F achieved acceptable ΔAFIs but failed to meet $\Delta\chi^2$ criteria and showed violations in their respective sensitivity analyses—loading invariance for Versions C and D, and threshold invariance for Versions E and F.

One key insight from the sensitivity analysis is the magnitude of invariance violations: as differences in expected response probabilities increase, so do discrepancies in the corresponding model parameters. This procedure is useful because it allows researchers to identify specific items and time points where invariance may be compromised. As previously noted, Version G exhibited several expected probability differences above the recommended cutoff, particularly at age 16. This pattern suggests that while the version may be relatively invariant over shorter time spans, the stability of its factor structure is likely affected over longer periods (e.g., >10 years). These findings align with previous studies [14–16,21].

Additionally, the sensitivity analysis identified the item "hopeful" as problematic. This may indicate that the item "I felt hopeful about the future" taps into a construct somewhat distinct from the rest of the CES-D items. It is important to note that this item primarily reflects the construct of "hope", which is more closely associated with positive affect [38]. Therefore, its relationship with depression measurement models may not be strong enough to yield invariant results.

Two factors may explain the lack of invariance observed in our study: the time elapsed between measurement waves and the content of specific items. Although depression is generally expected to be stable over time in adults—including maternal caregivers—there is limited evidence regarding whether its latent structure remains constant throughout adulthood. Significant affective changes occur during the transition from parenting an infant to parenting an adolescent [39], particularly given that parenting stress tends to

increase during adolescence [40]. In our study, several years passed between assessments (e.g., eight years between the second and third waves), and the evolving role of maternal caregiving, along with other stressors, may have altered the associations among depressive symptoms. Recent research has shown that such changes can occur during developmental transitions or following acute stress events [41,42], and may be reflected in the item covariances of the CES-D, resulting in non-invariance over time.

A second explanation for non-invariance concerns the content of certain CES-D items. The CES-D was developed approximately 50 years ago [12], based on items from earlier questionnaires. As a result, its conceptualization of depressive syndrome may not fully align with more recent classification systems, such as the DSM-5 or ICD-11 [43]. While critiques exist regarding the clinical utility and validity of these systems, numerous field studies and revisions have shaped the contemporary understanding of depression [44]. Some scholars have recommended updating legacy measures like the CES-D, Hamilton Rating Scale for Depression, and Beck Depression Inventory to reflect current conceptualizations [45]. Moreover, a weak theoretical foundation in a measure has been shown to affect its factor structure [46]

Additional evidence regarding item content comes from a previous study that tested the invariance of the CES-D by classifying items into four domains: depressed affect, positive affect, interpersonal, and somatic. That study found that longitudinal invariance was achieved only after removing the positive affect items ("*as good*," "*hopeful*," "*happy*," and "*enjoyed*") [47]. In our analysis, Version A was the only version that excluded all of these items, consisting solely of depressed affect and somatic symptoms—domains more closely aligned with DSM-based conceptualizations of depression [47]. From this perspective, it is plausible that a CES-D version focusing exclusively on the "core characteristics" of depression, while excluding positive affect dimensions, may offer greater stability over time.

Taken together, our findings and prior evidence support two recommendations: (1) prioritize the use of short CES-D versions that exclude positive affect items and ensure alignment with contemporary definitions of depression; and (2) exercise caution when using versions other than Version A in studies spanning more than 10 years. These recommendations should be interpreted carefully, as they are based on a subset of CES-D versions and may not generalize to other instruments (e.g., Patient Health Questionnaire-9, Beck Depression Inventory).

Most previous studies assessing longitudinal invariance have based their conclusions primarily on changes in approximate fit indices (ΔAFIs). Following these recommendations, most of the short versions analyzed in our study could be considered longitudinally invariant. However, the behavior of fit indices for ordered categorical responses differs from that observed with continuous numeric data. Therefore, the use of the chi-square difference test is recommended. This procedure, in conjunction with the use of the WLSMV estimator (also known as diagonally weighted least squares)—which is considered the most appropriate for this type of data [48]—enhances the robustness of the analysis. We consider this a strength of our study, as we employed a comprehensive approach that incorporated ΔAFIs, $\Delta\chi^2$, and sensitivity analyses, in alignment with current methodological recommendations in the field [49]

It is important to note that our study was not without limitations. First, due to the prospective nature of the study—spanning more than 10 years—participant attrition reduced the overall sample size. While the initial sample included over 1000 participants [50], the final sample analyzed in this study consisted of 452 maternal caregivers. However, we believe that the comprehensive approach used to assess measurement invariance helps mitigate this limitation. A secondary limitation is that approximately half of the sample

was composed of maternal caregivers of children at risk of maltreatment. As a result, the sample represents a population with characteristics that differ from the general population. Therefore, further studies are needed to determine whether these findings can be generalized to more diverse populations.

A final insight emerging from our results relates to studies employing a network analysis approach. Unlike latent variable modeling, which assumes depression to be an unobservable construct explaining variation in a set of indicators, network analysis conceptualizes depression as the result of interrelations among those indicators (i.e., depressive symptoms) [51]. One previous study applied this approach to maternal caregivers using the CES-D. When network stability—an analog to longitudinal invariance (LI) in this framework—was examined, the results were inconclusive: adjacency analyses showed moderate correlations, but the overall network structure was not invariant across waves [52]. These findings suggest that the structural stability of the CES-D requires further investigation using diverse methodological approaches.

Alongside prior findings evaluating other depression measures [53], our results reinforce the notion that testing longitudinal invariance in depression scales is a necessary prerequisite for developmental and treatment research. This step helps prevent misleading or biased conclusions. Furthermore, assessing longitudinal invariance may be essential to determine whether the classical assumption of structural stability in depression over time holds true in adults, or whether the factor structure is susceptible to changes driven by environmental influences (e.g., treatment-induced changes in symptom correlations, as suggested by [54]).

5. Conclusions

We believe that this study raises awareness about the use of brief versions of the CES-D in longitudinal research. Version A—a five-item form including *Depressed*, *Sleep*, *Lonely*, *Crying*, and *Get going*—was more closely aligned with the "core characteristics" of depression. It demonstrated superior model fit and met the criteria for all levels of longitudinal invariance. In contrast, the remaining versions, which included items measuring Positive Affect, did not yield results supporting longitudinal invariance, with violations primarily occurring at the later assessment points (ages 14 and 16). Building on previous evidence, these findings suggest that longitudinal invariance may be compromised when depression measures incorporate content beyond core symptomatology. Assessing longitudinal invariance is therefore strongly recommended as a critical step to enhance researchers' confidence in the validity of their conclusions, particularly when using scales in developmental or treatment studies. Moreover, evaluating longitudinal invariance may be essential to determine whether the classical assumption of structural stability in depression across time holds true in adults, or whether the factor structure is susceptible to changes driven by environmental influences.

Supplementary Materials: The following supporting information can be downloaded at https://www.mdpi.com/article/10.3390/psychiatryint6040126/s1.

Author Contributions: L.V.-G. conceptualized the manuscript; L.V.-G. and S.T. performed statistical analysis; L.V.-G. and A.T.-F. draft drafted the manuscript. A.T.-F. and D.M.-C. performed a critical review of the manuscript. All authors contributed to manuscript revision, read, and approved the submitted version. All authors have read and agreed to the published version of the manuscript.

Funding: The data for Longitudinal Studies of Child Abuse and Neglect (LONGSCAN) Assessments 0–12, have been given to the National Data Archive on Child Abuse and Neglect for public distribution by Desmond K. Runyan, Howard Dubowitz, Diana J. English, Jonathan Kotch, Alan Litrownik, Richard Thompson and Terri Lewis, and The LONGSCAN Investigator Group. Funding for the project

was provided by the Office on Child Abuse and Neglect (OCAN), Children's Bureau, Administration for Children and Families, Dept. of Health and Human Services (The National Center on Child Abuse and Neglect (NCCAN), under the Office of Human Services funded this consortium of studies during the early years of data collection from 04/01/1991 until NCCAN became part of OCAN in 1998.) (Award Numbers: 90CA1467, 90CA1481, 90CA1466, 90CA1458, 90CA1572, 90CA1569, 90CA1568, 90CA1566, 90CA1678, 90CA1681, 90CA1680, 90CA1676, 90CA1677, 90CA1679, 90CA1744, 90CA1745, 90CA1746, 90CA1747, 90CA1748, 90CA1749). The collector of the original data, the funder, NDACAN, Cornell University and their agents or employees bear no responsibility for the analyses or interpretations presented here.

Institutional Review Board Statement: Ethical review and approval were waived for this study because it was a secondary data analysis. The original study was conducted in accordance with the Declaration of Helsinki, and approved by University of North Carolina at Chapel Hill Public Health–Nursing Institutional Review Board (IRB #99-0449). For more information please refer to the source of the data at https://www.ndacan.acf.hhs.gov/datasets/dataset-details.cfm?ID=170 (accessed on 13 October 2025).

Informed Consent Statement: This study was a secondary data analysis. Every caregiver provided an individual consent and their respective children also provide their assent for participation in the study. Individual site consent, assent, and related human subjects' protocols were approved by local IRBs and the study Coordinator Center.

Data Availability Statement: The original datasets used in this analysis can be found in the National Data Archive on Child Abuse and Neglect (NDACAN) website: https://www.ndacan.acf.hhs.gov/datasets/dataset-details.cfm?ID=170 (accessed on 13 October 2025).

Acknowledgments: The data used in this publication were made available by the National Data Archive on Child Abuse and Neglect, Cornell University, Ithaca, NY, and have been used with permission. Data from Longitudinal Studies of Child Abuse and Neglect (LONGSCAN) Assessments were originally collected by Desmond K. Runyan, Howard Dubowitz, Diana J. English, Jonathan Kotch, Alan Litrownik, Richard Thompson and Terri Lewis and The LONGSCAN Investigator Group. Funding for the project was provided by the Office on Child Abuse and Neglect (OCAN), Children's Bureau, Administration for Children and Families, Dept. of Health and Human Services (The National Center on Child Abuse and Neglect (NCCAN), under the Office of Human Services funded this consortium of studies during the early years of data collection from 04/01/1991 until NCCAN became part of OCAN in 1998).

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

References

1. Morris, A.S.; Silk, J.S.; Steinberg, L.; Myers, S.S.; Robinson, L.R. The Role of the Family Context in the Development of Emotion Regulation. *Soc. Dev.* **2007**, *16*, 361–388. [CrossRef]
2. Wu, V.; East, P.; Delker, E.; Blanco, E.; Caballero, G.; Delva, J.; Lozoff, B.; Gahagan, S. Associations Among Mothers' Depression, Emotional and Learning-Material Support to Their Child, and Children's Cognitive Functioning: A 16-Year Longitudinal Study. *Child Dev.* **2018**, *90*, 1952–1968. [CrossRef] [PubMed]
3. Tully, E.C.; Iacono, W.G.; McGue, M. An Adoption Study of Parental Depression as an Environmental Liability for Adolescent Depression and Childhood Disruptive Disorders. *Am. J. Psychiatry* **2008**, *165*, 1148–1154. [CrossRef]
4. Meyer, A.; Bress, J.N.; Hajcak, G.; Gibb, B.E. Maternal Depression Is Related to Reduced Error-Related Brain Activity in Child and Adolescent Offspring. *J. Clin. Child. Adolesc. Psychol.* **2018**, *47*, 324–335. [CrossRef]
5. Reising, M.M.; Bettis, A.H.; Dunbar, J.P.; Watson, K.H.; Gruhn, M.; Hoskinson, K.R.; Compas, B.E. Stress, Coping, Executive Function, and Brain Activation in Adolescent Offspring of Depressed and Nondepressed Mothers. *Child Neuropsychol.* **2018**, *24*, 638–656. [CrossRef]
6. Withers, M.C.; Cooper, A.; Rayburn, A.D.; McWey, L.M. Parent-Adolescent Relationship Quality as a Link in Adolescent and Maternal Depression. *Child. Youth Serv. Rev.* **2016**, *70*, 309–314. [CrossRef]
7. Corfield, E.C.; Yang, Y.; Martin, N.G.; Nyholt, D.R. A Continuum of Genetic Liability for Minor and Major Depression. *Transl. Psychiatry* **2017**, *7*, e1131. [CrossRef] [PubMed]

8. Matijasevich, A.; Murray, J.; Cooper, P.J.; Anselmi, L.; Barros, A.J.D.; Barros, F.C.; Santos, I.S. Trajectories of Maternal Depression and Offspring Psychopathology at 6 Years: 2004 Pelotas Cohort Study. *J. Affect. Disord.* **2015**, *174*, 424–431. [CrossRef]
9. Meredith, W.; Teresi, J.A. An Essay on Measurement and Factorial Invariance. *Med. Care* **2006**, *44*, S69–S77. [CrossRef]
10. Putnick, D.L.; Bornstein, M.H. Measurement Invariance Conventions and Reporting: The State of the Art and Future Directions for Psychological Research. *Dev. Rev.* **2016**, *41*, 71–90. [CrossRef]
11. Widaman, K.F.; Ferrer, E.; Conger, R.D. Factorial Invariance within Longitudinal Structural Equation Models: Measuring the Same Construct across Time. *Child Dev. Perspect.* **2010**, *4*, 10–18. [CrossRef]
12. Radloff, L.S. The CES-D Scale: A Self-Report Depression Scale for Research in the General Population. *Appl. Psychol. Meas.* **1977**, *1*, 385–401. [CrossRef]
13. Contrada, R.J.; Boulifard, D.A.; Idler, E.L.; Krause, T.J.; Labouvie, E.W. Course of Depressive Symptoms in Patients Undergoing Heart Surgery: Confirmatory Analysis of the Factor Pattern and Latent Mean Structure of the Center for Epidemiologic Studies Depression Scale. *Psychosom. Med.* **2006**, *68*, 922–930. [CrossRef]
14. Ferro, M.A.; Speechley, K.N. Factor Structure and Longitudinal Invariance of the Center for Epidemiological Studies Depression Scale (CES-D) in Adult Women: Application in a Population-Based Sample of Mothers of Children with Epilepsy. *Arch. Womens Ment. Health* **2013**, *16*, 159–166. [CrossRef]
15. Motl, R.W.; Dishman, R.K.; Birnbaum, A.S.; Lytle, L.A. Longitudinal Invariance of the Center for Epidemiologic Studies-Depression Scale among Girls and Boys in Middle School. *Educ. Psychol. Meas.* **2005**, *65*, 90–108. [CrossRef]
16. Verhoeven, M.; Sawyer, M.G.; Spence, S.H. The Factorial Invariance of the CES-D during Adolescence: Are Symptom Profiles for Depression Stable across Gender and Time? *J. Adolesc.* **2013**, *36*, 181–190. [CrossRef]
17. Shrout, P.E.; Yager, T.J. Reliability and Validity of Screening Scales: Effect of Reducing Scale Length. *J. Clin. Epidemiol.* **1989**, *42*, 69–78. [CrossRef] [PubMed]
18. Santor, D.A.; Coyne, J.C. Shortening the CES-D to Improve Its Ability to Detect Cases of Depression. *Psychol. Assess.* **1997**, *9*, 233–243. [CrossRef]
19. Cole, J.C.; Rabin, A.S.; Smith, T.L.; Kaufman, A.S. Development and Validation of a Rasch-Derived CES-D Short Form. *Psychol. Assess.* **2004**, *16*, 360–372. [CrossRef]
20. Kohout, F.J.; Berkman, L.F.; Evans, D.A.; Cornoni-Huntley, J. Two Shorter Forms of the CES-D Depression Symptoms Index. *J. Aging Health* **1993**, *5*, 179–193. [CrossRef] [PubMed]
21. Park, B.S.; Lee, K.; Shin, C.; Choi, K.; Bae, S.W. Longitudinal Measurement Invariance of the Korean Version of the CES-D-11 Scale. *Sage Open* **2022**, *12*, 21582440221117799. [CrossRef]
22. Raina, P.; Wolfson, C.; Kirkland, S.; Griffith, L.E.; Balion, C.; Cossette, B.; DIonne, I.; Hofer, S.; Hogan, D.; van den Heuvel, E.R.; et al. Cohort Profile: The Canadian Longitudinal Study on Aging (CLSA). *Int. J. Epidemiol.* **2019**, *48*, 2066. [CrossRef] [PubMed]
23. Ni, Y.; Tein, J.Y.; Zhang, M.; Yang, Y.; Wu, G. Changes in Depression among Older Adults in China: A Latent Transition Analysis. *J. Affect. Disord.* **2017**, *209*, 3–9. [CrossRef]
24. Kim, A.J.; Sherry, S.B.; Nealis, L.J.; Mushquash, A.; Lee-Baggley, D.; Stewart, S.H. Do Symptoms of Depression and Anxiety Contribute to Heavy Episodic Drinking? A 3-Wave Longitudinal Study of Adult Community Members. *Addict. Behav.* **2022**, *130*, 107295. [CrossRef]
25. Graham, A.R.; Sherry, S.B.; Stewart, S.H.; Sherry, D.L.; McGrath, D.S.; Fossum, K.M.; Allen, S.L. The Existential Model of Perfectionism and Depressive Symptoms: A Short-Term, Four-Wave Longitudinal Study. *J. Couns. Psychol.* **2010**, *57*, 423–438. [CrossRef]
26. Krauss, S.; Orth, U.; Robins, R.W. Family Environment and Self-Esteem Development: A Longitudinal Study from Age 10 to 16. *J. Pers. Soc. Psychol.* **2020**, *119*, 457–478. [CrossRef] [PubMed]
27. Larsen, J.K.; van den Broek, N.; Verhagen, M.; Burk, W.J.; Vink, J.M. A Longitudinal Study on Changes in Food Parenting Practices during COVID-19 and the Role of Parental Well-Being. *Appetite* **2023**, *180*, 106331. [CrossRef]
28. Runyan, D.K.; Curtis, P.A.; Hunter, W.M.; Black, M.M.; Kotch, J.B.; Bangdiwala, S.; Dubowitz, H.; English, D.; Everson, M.D.; Landsverk, J. Longscan: A Consortium for Longitudinal Studies of Maltreatment and the Life Course of Children. *Aggress. Violent Behav.* **1998**, *3*, 275–285. [CrossRef]
29. Vilagut, G.; Forero, C.G.; Barbaglia, G.; Alonso, J. Screening for Depression in the General Population with the Center for Epidemiologic Studies Depression (Ces-d): A Systematic Review with Meta-Analysis. *PLoS ONE* **2016**, *11*, e0155431. [CrossRef]
30. Andresen, E.M.; Malmgren, J.A.; Carter, W.B.; Patrick, D.L. Screening for depression in well older adults: Evaluation of a short form of the CES-D. *Am. J. Prev. Med.* **1994**, *10*, 77–84. [CrossRef]
31. Muthén, L.K.; Muthén, B.O. *Mplus User's Guide*, 8th ed.; Muthén & Muthén: Los Angeles, CA, USA, 2017.
32. Liu, Y.; Millsap, R.E.; West, S.G.; Tein, J.Y.; Tanaka, R.; Grimm, K.J. Testing Measurement Invariance in Longitudinal Data with Ordered-Categorical Measures. *Psychol. Methods* **2017**, *22*, 486–506. [CrossRef]
33. Sass, D.A.; Schmitt, T.A.; Marsh, H.W. Evaluating Model Fit With Ordered Categorical Data Within a Measurement Invariance Framework: A Comparison of Estimators. *Struct. Equ. Model.* **2014**, *21*, 167–180. [CrossRef]

34. R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2019; ISBN 3900051003.
35. Knight, E.D.; Smith, J.B.; Dubowitz, H.; Litrownik, A.J.; Kotch, J.B.; English, D.; Everson, M.D.; Runyan, D.K. Reporting Participants in Research Studies to Child Protective Services: Limited Risk to Attrition. *Child Maltreat.* **2006**, *11*, 257–262. [CrossRef] [PubMed]
36. Furukawa, T.; Anraku, K.; Hiroe, T.; Takahashi, K.; Kitamura, T.; Hirai, T.; Takahashi, K.; Iida, M. Screening for Depression among First-Visit Psychiatric Patients: Comparison of Different Scoring Methods for the Center for Epidemiologic Studies Depression Scale Using Receiver Operating Characteristic Analyses. *Psychiatry Clin. Neurosci.* **1997**, *51*, 71–78. [CrossRef] [PubMed]
37. Kennedy, S.H. Core Symptoms of Major Depressive Disorder: Relevance to Diagnosis and Treatment. *Dialog. Clin. Neurosci.* **2008**, *10*, 271–277. [CrossRef] [PubMed]
38. Ritschel, L.A.; Cassiello-Robbins, C. Hope and Depression and Personality Disorders. *Curr. Opin. Psychol.* **2023**, *49*, 101507. [CrossRef]
39. van der Giessen, D.; Hollenstein, T.; Hale, W.W.; Koot, H.M.; Meeus, W.; Branje, S. Emotional Variability in Mother-Adolescent Conflict Interactions and Internalizing Problems of Mothers and Adolescents: Dyadic and Individual Processes. *J. Abnorm. Child. Psychol.* **2014**, *43*, 339–353. [CrossRef]
40. Daundasekara, S.S.; Beauchamp, J.E.S.; Hernandez, D.C. Parenting Stress Mediates the Longitudinal Effect of Maternal Depression on Child Anxiety/Depressive Symptoms. *J. Affect. Disord.* **2021**, *295*, 33–39. [CrossRef]
41. Liu, D.; Yu, M.; Zhang, X.; Cui, J.; Yang, H. Adolescent Anxiety and Depression: Perspectives of Network Analysis and Longitudinal Network Analysis. *BMC Psychiatry* **2024**, *24*, 619. [CrossRef]
42. Wang, Y.; Hu, Z.; Feng, Y.; Wilson, A.; Chen, R. Changes in Network Centrality of Psychopathology Symptoms between the COVID-19 Outbreak and after Peak. *Mol. Psychiatry* **2020**, *25*, 3140–3149. [CrossRef]
43. Monroe, S.M.; Anderson, S.F. Depression: The Shroud of Heterogeneity. *Curr. Dir. Psychol. Sci.* **2015**, *24*, 227–231. [CrossRef]
44. Blashfield, R.K.; Keeley, J.W.; Flanagan, E.H.; Miles, S.R. The Cycle of Classification: DSM-I Through DSM-5. *Annu. Rev. Clin. Psychol.* **2014**, *10*, 25–51. [CrossRef]
45. Fried, E.I.; Flake, J.K.; Robinaugh, D.J. Revisiting the Theoretical and Methodological Foundations of Depression Measurement. *Nat. Rev. Psychol.* **2022**, *1*, 358–368. [CrossRef]
46. Fischer, R.; Karl, J.A.; Luczak-Roesch, M.; Hartle, L. Why We Need to Rethink Measurement Invariance: The Role of Measurement Invariance for Cross-Cultural Research. *Cross Cult. Res.* **2025**, *59*, 147–179. [CrossRef]
47. Bergenfeld, I.; Kaslow, N.J.; Yount, K.M.; Cheong, Y.F.; Johnson, E.R.; Clark, C.J. Measurement Invariance of the Center for Epidemiologic Studies Scale–Depression Within and Across Six Diverse Intervention Trials. *Psychol. Assess.* **2023**, *35*, 805–820. [CrossRef]
48. Li, C.H. Confirmatory Factor Analysis with Ordinal Data: Comparing Robust Maximum Likelihood and Diagonally Weighted Least Squares. *Behav. Res. Methods* **2016**, *48*, 936–949. [CrossRef]
49. Tse, W.W.Y.; Lai, M.H.C.; Zhang, Y. Does Strict Invariance Matter? Valid Group Mean Comparisons with Ordered-Categorical Items. *Behav. Res. Methods* **2024**, *56*, 3117–3139. [CrossRef] [PubMed]
50. Knight, E.D.; Runyan, D.K.; Dubowitz, H.; Brandford, C.; Kotch, J.; Litrownik, A.; Hunter, W. Methodological and Ethical Challenges Associated With Child Self-Report of Maltreatment. *J. Interpers. Violence* **2000**, *15*, 760–775. [CrossRef]
51. Borsboom, D.; Cramer, A.O.J. Network Analysis: An Integrative Approach to the Structure of Psychopathology. *Annu. Rev. Clin. Psychol.* **2013**, *9*, 91–121. [CrossRef]
52. Santos, H.P.; Kossakowski, J.J.; Schwartz, T.A.; Beeber, L.; Fried, E.I. Longitudinal Network Structure of Depression Symptoms and Self-Efficacy in Low-Income Mothers. *PLoS ONE* **2018**, *13*, e0191675. [CrossRef]
53. Fried, E.I.; van Borkulo, C.D.; Epskamp, S.; Schoevers, R.A.; Tuerlinckx, F.; Borsboom, D. Measuring Depression over Time Or Not? Lack of Unidimensionality and Longitudinal Measurement Invariance in Four Common Rating Scales of Depression. *Psychol. Assess.* **2016**, *28*, 1354–1367. [CrossRef] [PubMed]
54. Blanken, T.F.; van der Zweerde, T.; van Straten, A.; van Someren, E.J.W.; Borsboom, D.; Lancee, J. Introducing Network Intervention Analysis to Investigate Sequential, Symptom-Specific Treatment Effects: A Demonstration in Co-Occurring Insomnia and Depression. *Psychother. Psychosom.* **2019**, *88*, 52–54. [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.