

Psychometric Properties of the Edinburgh Postpartum Depression Scale in Mexican Mothers

José Moral de la Rubia, Ana Luisa Rodríguez López

School of Psychology, Universidad Autónoma de Nuevo León, Monterrey, Mexico
Email: jose.morald@uanl.edu.mx

How to cite this paper: Moral de la Rubia, J., & Rodríguez López, A. L. (2025). Psychometric Properties of the Edinburgh Postpartum Depression Scale in Mexican Mothers. *Psychology*, 16, 905-944.
<https://doi.org/10.4236/psych.2025.167052>

Received: June 15, 2025

Accepted: July 28, 2025

Published: July 31, 2025

Copyright © 2025 by author(s) and Scientific Research Publishing Inc.
This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).
<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

The Edinburgh Postnatal Depression Scale (EPDS) is widely used in epidemiological, research, and clinical contexts. Although it has clearly established cut-off points, its psychometric properties have not been thoroughly analyzed—particularly considering the ordinal nature of its items. To address this gap, the present psychometric study aimed to analyze the items, test a unidimensional (single-factor) model, estimate overall internal consistency, describe score distribution, and evaluate concurrent construct validity in relation to perceived social support (negative association), depression during pregnancy (risk factor), and marital status (being a single mother as a risk factor). The EPDS, along with a Perceived Social Support Scale (PSSS), was administered to an incidental sample of 142 Mexican women who were at least two months postpartum. The ten items demonstrated discriminating power and internal consistency. The one-factor model was validated with an acceptable to good fit using the weighted least squares mean and variance adjusted estimation method. Internal consistency was very high (McDonald's $\omega = .93$; Green-Yang's $\omega = .91$). The total score distribution deviated from normality due to positive skewness. Expectations were met for all three aspects of validation. The correlation between the EPDS total score and the PSSS total score was negative ($r = -.66$, 95% BCa $CI[-.75, -.54]$). The mean score of mothers with depression during pregnancy was significantly higher than that of mothers without depression during pregnancy ($md = 6.15$, 95% BCa $CI[3.96, 8.83]$). Single mothers without partners had a significantly higher mean rank than married and cohabiting mothers. It is concluded that the EPDS is a reliable and valid instrument and should be interpreted using percentile ranks. Future studies are encouraged to replicate these findings and to treat marital status as a polytomous qualitative variable, nuanced by the presence or absence of a partner, when included in statistical analyses.

Keywords

Psychometrics, Item Analysis, Reliability, Confirmatory Factor Analysis, Concurrent Validity

1. Introduction

1.1. Definition and Prevalence of Postpartum Depression

Postpartum depression, also known as postnatal depression, is a mood disorder that can affect women after childbirth (Saharoy et al., 2023). It is characterized by the persistent presence, lasting at least two weeks, of clinically significant depressive symptoms, such as profound sadness, loss of interest or pleasure in usual activities, fatigue, sleep and appetite disturbances, feelings of worthlessness or excessive guilt, difficulties in establishing an emotional bond with the newborn, and, in severe cases, suicidal thoughts or delusional ideation (World Health Organization [WHO], 2024).

Unlike postpartum maternal blues, which lasts only a few days and begins within the first six weeks after delivery, postpartum depression can begin at any time within the first year after childbirth and persists for weeks or months, requiring professional evaluation and treatment (Santiago Sanabria et al., 2022). Its etiology is multifactorial, encompassing biological factors, such as hormonal changes; psychological factors, including depression during pregnancy, a history of affective disorders, or a depressive personality; and social factors, such as the level of perceived support, marital dissatisfaction, marital status (being a single mother without a partner or recently divorced), or economic hardship (Agrawal et al., 2022; Quesada & Chinchilla, 2022).

Malpartida-Ampudia (2020) states that postpartum depression typically manifests between two and four days after childbirth and may last approximately two to three weeks. However, Jadresic (2017) notes that, in hospital settings, a significant percentage of postpartum depression cases do not arise immediately but rather within the first month or later. The author highlights that symptoms generally peak in intensity between the eighth and twelfth weeks postpartum. During the first six weeks, distinguishing between postpartum maternal blues and postpartum depression syndrome is more challenging (Manurung & Setyowati, 2021).

The fifth edition of the American Psychiatric Association's Diagnostic and Statistical Manual of Mental Disorders (DSM-V; APA, 2013) establishes a four-week period after childbirth as the timeframe for the onset of postpartum depression. The International Classification of Diseases, in its tenth revision (ICD-10; WHO, 2024), defined a six-week period after childbirth, whereas in its eleventh revision, it adopted the four-week criterion, aligning with the DSM-5 (WHO, 2024). Meanwhile, the Centers for Disease Control and Prevention (CDC, 2008) extended the risk period to 12 months postpartum.

Wang et al. (2021a, 2021b) report that postpartum depression affects 17.22% (95% *CI* [16, 18.51]) of women worldwide, with a higher prevalence in underdeveloped countries than in developed ones. In Mexico, prevalence ranges between 12% and 14%, according to figures from the *Secretaría de Salud* (2016), and two out of ten women experience clinical depression during pregnancy and postpartum based on ICD-10 criteria (*Secretaría de Salud*, 2016). Santiago Sanabria et al. (2023) found a prevalence of postpartum depression similar to that reported by the Ministry of Health using the Edinburgh Postnatal Depression Scale (EPDS; Cox et al., 1987) in a sample of 717 women. These researchers observed that 106 women had an EPDS score of 10 or higher (Cox et al., 1987), representing a postpartum depression prevalence of 14.9%. Similarly, Flores-Ramos et al. (2025), using 12 as the EPDS cut-off score (Alvarado-Esquivel et al., 2006; Macías-Cortés et al., 2020), found a perinatal depression prevalence of 14.18% in a sample of 141 women from Sinaloa.

Organización Panamericana de la Salud (2016) states that the first step in treating depression during pregnancy or postpartum is detection and identification. In this regard, de Castro et al. (2016) analyzed the level of coverage in perinatal depression care in Mexico. They evaluated 211 obstetric units at the secondary and tertiary levels of care belonging to the Ministry of Health, the Institute of Security and Social Services for State Workers (ISSSTE), the Mexican Institute of Social Security (IMSS), and IMSS-Opportunities. The results showed that although 64% of obstetric hospitals provided some form of mental health care, only 37% had protocols for detecting perinatal mental disorders, and 40% provided primary care.

On the other hand, Jiménez-Brito (2022) presents data indicating that most mothers in Mexico lack Social Security benefits. The author reports that 69% of women who are mothers do not engage in any economic activity, meaning there is no guarantee of access to maternal health services, paid leave, childcare services (daycare centers), among others. This lack of support exposes them to a higher risk of postpartum depression due to the vulnerable conditions in which motherhood is experienced.

1.2. Measuring Postpartum Depression

The Edinburgh Postnatal Depression Scale (EPDS), developed by Cox et al. (1987), is the most widely used instrument specifically designed for detecting symptoms of postpartum depression. It is a self-administered questionnaire consisting of 10 items, each with four response options scored from 0 to 3, resulting in a total score ranging from 0 to 30. The EPDS is unidimensional and demonstrates good internal consistency, with a Cronbach's alpha coefficient of .87 reported in the original study (Cox et al., 1987). This high internal consistency has been replicated in several studies, with alpha values ranging from .80 to .87 (Gibson et al., 2009). In a validation study conducted with a Mexican sample, Alvarado-Esquivel et al. (2006) reported a Cronbach's alpha of .82, while a study

carried out in a Spanish population by Garcia-Esteve et al. (2003) found a value of .81. It has also demonstrated temporal reliability, with an intraclass correlation coefficient (ICC) of .92 over a three-day period (Kernot et al., 2015), as well as a Pearson product-moment correlation of .40 between pre- and post-pregnancy scores over approximately five months (Alfayumi-Zeadna et al., 2022).

Precisely, in a study of temporal stability, a hierarchical three-factor model was proposed—anhedonia (defined by items 1 and 2), anxiety (items 3 to 6), and sadness (items 7 to 10)—organized under a general factor of depression (Alfayumi-Zeadna et al., 2022; Song et al., 2024). Israeli researchers Alfayumi-Zeadna et al. (2022) reported that the hierarchical model, estimated using maximum likelihood, showed a good fit in a paired-sample dataset of 332 Bedouin women. The model fit was satisfactory both at the first measurement, conducted between the 26th and 38th weeks of pregnancy ($\chi^2[26] = 30.90, p = .23; SRMR = .034; RMSEA = .025; CFI = .99$), and at the second measurement, conducted between the second and fourth months postpartum ($\chi^2[26] = 29.59, p = .29; SRMR = .031; RMSEA = .020; CFI = .99$). Good fit was also observed in the multigroup analysis ($\chi^2[52] = 60.47, p = .20$) for the unconstrained model. Chinese researchers Song et al. (2024) similarly reported that the hierarchical three-factor model met the criteria for metric, scalar, and strict invariance.

Another widely used measurement instrument is the Postpartum Depression Screening Scale (PDSS), developed by Beck and Gable (2000). It consists of 35 Likert-type items with five response options, grouped into seven factors: sleeping/eating disturbances, anxiety/insecurity, emotional lability, cognitive impairment, loss of self, guilt/shame, and thoughts of self-harm. The total scale shows excellent internal consistency, with a Cronbach's alpha coefficient of .95; reliability coefficients for the subscales range from .83 (sleeping/eating disturbances) to .94 (loss of self). A Spanish version of the PDSS is available (Beck & Gable, 2003) and has been validated in Mexico by Lara et al. (2013). These authors proposed a cut-off score of 77 for subclinical postpartum depression and 95 for clinical depression.

Another option is to use structured interviews developed in accordance with major mental health diagnostic systems, such as the Structured Clinical Interview for DSM-5 (Lyubanova et al., 2021), the Structured Clinical Interview for ICD-11 (Bai et al., 2023), or the WHO Flexible Interview for ICD-11 (FLII-11) developed by Reed et al. (2024). Finally, a general depression scale may also be used; among these, the Revised Beck Depression Inventory (BDI-II) is the most widely used instrument (Yang et al., 2023).

The EPDS has been translated into multiple languages and validated in several countries, generally yielding positive results. These studies have primarily focused on assessing linguistic equivalence, estimating internal consistency, and establishing clinical cut-off points, using structured interviews or the BDI-II as reference criteria (Shafian et al., 2022). In other words, they have been oriented towards applied and clinically relevant aspects. However, no study has examined the as-

sumed unidimensional structure underlying its 10 items, nor have the psychometric properties of individual items or the distribution of total scores been described—even though these elements are fundamental in psychometric analysis (Furr, 2021).

1.3 Objectives of the Study

The purpose of this study is to validate the EPDS by addressing the existing research gap. Its objectives are: 1) to analyze the properties of the 10 items that compose the scale using Classical Test Theory (Gorham & Randall, 2022), describing their distributions, assessing their discriminative capacity, and evaluating reliability; 2) to test the one-factor model; 3) to estimate overall internal consistency; 4) to describe the distribution of the total EPDS score; and 5) to verify its convergent construct validity by assessing whether an inverse relationship exists between EPDS and perceived social support (Cho et al., 2022; Hajipoor et al., 2021), whether women who experienced depression during pregnancy score higher on the EPDS than those who did not (Agrawal et al., 2022; Flores-Ramos et al., 2025; Quesada & Chinchilla, 2022), and whether marital status is associated with postpartum depression, with single mothers without a partner being at higher risk. In all these analyses, the 10 items that compose the scale are considered ordinal variables rather than discrete or continuous quantitative variables (Kyriazos & Poga-Kyriazou, 2023).

The present study proposes a one-factor model with ten indicators as the hypothesized structure. If this model fails to demonstrate a good fit, the hierarchical three-factor model proposed by Alfayumi-Zeadna et al. (2022) will be evaluated. Alternatively, a different model will be explored.

The EPDS was chosen over the PDSS and BDI-II for this study because of its brevity (10 items compared to 35 and 21, respectively) and its unidimensional structure (as opposed to two or more factors), making it quick and easy to administer and complete. It is freely available, providing a cost-effective option for healthcare settings. Unlike other instruments that emphasize somatic symptoms (e.g., weight changes, sleep disturbances), the EPDS focuses on the core symptoms of depression, such as depressed mood, anhedonia (loss of interest or pleasure), and anxiety. Importantly, it includes a dedicated item on suicidal ideation (item 10), allowing for the immediate identification of this critical risk factor. Its widespread use also facilitates communication among healthcare providers. Additionally, the EPDS has been validated in numerous languages and cultural contexts, enhancing its value as a flexible and reliable screening tool.

2. Method

2.1. Participants

A non-probabilistic incidental sampling method was used. The inclusion criteria were: being female, at least 18 years old, and in the postpartum period (between 2 and 12 months after delivery). The exclusion criteria were: having a visual impair-

ment that prevented completion of the questionnaire, being underage, being less than two months or more than 12 months postpartum, and being pregnant. Participants who met any of the exclusion criteria did not complete the questionnaire or the additional items. Consequently, missing data were handled by excluding those cases. Mothers within the first six weeks postpartum were excluded to obtain a sample reflecting pure postpartum depression, uncontaminated by the postpartum blues syndrome (Jadresic, 2017; Manurung & Setyowati, 2021).

The invitation to participate in the research was sent via social networks (Instagram, Facebook and WhatsApp) to specific groups of mothers with different interests, such as breastfeeding, sales, complementary feeding, parenting, food allergies, and baby dance classes. The invitation, disseminated by the second author, included details about the study and a link to access the questionnaire. The most challenging sample to obtain was that of women engaged in unpaid work (i.e. home duties), so the second author visited a health center in the municipality of General Escobedo to collect data in person.

The questionnaire was completed by 165 women; however, 21 were excluded for meeting one or more exclusion criteria, and two additional cases were removed based on the elimination criterion. The final sample therefore consisted of 142 participants.

We considered this sample size to be adequate, as it allows for approximately 14 participants per item ($142:10 \approx 14:1$) and 7 participants per parameter to be estimated in a single-factor model ($142:20 \approx 7:1$) with high measurement weights ($\hat{\lambda}$ ranging from .641 to .934, with an average of .758).

2.2. Instruments of Measurement

The electronic form included the informed consent, questions on sociodemographic, occupational, and pregnancy history, as well as the Edinburgh Postnatal Depression Scale (EPDS; Cox et al., 1987) and the Perceived Social Support Scale from Partner and Significant Others (PSSS; Cienfuegos-Martínez, 2010). The form was designed using the Google Forms platform.

The EPDS, developed by Cox et al. (1987), is a self-report scale created in health centers in Livingston and Edinburgh to assist primary care professionals in detecting postpartum depression in mothers. It is recommended that the scale be administered after the sixth postpartum week, as symptoms during the initial stage (the first six weeks after delivery) may reflect postpartum blues associated with hormonal changes. The scale is available in both English and Spanish; the Spanish version was used in this study. See Appendix.

Each EPDS item offers four response options. Mothers are asked to select the option that best reflects how they felt during the previous seven days. Items 1 and 2 are scored from 0 (first option) to 3 (last option), with higher scores indicating greater symptom severity. Items 3 through 10 are reverse-keyed and scored from 3 (first option) to 0 (last option). The total score is obtained by summing the scores for all items, resulting in a possible range from 0 to 30. Scores from 0 to 11 indicate no depression, 12 to 19 suggest mild postpartum depression, 20 to 29

indicate moderate depression, and a score of 30 indicates severe depression (Alvarado-Esquivel et al., 2006; Macías-Cortés et al., 2020).

The PSSS, developed by Cienfuegos-Martínez (2010), consists of 44 Likert-type items with five response options (“never,” “almost never,” “sometimes,” “almost always,” and “always”), scored from 1 to 5. The total score ranges from 44 to 220, with higher scores indicating greater perceived social support. The scale shows excellent internal consistency (44 items; Cronbach’s $\alpha = .90$). It comprises two subscales: support from the partner (24 items; Cronbach’s $\alpha = .87$) and support from significant others (20 items; Cronbach’s $\alpha = .88$).

2.3. Procedures and Ethical Aspects

The research design was non-experimental and cross-sectional. The study was approved by the Research Sub-directorate of the School of Psychology at the Autonomous University of Nuevo León.

To comply with ethical standards, the Code of Ethics of the Mexican Society of Psychology (Sociedad Mexicana de Psicología, 2010), and the guidelines of the American Psychological Association (APA, 2017), were used as a basis for informed consent procedures and privacy notices.

This study classifies it as risk-free research, as no intervention or intentional modification was made to the physiological, psychological, or social variables of the participants, in accordance with Article 17 of the General Health Law on Health Research (Cámara de Diputados del H. Congreso de la Unión, 2014).

2.4. Data Analysis

The analysis of the items began with a description of their distributions, followed by an assessment of their discriminative capacity, and concluded with an evaluation of their reliability. The distributions of the items, which are ordinal variables, were visualized using bar plots. Additionally, they were described using measures of central tendency (arithmetic mean, median, and mode), variability (mean absolute deviation, semi-interquartile range, and quartile coefficient of variation), and shape—specifically, Bowley’s (Bowley, 1901) quartile coefficient of skewness and Kelley’s (Kelley, 1923) percentile coefficients of skewness and kurtosis.

To assess the discriminating power of each item, high and low scoring groups were created based on the EPDS total score (the sum of the 10 items). The low-score group consisted of scores less than or equal to the 27th percentile of the total score, while the high-score group included scores greater than or equal to the 73rd percentile (Kelley, 1939). Differences in central tendency for each item between these two groups were tested using the Mann-Whitney U test (Mann & Whitney, 1947). In addition, mean differences were computed along with 95% bootstrap confidence intervals using the bias-corrected and accelerated (BCa) percentile method, based on 1,000 resamples with replacement. A significant mean difference of at least one quarter of the item range ($[3 - 0]/4 = .75$) was considered evidence of discriminative power (Moral, 2006).

The reliability of each item was assessed based on three criteria: 1) its correlation with the sum of the remaining items, 2) the reduction in overall internal consistency when the item was removed, and 3) its communality, or the variance it shared with the other items (Furr, 2021; Moral, 2006).

The correlation between each item and the sum of the remaining items was estimated using the polyserial correlation coefficient when the assumption of bivariate normality was met (Zheng & Cao, 2022), or the Spearman rank-order correlation coefficient (Spearman, 1904) otherwise. The normality assumption was tested using the chi-square goodness-of-fit test (Li, 2021).

Due to the violation of the tau-equivalence assumption, internal consistency was estimated using McDonald's omega coefficient (McDonald, 1999) and Green-Yang's categorical omega coefficient (Green & Yang, 2009). The assumption of tau-equivalence—or equal factor loadings on a general factor—was tested via confirmatory factor analysis (CFA). A one-factor model with 10 indicators was specified with the constraint that all factor loadings were equal. The model was estimated using the Weighted Least Squares Mean and Variance adjusted (WLSMV) method, which is appropriate for ordinal data and relies on polychoric correlations. These correlations assume an underlying bivariate normal distribution for the ordinal variables (Han, 2022). This assumption was tested using the chi-square goodness-of-fit test (Li, 2021).

Communality was estimated using the squared multiple correlation obtained by predicting each item from the remaining items. The predictive model was specified through path analysis, with parameters estimated using the WLSMV method. Communalities below .16 are generally considered low. Ideally, communalities should be at least .25, and preferably .50 or higher (Furr, 2021; Moral, 2006).

Regarding the second objective, a one-factor model with freely estimated factor loadings (with one loading fixed for identification) was tested using the WLSMV method. The significance of the factor loadings was assessed using the z-test. In addition, 95% bootstrap confidence intervals for the loadings were computed using the percentile method based on 1000 bootstrap samples (Efron & Narasimhan, 2020).

Model fit is considered good when the null hypothesis of exact fit is not rejected by the likelihood ratio chi-square test at the .05 significance level. Good fit is also indicated by a standardized likelihood ratio chi-square statistic (χ^2/df) value less than or equal to 2, Comparative Fit Index (*CFI*) and the Tucker-Lewis index (*TLI*)—or Normed Fit Index (*NFI*)—values greater than .95, a Root Mean Square Error of Approximation (*RMSEA*) value less than or equal to .05, and a Standardized Root Mean squared Residual (*SRMR*) value less than or equal to .08. An acceptable fit corresponds to values such as $p \geq .01$ for χ^2 , $\chi^2/df \leq 3$, *CFI* and *TLI* $\geq .90$, *RMSEA* $\leq .08$, and *SRMR* $\leq .1$. In contrast, a poor fit is defined by values such as $p < .01$ for χ^2 , $\chi^2/df > 3$, *CFI* and *TLI* $> .90$, *RMSEA* $> .08$, and *SRMR* $> .1$ (McNeish & Wolf, 2023; Rosseel, 2012; Rosseel et al., 2024). An Average Variance Extracted (*AVE*) greater than .5 and a McDonald's omega coefficient value equal to or higher than .7 indicate convergent validity in a measurement model (Cheung et

al., 2024).

Regarding the third objective, the internal consistency of the scale was estimated using both McDonald's omega coefficient (McDonald, 1999) and the categorical omega coefficient proposed by Green and Yang (2009). The latter was computed using the MBESS package in R (Kelley, 2023). Analogous to Cronbach's alpha, omega values in the range [.5, .6) were considered indicative of unacceptable internal consistency; values in the range [.6, .7), questionable; [.7, .8), acceptable; [.8, .9), good; and [.9, 1], excellent (Izah et al., 2023). However, omega values $\geq .95$ and $\chi^2/df < 1$ may indicate the presence of redundant items and an overfitted model (Byrne, 2016; Kline, 2016).

With respect to the fourth objective, the distribution of the EPDS total score was described using measures of central tendency (arithmetic mean, median, and Grenander's (Grenander, 1965) mode), variability (variance, standard deviation, and mean absolute deviation) and shape (coefficients of skewness and kurtosis based on Fisher's (Fisher, 1930) cumulants). Outliers were tested using Grubbs' (Grubbs, 1969) test, symmetry was examined using D'Agostino's (D'Agostino, 1970) test, and kurtosis was assessed using Anscombe & Glynn (1983) test. The distribution was visualized through a table of class intervals, a box plot, and a histogram with the estimated density and normal curves overlaid. Normality was assessed using the tests by D'Agostino, Belanger, and D'Agostino (D'Agostino et al., 1990); Shapiro and Wilk (1965), with Royston's (Royston, 1992) standardization; Anderson and Darling (1952); and Kolmogorov (1933) and Smirnov (1948), with p-values computed via Lilliefors' Monte Carlo simulation (Lilliefors, 1967).

Regarding the fifth objective, convergent construct validity was assessed by examining whether there is a negative correlation between the EPDS and the EAP using Pearson's product-moment correlation. To test its significance, the bootstrap BCa confidence interval and a one-sided bootstrap p-value ($H_0: \rho \geq 0$; $H_1: \rho < 0$) were computed. The significant difference in central tendency of the EPDS total score between groups of women with and without depression during pregnancy was tested using a one-sided Mann-Whitney U test (Mann & Whitney, 1947) at a 5% significance level. Tie correction was applied, but not Yates' continuity correction (Yates, 1934). Both asymptotic and bootstrap p-values were calculated (Conover, 1999).

The effect size was estimated using Rosenthal's (Rosenthal, 1991) r coefficient, Spearman's (Spearman, 1904) rank correlation coefficient (r_s), and Kerby's (Kerby, 2014) difference, which corresponds to Cureton's (Cureton, 1968) biserial rank correlation coefficient. This last measure represents the probability that a randomly selected value from the group with depression during pregnancy exceeds a randomly selected value from the group without depression. It takes values between -1 and 1 , where negative values indicate superiority of the second group (without depression) over the first group (with depression), and positive values indicate superiority of the first group over the second. It can be interpreted analogously to McGraw and Wong's (McGraw and Wong, 1992) Common Language

Effect Size (*CLES*) measure for comparing means between two independent groups. Considering the correspondence between *CLES* and Cohen's *d* statistic (Cohen, 1988), given by $CLES = \Phi(d / \sqrt{2})$, and the conventional cut-off points of .2, .5, and .8 for *d* stipulated by Cohen (1988), absolute values of Kerby's *d* in the interval [.5, .579] indicate a trivial effect size, [.579, .691] small, [.691, .788] medium, and [.788, 1] large.

The difference between the marital status groups was tested using the Kruskal & Wallis' (Kruskal & Wallis, 1952) test. The effect size was estimated by the epsilon squared (ϵ^2) and eta squared (η^2) coefficients. According to Cohen (1988), values of ϵ^2 and η^2 less than .01 are considered trivial, values between .01 and .06 are considered small, values between .06 and .14 are considered medium, and values of .14 or greater are considered large. Pairwise comparisons were conducted with Dunn's (Dunn, 1964) test applying Bonferroni's (Bonferroni, 1936) correction.

3. Results

3.1. Sample Description

The mean age of the mothers was 31.56, 95% BCa CI [30.73, 32.34] ranging from 20 to 44 with a sample standard deviation of 5.049. Among the 142 women, 110 (77.5%) were married, 22 (15.5%) were cohabiting, 6 (4.2%) were single without a partner, and 4 (2.8%) were single with a partner. In terms of educational level, one out of 142 women (.7%) reported having a primary education; 11 (7.7%) reported having a secondary education; seven (4.9%) reported having a high school education; nine (6.3%) reported having technical studies; 80 (56.3%) reported having an undergraduate or engineering degree; and 34 (23.9%) reported having postgraduate studies. Regarding work activity, 38% of the women were housewives, 43% had paid work in a company, and 19% were entrepreneurs, either through catalog sales or owning their own business.

Among the 142 participants, 71 (50%) were primiparous, while 71 (50%) were multiparous. Among the multiparous women, 42 (59.2%) had two children, 23 (32.4%) had three children, 4 (5.6%) had four children, and 2 (2.8%) had five children. When asked whether they had experienced at least one previous abortion, 38 (26.8%) responded yes, while 104 (73.2%) responded no. Additionally, 10 (7%) of the 142 women reported having been diagnosed with depression during pregnancy, whereas 132 (93%) had not.

The infants' ages ranged from 2 to 12 months. At the time the mothers completed the questionnaire, 21 (14.8%) were 2 months old, 15 (10.6%) were 3 months, 16 (11.3%) were 4 months, 14 (9.9%) were 5 months, 13 (9.2%) were 6 months, 13 (9.2%) were 7 months, 12 (8.5%) were 8 months, 9 (6.3%) were 9 months, 9 (6.3%) were 10 months, 12 (8.5%) were 11 months, and 8 (5.6%) were 12 months.

3.2. Description of the Distribution of the Items

Table 1 shows the measures of central tendency, variability, and shape of the 10

items that make up the EPDS, whose distributions are represented by a bar chart in **Figures 1-10**.

The central tendency of item 1 (“I have been able to laugh and see the funny side of things”), as assessed by mode and median, was “as much as always” (value 0). The arithmetic mean was .394, which rounded also is 0. Absolute variability ($MAD = .394$ and $RSI = .5$), and relative variability ($QCV = 1/3$) indicate a moderate-to-low variability. The distribution exhibited a long right tail ($PCS = 1$), and platykurtosis, characterized by widened shoulders ($PCK = .237$). Thus, the interquartile range coincided with the percentile range. See **Figure 1** and **Table 1**.

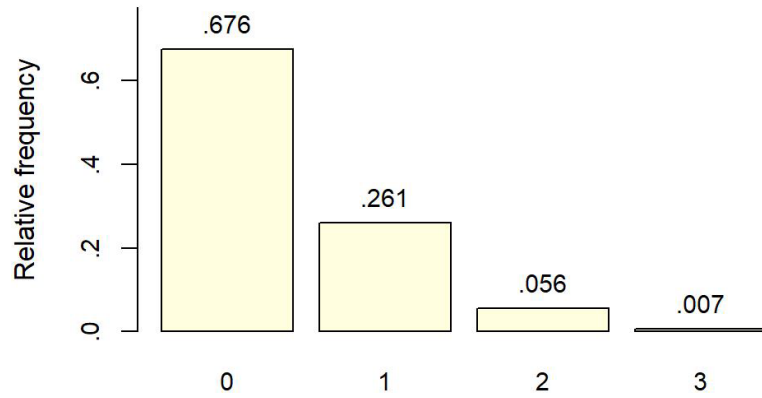
Items 2 and 7 were similar to item 1 (see **Figure 2** and **Figure 3**). Additionally, the median and mode coincided at 0 for item 10. The arithmetic mean of item 10 was very small ($m = .239$). Its variability was null, as measured by the semi-interquartile range and the coefficient of quartile variation, and very low, as measured by the mean absolute deviation ($MAD = .239$). It concentrated more on 0 than the previous three. Its profile also showed platykurtosis, but it was symmetrical (see **Figure 4**).

If a floor effect is defined as more than half of the responses being concentrated in the lowest category, then these four items exhibited a floor effect. If the threshold is raised to three-quarters, only item 10 would show a floor effect. In the remaining six items, the first category accounted for less than half of the responses. On the other hand, none of the items showed a ceiling effect.

Table 1. Descriptive measures of the 10 items of EPDS.

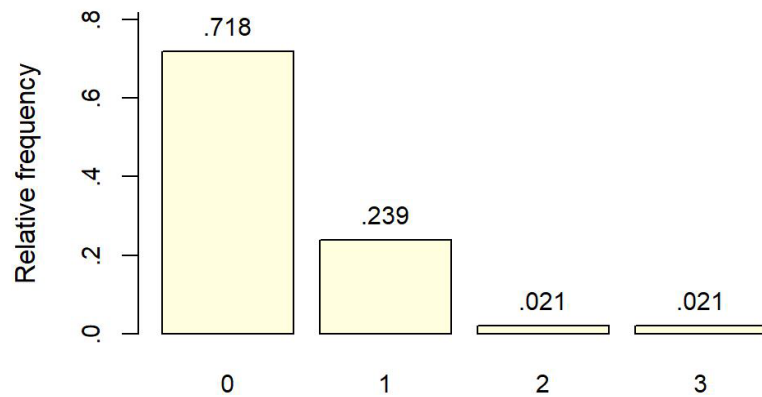
| Item of EPDS | Central tendency | | | Variation | | | Shape | | | Percentiles | |
|--------------|------------------|-----------|----------|------------|------------|------------|------------|------------|------------|-------------|------------|
| | <i>mdn</i> | <i>mo</i> | <i>m</i> | <i>MAD</i> | <i>SQR</i> | <i>QCV</i> | <i>QCS</i> | <i>PCS</i> | <i>PCK</i> | <i>P80</i> | <i>P90</i> |
| 1 | 0 | 0 | .394 | .394 | .5 | .333 | 1 | 1 | .237 | 1 | 1 |
| 2 | 0 | 0 | .345 | .345 | .5 | .333 | 1 | 1 | .237 | 1 | 1 |
| 3 | 2 | 2 | 2.014 | .507 | .5 | .143 | 1 | 0 | -.013 | 3 | 3 |
| 4 | 2 | 2 | 1.754 | .613 | .5 | .200 | -1 | -.333 | -.096 | 2 | 3 |
| 5 | 1 | 2 | 1.246 | .810 | 1 | .500 | 0 | 0 | .237 | 2 | 2 |
| 6 | 2 | 2 | 1.704 | .648 | .5 | .200 | -1 | -.333 | -.096 | 2 | 3 |
| 7 | 0 | 0 | .697 | .697 | .5 | .333 | 1 | 1 | -.013 | 1 | 2 |
| 8 | 1 | 1 | .718 | .577 | .5 | .333 | -1 | 0 | -.013 | 1 | 2 |
| 9 | 1 | 1 | .669 | .528 | .5 | .333 | -1 | -1 | .237 | 1 | 1 |
| 10 | 0 | 0 | .239 | .239 | 0 | 0 | Ind | 1 | -.263 | 0 | 1 |

Note. Measures of central tendency: *mdn* = median, *mo* = mode, and *m* = arithmetic mean. Measures of variation: *MAD* = median absolute deviation (from the median), *SQR* = semi-quartile range, and *QCV* = quartile coefficient of variation. Measures of shape: *QCS* = quartile coefficient of skewness, *PCS* = percentile coefficient of skewness, and *PCK* = percentile coefficient of kurtosis centered at 0. Criteria of high value (case): *P80* = percentile 80 and *P90* = percentile 90.



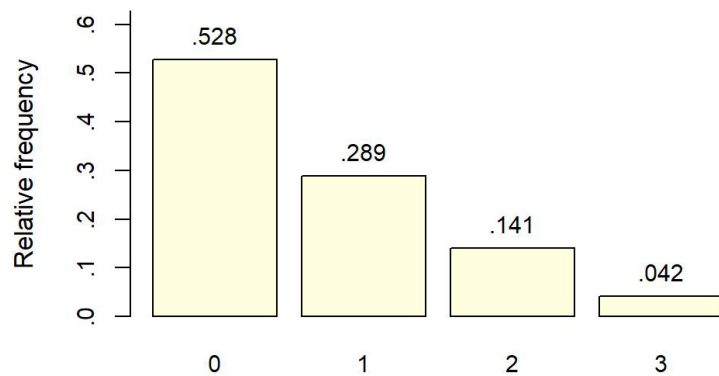
i1: I have been able to laugh and see the funny side

Figure 1. Bar diagram of the sample distribution of Item 1. Value labels: 0 = “As much as I always could”, 1 = “Not quite so much now”, 2 = “Definitely not so much now”, and 3 = “Not at all”.



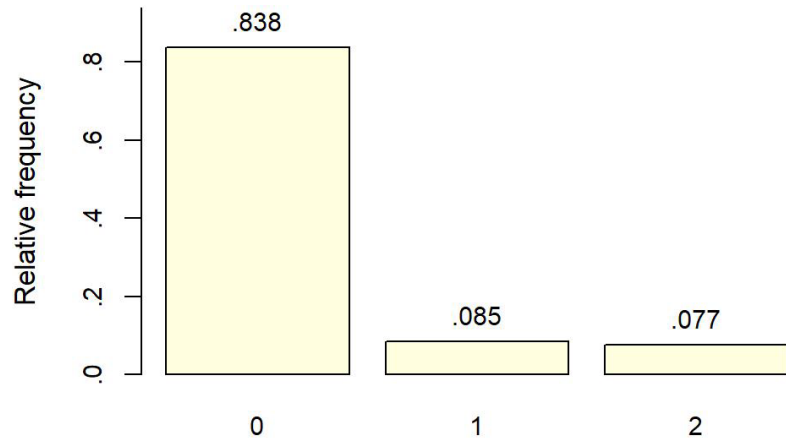
i2: I have looked forward with enjoyment to things

Figure 2. Bar diagram of the sample distribution of Item 2. Value labels: 0 = “As much as I ever did”, 1 = “Rather less than I used to”, 2 = “Definitely less than I used to”, and 3 = “Hardly at all”.



i7: I have been so unhappy that I have had difficulty sleeping

Figure 3. Bar diagram of the sample distribution of Item 7. Value labels: 0 = “No, not at all”, 1 = “No, not very often”, 2 = “Yes, sometimes”, and 3 = “Yes, most of the time”.

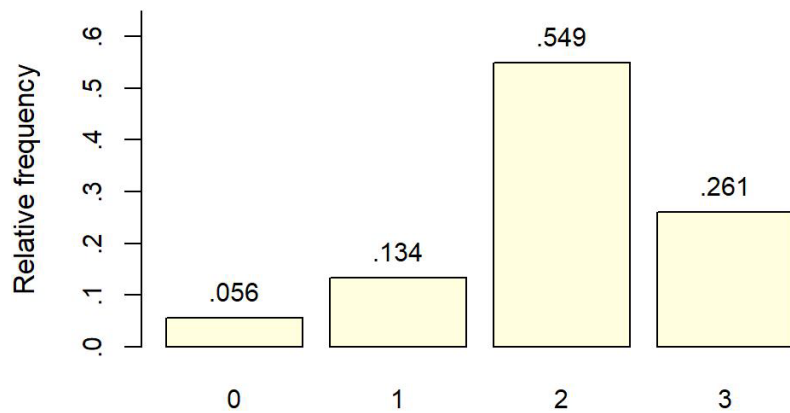


i10: The thought of harming myself has occurred to me

Figure 4. Bar diagram of the sample distribution of Item 10. Value labels: 0 = “Never”, 1 = “Hardly ever”, 2 = “Sometimes”. The fourth level does not appear: 3 = “Yes, quite often”.

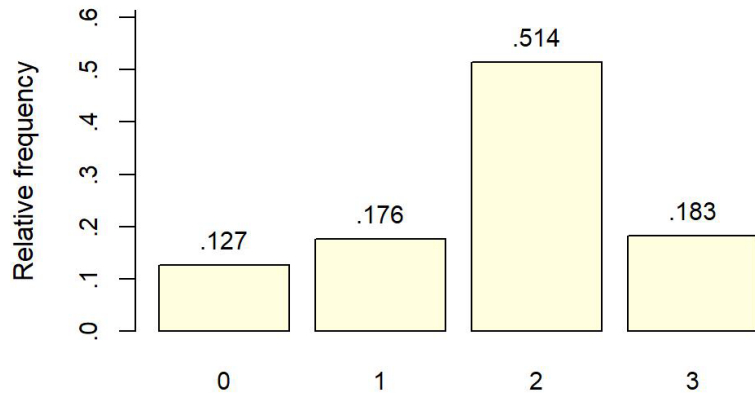
The median and mode values of Item 3 were both 2, as was its rounded arithmetic mean ($m = 2.014$). The variability was medium-low ($MAD = .507$, $SQR = .5$, and $QCV = .143$). The distribution was symmetric ($PCS = 0$) with mean tails ($PCK = -.013$) (Figure 5). Item 3 was similar to items 4 and 6 in terms of central tendency and variability. However, the latter two showed negative asymmetry and a long tail towards the left side ($PCS = -.333$), as well as slight pointing ($PCK = -.096$). See Figure 6 and Figure 7.

The central tendency of item 5 was concentrated between 1 and 2 ($mdn = 1$, $mo = 2$ and $m = 1.246$). It exhibited the highest variability among the ten items ($MAD = .810$, $SQR = 1$, and $QCV = .5$). Its profile was leptokurtic, that is, it had widened shoulders ($PCK = .237$), where the interquartile range coincided with the percentile range. Item 5 was the only symmetrical item ($QCS = 0$ and $PCS = 0$). See Figure 8.



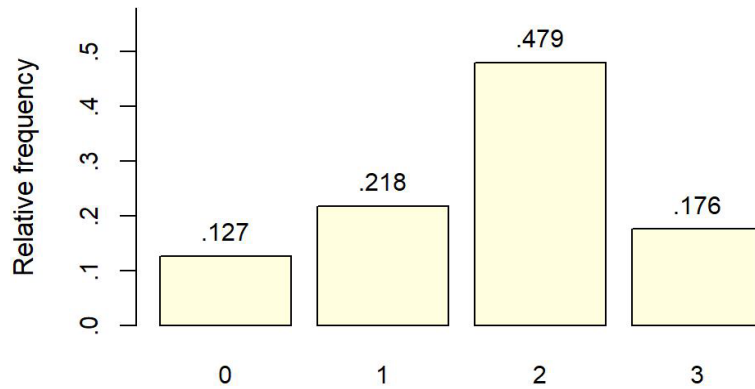
i3: I have blamed myself unnecessarily

Figure 5. Bar chart of the sample distribution of Item 3. Value labels: 0 = “No, never”, 1 = “Not very often”, 2 = “Yes, some of the time”, and 3 = “Yes, most of the time”.



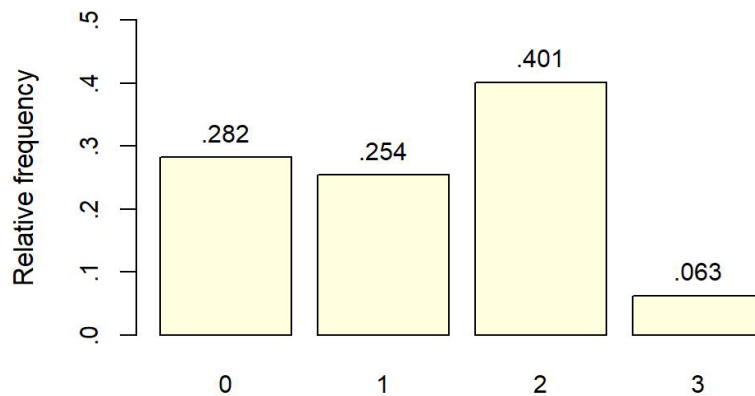
i4: I have been anxious or worried for no good reason

Figure 6. Bar diagram of the sample distribution of Item 4. Value labels: 0 = “Yes, very often”, 1 = “Yes, sometimes”, 2 = “Hardly ever”, and 3 = “No, not at all”.



i6: Things have been getting to me

Figure 7. Bar chart of the sample distribution of Item 6. Value labels: 0 = “No, I have been coping as well as ever”, 1 = “No, most of the time I have coped quite well”, 2 = “Yes, sometimes I haven’t been coping as well as usual”, and 3 = “Yes, most of the time I haven’t been able to cope at all”.



i5: I have felt scared or panicky for no good reason

Figure 8. Bar chart of the sample distribution of Item 5. Value labels: 0 = “No, not at all”, 1 = “No, not much”, 2 = “Yes, sometimes”, and 3 = “Yes, quite a lot”.

The central tendency of items 8 and 9 was 1, with medium-low variability. Item 8 was symmetrical ($PCS = 0$) and mesokurtic ($PCK = -.013$), whereas item 9 exhibited a long tail towards the left side ($PCS = -1$) and widened shoulders ($PCK = .237$) due to a shift in responses from category 2 to category 1. However, both items had a longer left tail than right tail, as indicated by the quartile coefficient of skewness ($QCS = -1$). Thus, their profiles were highly similar. See **Figures 9-10**.

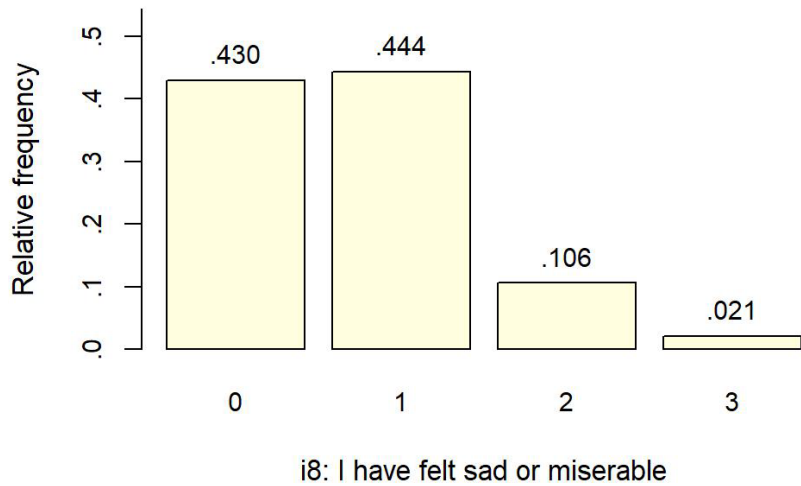


Figure 9. Bar chart of the sample distribution of Item 8. Value labels: 0 = “No, not at all”, 1 = “Not very often”, 2 = “Yes, quite often”, and 3 = “Yes, most of the time”.

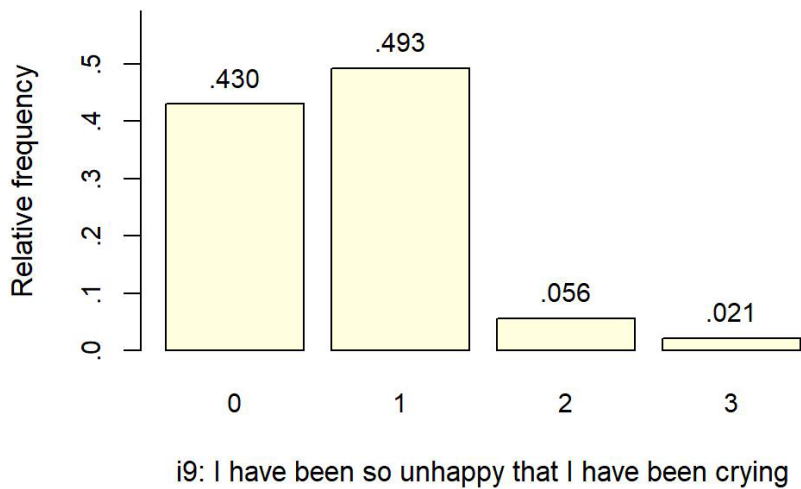


Figure 10. Bar chart of the sample distribution of Item 9. Value labels: 0 = “No, never”, 1 = “Only occasionally”, 2 = “Yes, quite often”, and 3 = “Yes, most of the time”.

3.3. Discriminating Power of Items

Assessed using the Mann-Whitney’s U test, the difference in central tendency was significant for all ten items ($p < .001$). The mean difference was greater than .75 (one-quarter of the item’s range) for eight items and greater than .5 (one-sixth of the item’s range) for items 2 and 10 (**Table 2**).

Table 2. Comparison of item central tendency between the groups with high and low EPDS total scores using the Mann-Whitney U test.

| Item | <i>n</i> _{LSG} | <i>n</i> _{HSG} | <i>md</i> [95% <i>CI</i>] | U_stat | exact_p_value | Z_stat | asympt_p_value |
|------|-------------------------|-------------------------|----------------------------|--------|---------------|--------|----------------|
| 1 | 41 | 49 | .874 [.652, 1.085] | 6.346 | < .001 | 6.346 | < .001 |
| 2 | 41 | 49 | .727 [.478, .960] | 5.624 | < .001 | 5.624 | < .001 |
| 3 | 41 | 49 | 1.108 [.796, 1.412] | 5.808 | < .001 | 5.808 | < .001 |
| 4 | 41 | 49 | 1.441 [1.126, 1.729] | 6.791 | < .001 | 6.791 | < .001 |
| 5 | 41 | 49 | 1.638 [1.378, 1.889] | 7.486 | < .001 | 7.486 | < .001 |
| 6 | 41 | 49 | 1.636 [1.365, 1.896] | 7.534 | < .001 | 7.534 | < .001 |
| 7 | 41 | 49 | 1.421 [1.205, 1.665] | 7.660 | < .001 | 7.660 | < .001 |
| 8 | 41 | 49 | 1.359 [1.176, 1.555] | 8.313 | < .001 | 8.313 | < .001 |
| 9 | 41 | 49 | 1.074 [.867, 1.295] | 7.402 | < .001 | 7.402 | < .001 |
| 10 | 41 | 49 | .633 [.416, .876] | 4.719 | < .001 | 4.719 | < .001 |

Note. *n*_{LSG} = size sample of the low score group: EPDS total score ≤ the 27th percentile of EPDS total score = 6 and *n*_{HSG} = size sample of the high score: EPDS total score ≥ the 73rd percentile of EPDS total score = 12, *md* = mean difference, U_stat = U statistic, exact_p_value = exact probability value, Z_stat = standardized U statistic, asympt_p_value = asymptotic probability value.

3.4. Internal Consistency of Items

3.4.1. Correlation between the Item and the Rest of the Scale

The correlation between each item and the rest of the scale was significantly greater than or equal to .5 in all cases (bootstrap p-values > .05), ranging from .592 to .833, with an average of .690, according to the polyserial correlation coefficient (Table 3). The null hypothesis of bivariate normality was not supported for items 1, 2, 3, and 6, based on the chi-square goodness-of-fit test (*p* < .05), making Spearman’s rank correlation coefficient a more appropriate choice for these items. For these four correlations, the null hypothesis of a correlation greater than or equal to .5 was also supported (bootstrap p-values > .05), with values ranging from .481 to .672 and an average of .572 (Table 3).

Table 3. Polyserial and Spearman correlation between the item with the rest of the scale.

| Items of EPDS | Chi-square test | | | | Correlation | | |
|---------------|-----------------|----|---------|------------------------|----------------|--------------|--------------|
| | χ^2 | df | p-value | <i>r</i> | Point estimate | (95% BCa CI) | boot p-value |
| 1 | 49.12 | 11 | < .001 | <i>r</i> _{PS} | .723 | [.586, .818] | .999 |
| | | | | <i>r</i> _S | .592 | [.481, .686] | .941 |
| 2 | 24.17 | 11 | .012 | <i>r</i> _{PS} | .624 | [.402, .746] | .998 |
| | | | | <i>r</i> _S | .481 | [.339, .591] | .402 |
| 3 | 9.48 | 11 | .578 | <i>r</i> _{PS} | .592 | [.458, .702] | .924 |
| | | | | <i>r</i> _S | .542 | [.392, .651] | .766 |
| 4 | 15.83 | 11 | .148 | <i>r</i> _{PS} | .659 | [.519, .750] | .996 |

Continued

| | | | | | | | |
|----|-------|----|------|----------|------|---------------|------|
| 5 | 8.67 | 11 | .653 | r_{ps} | .681 | [.516, .771] | .996 |
| 6 | 26.43 | 11 | .006 | r_{ps} | .720 | [.604, .793] | 1 |
| | | | | r_s | .672 | [.564, .751] | .998 |
| 7 | 18.73 | 11 | .066 | r_{ps} | .668 | [.523, .766] | .993 |
| 8 | 11.11 | 11 | .434 | r_{ps} | .833 | [.756, .885] | 1 |
| 9 | 10.79 | 11 | .461 | r_{ps} | .762 | [.745, .893] | 1 |
| 10 | 4.34 | 8 | .826 | r_{ps} | .634 | [.459, .7583] | .974 |

Note. Chi-square goodness-of-fit test: χ^2 = test statistic, df = degrees of freedom, and p -value = probability of obtaining a value equal to or greater than the observed statistic under the chi-square distribution with df degrees of freedom, for testing the null hypothesis of goodness of fit to the bivariate normality model. Correlations: r_{ps} = polyserial correlation; r_s = Spearman rank correlation coefficient; 95% BCa CI = 95% confidence interval based on the Bias-Corrected and accelerated (BCa) percentile method with 1000 bootstrap replications (except for item 1, for which 250 replications were used, and item 2, for which 900 replications were used in the r_{ps} estimation); boot p -value = bootstrap probability for the null hypothesis that $\rho \geq .5$, based on 1000 replications.

3.4.2. Internal Consistency of the Scale after Deleting the Item

Consistency can be assessed using the ordinal coefficient alpha (Zumbo et al., 2007) if the tau-equivalence assumption holds. If this assumption is not met, consistency can be evaluated using omega coefficients, either McDonald's (McDonald, 1999) or Green and Yang's (Green and Yang, 2009).

It is said that tau-equivalence exists when the weights of measurement of a general factor that determines the items are equivalent. Thus, a general factor model with ten indicators was defined with the restriction that the measurement weights are homogeneous (Figure 11). It was estimated by the WLSMV method that appears by default for ordinal variables in the R Lavaan package.

The iterative estimation process converged after the third iteration. The estimated homogeneous factor loading was .777, with 39.7% of the variance in each item explained by the factor. The factor variance was significant ($s^2 = .603$, $se = .033$, $z = 18.526$, $p < .001$). Although the Comparative Fit Index ($CFI = .971$) and the Tucker-Lewis Index ($TLI = .970$) exceeded the .95 threshold, the null hypothesis of tau-equivalence was rejected based on several fit criteria: the likelihood ratio chi-square test ($\chi^2[44] = 133.6$, $p < .001$), the standardized chi-square statistic ($\chi^2/df = 3.036 > 3$), the Root Mean Square Error of Approximation ($RMSEA = .120 > .08$, 95% CI [.097, .144], $p < .001$ for $H_0: RMSEA \leq .05$), and the Standardized Root Mean Squared Residual ($SRMR = .110 > .08$). Consequently, omega coefficients were used.

Internal consistency, as estimated by McDonald's omega coefficient, was very high ($\omega = .932$, 95% BCa CI [.908, .947]). The item-deleted omega coefficients were also high, ranging from .916 to .930, with an average of .925. In all ten cases, removing an item led to a decrease in overall internal consistency. The largest decrease (.017) occurred with item 8, and the smallest (.002) with item 3.

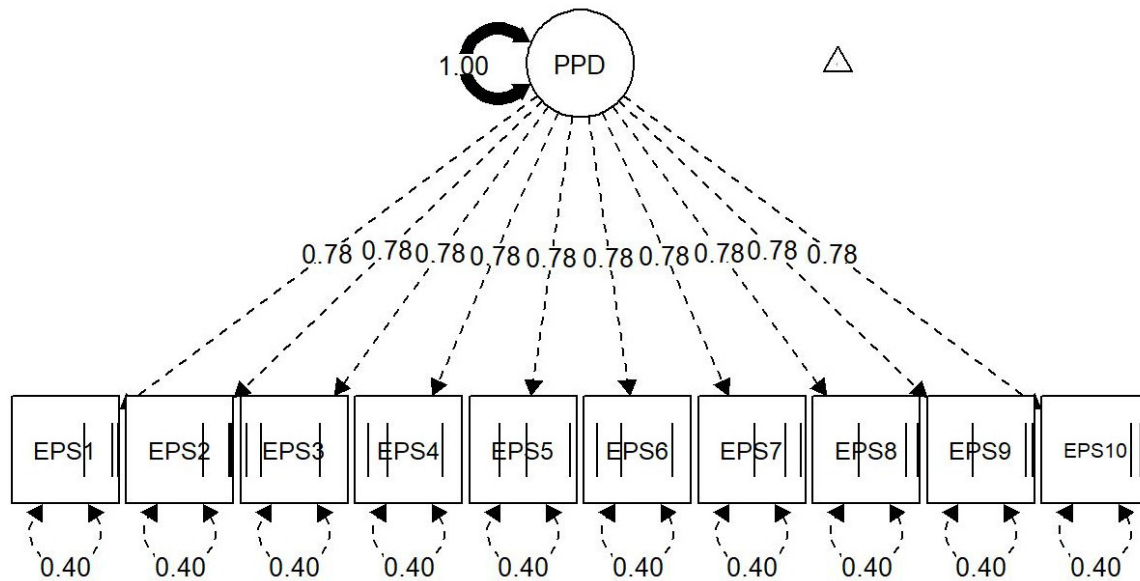


Figure 11. Standardized model for testing tau-equivalence, estimated using the WLSMV method.

However, these decreases were not statistically significant in two-tailed tests at the 5% significance level, as the 95% confidence intervals with and without the item overlapped (**Table 4**).

Table 4. Item-deleted internal consistency estimates using McDonald omega and Green-Yang categorical omega.

| Items | ω^M | 95% BCa CI | Diff. | Int. | ω^{GY} | 95% BCa CI | Diff. | Int. |
|--------------|------------|--------------|-------|------|---------------|--------------|-------|------|
| total | .932 | [.908, .947] | | | .912 | (.882, .926) | | |
| total - [1] | .923 | [.917, .952] | .009 | ↓ ns | .900 | (.883, .931) | .012 | ↓ ns |
| total - [2] | .928 | [.901, .943] | .004 | ↓ ns | .906 | (.851, .929) | .006 | ↓ ns |
| total - [3] | .930 | [.905, .946] | .002 | ↓ ns | .908 | (.856, .933) | .003 | ↓ ns |
| total - [4] | .926 | [.899, .943] | .006 | ↓ ns | .888 | (.831, .913) | .023 | ↓ ns |
| total - [5] | .926 | [.897, .942] | .007 | ↓ ns | .893 | (.837, .919) | .018 | ↓ ns |
| total - [6] | .924 | [.897, .942] | .008 | ↓ ns | .899 | (.842, .927) | .012 | ↓ ns |
| total - [7] | .927 | [.897, .942] | .006 | ↓ ns | .906 | (.853, .931) | .006 | ↓ ns |
| total - [8] | .916 | [.884, .935] | .017 | ↓ ns | .884 | (.824, .909) | .028 | ↓ ns |
| total - [9] | .920 | [.890, .938] | .012 | ↓ ns | .895 | (.840, .918) | .017 | ↓ ns |
| total - [10] | .928 | [.902, .946] | .004 | ↓ ns | .907 | (.857, .931) | .005 | ↓ ns |

Note. total - [j] = the sum of the 10 items excluding the item in square brackets (i = 1, 2, ..., 10). ω^M = McDonald's omega and ω^{GY} = Green-Yang categorical omega, with their point estimates and 95% bootstrap confidence intervals; the latter were calculated using the bias-corrected and accelerated percentile method based on 1,000 resamples with replacement (95% BCa CI). Diff. = difference between the internal consistency estimate with all 10 items and the estimate obtained after removing the item in square brackets. Int.: ↓ if $\omega_t > \omega_{t-i}$, ↑ if $\omega_t < \omega_{t-i}$; ns = non-significant change (confidence intervals for ω_t and ω_{t-i} overlap); s = significant change (confidence intervals do not overlap).

Overall internal consistency was also very high, as estimated by Green and Yang's omega coefficient ($\omega = .912$, 95% BCa CI [.882, .926]). In all ten cases, item removal led to a decrease in omega values, ranging from .003 to .028, with an average decrease of .013. However, these decreases were not statistically significant in two-tailed tests at the 5% significance level, as the 95% confidence intervals with and without the item overlapped (Table 4).

3.4.3. Commonalities of Items

The proportion of variance in each item, explained by the predictive model (which includes the other nine items), ranged from .423 (Item 3) to .786 (Item 8), with a mean of .566. Seven out of ten items (70%) had explained variances greater than .5. The predictive model was specified through path analysis and estimated using the WLSMV method (Table 5).

Table 5. Commonality or proportion of shared variance between each item and the other nine items.

| R-squared | Value |
|--|-------|
| R ² (Item 1 The other nine items) | .5699 |
| R ² (Item 2 The other nine items) | .4585 |
| R ² (Item 3 The other nine items) | .4225 |
| R ² (Item 4 The other nine items) | .5719 |
| R ² (Item 5 The other nine items) | .5732 |
| R ² (Item 6 The other nine items) | .5596 |
| R ² (Item 7 The other nine items) | .4876 |
| R ² (Item 8 The other nine items) | .7855 |
| R ² (Item 9 The other nine items) | .6873 |
| R ² (EPS10 Los otros ítems) | .5442 |

Note. R² = squared multiple correlation or proportion of variance explained by the model (item commonality). The model was calculated through path analysis with its parameters estimated using the Weighted Least Squares with Means and Variances adjusted (WLSMV) method.

3.5. Factor Structure with the Initial Items

3.5.1. Polychoric Correlations between 10 Items

The 45 correlations among the 10 EPDS items were estimated using polyserial correlation via the two-step method. In all 45 cases, the null hypothesis of bivariate normality was retained according to the chi-square test; therefore, the 95% confidence intervals were computed using the asymptotic approximation. Correlations ranged from .287 to .878, with a mean of .552. One correlation was below .3, 13 were in the range (.3, .5), 28 in the range (.5, .7], and 3 were above .7 (Table 6).

Table 6. Test of bivariate normality and point estimates with 95% asymptotic confidence intervals for the polychoric correlations among the 10 EPDS items.

| Ítems | $\chi^2[df, n=142]$ | <i>df</i> | p-value | <i>r_{pc}</i> | 95% <i>LL</i> | 95% <i>UL</i> | <i>ase</i> | | |
|-------|---------------------|-----------|---------|-----------------------|---------------|---------------|------------|-------|-------|
| 1 | 2 | 9.410 | 8 | .3089 | .6952 | .0697 | .5324 | .8084 | |
| | 3 | 5.574 | 8 | .6949 | .4674 | .0916 | .2701 | .6271 | |
| | 4 | 6.430 | 8 | .5992 | .5767 | .0774 | .4054 | .7089 | |
| | 5 | 15.020 | 8 | .0588 | .5270 | .0818 | .3488 | .6685 | |
| | 6 | 1.5420 | 8 | .9920 | .5712 | .0800 | .3942 | .7074 | |
| | 7 | 1.210 | 8 | .2508 | .5564 | .0812 | .3776 | .6952 | |
| | 8 | 7.381 | 8 | .4961 | .7003 | .0619 | .5580 | .8026 | |
| | 9 | 9.826 | 8 | .2775 | .6824 | .0673 | .5276 | .7934 | |
| | 10 | 4.981 | 5 | .4182 | .4195 | .1222 | .1552 | .6277 | |
| | 2 | 3 | 1.880 | 8 | .2085 | .2872 | .1077 | .0654 | .4820 |
| 4 | | 9.393 | 8 | .3102 | .4590 | .0939 | .2569 | .6226 | |
| 5 | | 14.570 | 8 | .0680 | .3816 | .0977 | .1759 | .5554 | |
| 6 | | 8.699 | 8 | .3683 | .5440 | .0864 | .3533 | .6913 | |
| 7 | | 6.882 | 8 | .5494 | .6033 | .0778 | .4290 | .7343 | |
| 8 | | 3.913 | 8 | .8648 | .5880 | .0823 | .4037 | .7264 | |
| 9 | | 6.208 | 8 | .6239 | .5572 | .0892 | .3586 | .7075 | |
| 10 | | 6.086 | 5 | .2979 | .4222 | .1260 | .1486 | .6358 | |
| 3 | | 4 | 2.364 | 8 | .9678 | .5506 | .0699 | .3992 | .6728 |
| | | 5 | 1.424 | 8 | .9939 | .4811 | .0778 | .3149 | .6187 |
| | 6 | 7.573 | 8 | .4762 | .5887 | .0645 | .4479 | .7010 | |
| | 7 | 8.144 | 8 | .4195 | .3690 | .0926 | .1754 | .5352 | |
| | 8 | 7.235 | 8 | .5115 | .5581 | .0739 | .3967 | .6861 | |
| | 9 | 3.070 | 8 | .9299 | .5124 | .0817 | .3354 | .6544 | |
| | 10 | 7.600 | 5 | .1797 | .4999 | .1101 | .2556 | .6842 | |
| | 4 | 5 | 12.570 | 8 | .1273 | .6996 | .0516 | .5839 | .7875 |
| | | 6 | 11.430 | 8 | .1783 | .6022 | .0619 | .4669 | .7080 |
| | | 7 | 15.760 | 8 | .0460 | .4263 | .0844 | .2479 | .5768 |
| 8 | | 5.220 | 8 | .7338 | .4905 | .0776 | .3242 | .6274 | |
| 9 | | 3.778 | 8 | .8766 | .4725 | .0830 | .2948 | .6186 | |
| 10 | | 5.230 | 5 | .3885 | .3471 | .1250 | .0835 | .5653 | |
| 5 | 6 | 4.959 | 8 | .7620 | .5553 | .0676 | .4091 | .6737 | |
| | 7 | 8.021 | 8 | .4314 | .4980 | .0771 | .3325 | .6337 | |
| | 8 | 8.175 | 8 | .4165 | .6035 | .0674 | .4549 | .7193 | |
| | 9 | 5.110 | 8 | .7458 | .5993 | .0720 | .4397 | .7222 | |
| | 10 | 5.144 | 5 | .3986 | .5207 | .0990 | .3014 | .6878 | |

Continued

| | | | | | | | | |
|---|----|--------|---|-------|-------|--------|-------|-------|
| 6 | 7 | 11.530 | 8 | .1734 | .5314 | .07630 | .3659 | .6643 |
| | 8 | 7.376 | 8 | .4966 | .6924 | .0550 | .5687 | .7854 |
| | 9 | 2.585 | 8 | .9577 | .5381 | .0758 | .3735 | .6699 |
| | 10 | 2.024 | 5 | .9577 | .5381 | .0758 | .3735 | .6699 |
| 7 | 8 | 13.320 | 8 | .1015 | .7127 | .0517 | .5958 | .8000 |
| | 9 | 1.410 | 8 | .2375 | .6141 | .0687 | .4616 | .7313 |
| | 10 | 3.376 | 5 | .6423 | .6240 | .0881 | .4208 | .7676 |
| 8 | 9 | 7.491 | 8 | .4847 | .8781 | .0358 | .7859 | .9320 |
| | 10 | 6.437 | 5 | .266 | .6915 | .0809 | .4982 | .8194 |
| 9 | 10 | 4.137 | 5 | .5299 | .6210 | .0942 | .4021 | .7728 |

Note. $\chi^2[df, n = 142]$ = test statistic value, df = degree of freedom, p -value = probability value calculated towards right tail of a chi-square distribution, r_{pc} = point estimate of the polychoric correlation, 95% LL = lower limit and UL = upper limit of the 95% asymptotic confidence interval, and ase = asymptotic standard error.

3.5.2. Measurement Weights and Factor Model Fit

The one-factor model with 10 indicators was estimated using the WLSMV method, based on the previously reported polychoric correlations (Table 6). The iterative process converged after 38 iterations. All measurement weights were statistically significant, ranging from .641 to .934, with a mean of .758. The factor variance was significant ($s^2 = .615, se = .071, z = 8.609, p < .001$). See Figure 12. Bootstrap 95% confidence intervals for the standardized measurement weights (factor loadings) were calculated using the percentile method based on 1000 bootstrap samples (Table 7). The BCa method could not be applied to these data.

The fit of the one-factor model with 10 indicators was adequate according to the likelihood ratio chi-square test ($\chi^2[35] = 54.252, p = .020$), the Root Mean Square Error of Approximation ($RMSEA = .062 < .08, 95\% CI [.025, .094], p = 249$ for $H_0: RMSEA \leq .05$), and the Standardized Root Mean squared Residual ($SRMR = .076 < .08$). Model fit was good according to the remaining indices: standardized likelihood ratio chi-square statistic ($\chi^2/df = 54.252/35 = 1.550 < 2$), Comparative Fit Index ($CFI = .994 > .95$), and Tucker-Lewis Index ($TLI = .992 > .95$).

The average variance extracted (AVE) exceeded the recommended threshold of .5 ($AVE = .582$), and McDonald’s omega coefficient was above .7 ($\omega = .932, 95\% BCa CI [.908, .947]$). These results indicate that the measurement model demonstrated convergent validity (Cheung et al., 2024).

Table 7. Measurement weights in a single-factor model with ten indicators.

| Indicators | Method | $\hat{\lambda}$ | se | z | p-value | $\hat{\lambda}$ [95% CI] |
|-------------------------------|--------|-----------------|------|-------|---------|--------------------------|
| $\lambda(F \rightarrow EPS1)$ | asint | 1 | | | | .784 |
| | boot | | .113 | | | [.658, .890] |
| $\lambda(F \rightarrow EPS2)$ | asint | .882 | .090 | 9.784 | <.001 | .692 |

Continued

| | | | | | | | |
|---------------------------------------|-------|-------|-----|--------|-------|--|--------------|
| | boot | | 122 | | | | [.495, .840] |
| $\lambda(F \rightarrow \text{EPS3})$ | asint | 817 | 077 | 10.609 | <.001 | | .641 |
| | boot | | 103 | | | | [.523, .766] |
| $\lambda(F \rightarrow \text{EPS4})$ | asint | 923 | 075 | 12.256 | <.001 | | .724 |
| | boot | | 115 | | | | [.593, .839] |
| $\lambda(F \rightarrow \text{EPS5})$ | asint | 949 | 076 | 12.542 | <.001 | | .745 |
| | boot | | 111 | | | | [.624, .848] |
| $\lambda(F \rightarrow \text{EPS6})$ | asint | 971 | 078 | 12.431 | <.001 | | .761 |
| | boot | | 110 | | | | [.658, .854] |
| $\lambda(F \rightarrow \text{EPS7})$ | asint | 923 | 083 | 11.141 | <.001 | | .724 |
| | boot | | 113 | | | | [.579, .843] |
| $\lambda(F \rightarrow \text{EPS8})$ | asint | 1.191 | 076 | 15.681 | <.001 | | .934 |
| | boot | | 122 | | | | [.864, .986] |
| $\lambda(F \rightarrow \text{EPS9})$ | asint | 1.110 | 075 | 14.878 | <.001 | | .871 |
| | boot | | 118 | | | | [.776, .942] |
| $\lambda(F \rightarrow \text{EPS10})$ | asint | 902 | 119 | 7.581 | <.001 | | .707 |
| | boot | | 121 | | | | [.563, .862] |
| $\sigma^2(F)$ | asint | 615 | 071 | 8.609 | <.001 | | 1 |

Note. Parameters: $\lambda(F \rightarrow \text{EPS}_i)$ = measurement weight of the indicator EPS_i ($i = 1, 2, \dots, 10$) and $\sigma^2 =$ variance. F = depression factor with ten indicators. Estimates: $\hat{\lambda}$ = unstandardized point estimate, se = standard error, z-value = test statistic, p-value = $2 \times (1 - P(Z \leq |z|))$ = two-tailed probability for the null hypothesis: $\Lambda = 0$, $\hat{\lambda}$ [95% CI] = standardized measurement weight, including its point estimate and its bootstrap confidence interval obtained using the percentile method.).

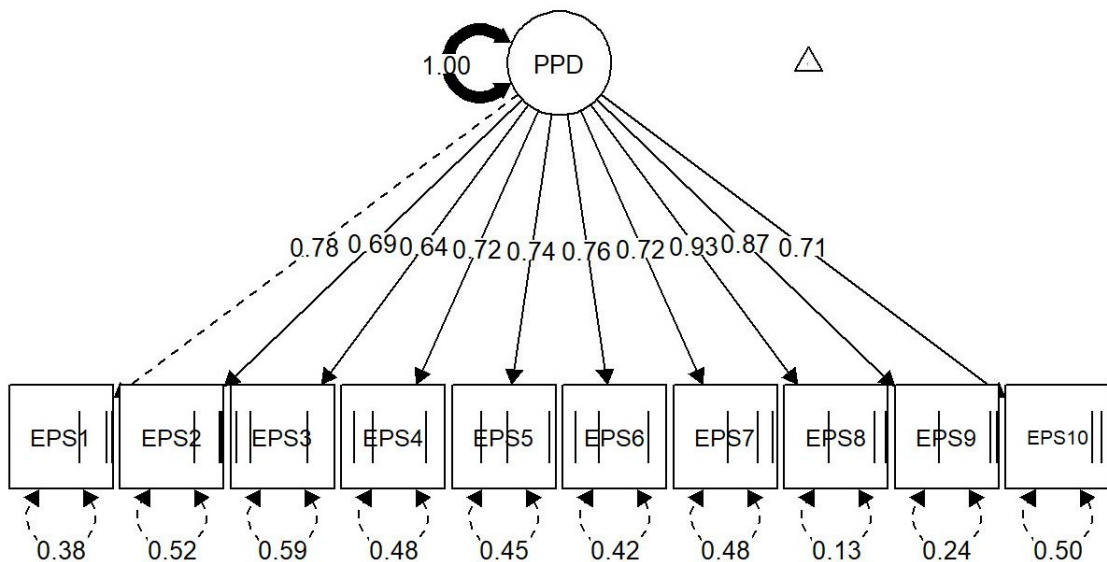


Figure 12. Standardized one-factor model with 10 indicators, estimated using the WLSMV method.

3.6. Decision on Item Deletion

No item warranted elimination based on distributional characteristics that deviated markedly from those of the other items, pronounced ceiling or floor effects, poor discriminating power, lack of internal consistency, or low measurement weight. The ten items were discriminative, internally consistent, and contributed significantly to the one-factor model, which demonstrated an acceptable-to-good fit and evidence of convergent validity ($AVE > .5$ and $\omega > .7$). Given the good fit of the one-factor model with ten indicators, the hierarchical three-factor model proposed by Alfayumi-Zeadna et al. (2022) was not examined.

3.7. Description of the EPDS Total Score Distribution

In this sample of 142 women, the mean of the scale was 9,782 (95% BCa CI [8,917, 10,642]), the median was 10 (95% BCa CI [9, 10]), and the mode, estimated using Grenader's (Grenader, 1960) method, was 8,995 (95% BCa CI [6,841, 10,590]). The mode based on the highest frequency sample value was 10; this was unique, with a frequency of .12. Thus, the measures of central tendency were very similar.

The sample minimum value was 0, the maximum was 28, and the sample standard deviation was 5.31. The coefficient of variation was 54.3% (95% BCa CI [48.52, 61.87]), which exceeds the expected value for a normal distribution. For this type of data, values of around one quarter or less are usually observed (Panichkitkosolkul, 2015).

Grubbs' (1969) outlier test identified the sample maximum of 28 as an outlier ($G = 3.431$, $U = .916$, $p = .033$, $\alpha = .05$). Thus, the distribution presented positive asymmetry with a long tail towards the right side ($\sqrt{b_1} = .431$, 95% BCa CI [.127, .994]) as tested by D'Agostino (1970) test ($z[\sqrt{b_1}] = 2.115$, $p = .034$), indicating non-normality. However, the null hypothesis of mesokurtosis ($H_0: \beta_2 = 3$) was maintained ($b_2 = 3.026$, 95% BCa CI [2.279, 4.822]) according to Anscombe & Glynn (1983) test ($z[b_2] = .345$, $p = .730$), which is one of the characteristics of the normal distribution (Figure 13).

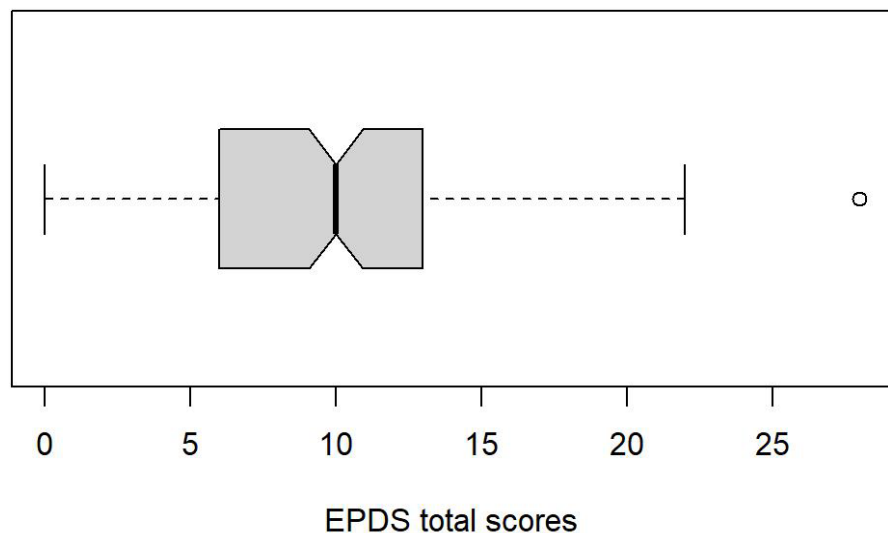


Figure 13. Box plot of the EPDS total score.

The null hypothesis of normality was rejected by the Shapiro & Wilk’s (Shapiro & Wilk, 1965) test with the Royston’s (Royston, 1992) standardization: $W = .976, p = .014 < \alpha = .05$. This result was confirmed by the Anderson and Darling’s (Anderson and Darling, 1952) test: $A^2 = .825, p = .032 < \alpha = .05$, and of Kolmogorov-Smirnov’s test with the probability calculated via Lilliefors’ (Lilliefors, 1967) Monte Carlo simulation: $D = .096, p = .003 < \alpha = .05$. The histogram, with the normal density curve overlaid, showed that the peak of the sample distribution is higher, the left tail was shorter, and the right tail was heavier than the peak and the tails corresponding to the normal density curve (Figure 14). Since the distribution of the EPDS total score deviated from normal, the scale should be scaled by percentile scores rather than by standard deviation units from the mean. Table 8 shows the sample deciles estimated using R’s type-8 rule.

Table 8. Deciles of the EPDS total score.

| D1 | D2 | D3 | D4 | D5 | D6 | D7 | D8 | D9 |
|----|----|----|-------|----|----|----|----|----|
| 3 | 5 | 7 | 8.267 | 10 | 10 | 12 | 14 | 18 |

Note. The sample deciles were estimated using the type-8 rule in R.

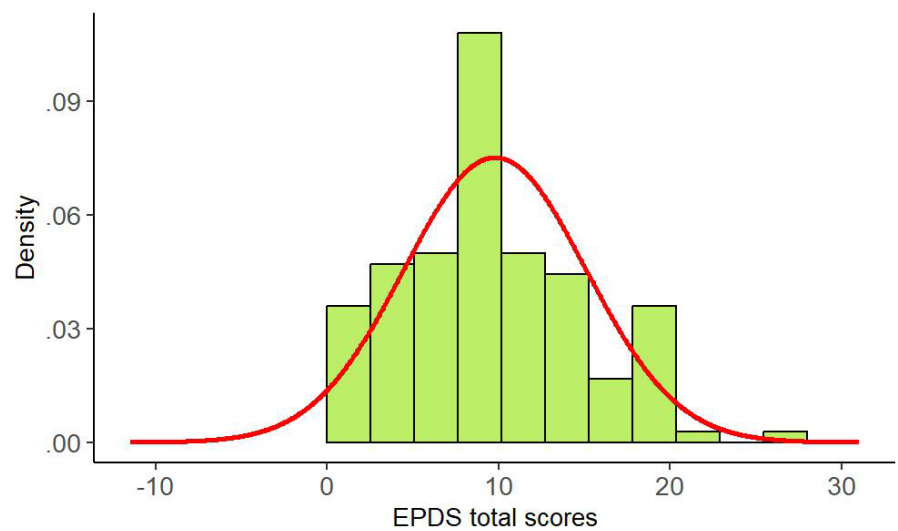


Figure 14. Histogram of the EPDS total score with an overlaid normal density curve.

3.8. Concurrent Construct Validity

3.8.1. Perceived Social Support

In the present sample, a correlated two-factor model was specified for the 44 items of PSSS, and its parameters and fit were estimated using the WLSMV method. The first factor, support from partner (items 1 to 24), showed an average variance extracted (AVE) above .50 (AVE = .712) and high composite reliability (McDoanld’s $\omega = .983$). The second factor, support from significant others (items 25 to 44), exhibited an AVE close to .50 (AVE = .586) and similarly high composite reliability (McDoanld’s $\omega = .966$). The correlation between the two factors was low ($r(F1,$

F2) = .281, 95% BCa CI [.065, .409]), yet statistically significant ($z = 4.297, p < .001$), indicating a shared variance of 7.9%. The overall composite reliability of the scale was very high (McDoanld’s $\omega = .988$).

Model fit was good according to several indices: $\chi^2/df = 1535.969/901 = 1.705 < 2$; $GFI = .980, AGFI = .978, CFI = .954, TLI$ (or $NNFI$) = .952, Bollen’s $IFI = .954$, and the Relative Noncentrality Index (RNI) = .954, all exceeding the .95 threshold. Fit was considered adequate based on Bentler-Bonett’s $NFI = .896$ and Bollen’s $RFI = .890$, both above .85, and $RMSEA = .071$, with a 90% confidence interval of [.065, .077], below the .08 cutoff. However, model fit was not supported by the likelihood ratio chi-square test ($\chi^2[901] = 1535.969, p < .001$), and the Standardized Root Mean Square Residual ($SRMR = .103$) exceeded the recommended threshold of .08, indicating poor fit on this index.

The distribution of EPDS total scores deviated from normality according to the Shapiro-Wilk test ($W = .979, z = 1.878, p = .030 < .05$), as did the distribution of the couple support factor scores ($W = .851, z = 6.328, p < .001$). However, the null hypothesis of normality was not rejected for the significant other support factor scores ($W = .982, z = 1.507, p = .066 > .05$).

Correlations between EPDS total scores and PSSS total scores, as well as their two factor scores, were calculated using Pearson’s product-moment coefficient with 95% BCa confidence intervals. Statistical significance was assessed using two-sided bootstrap probabilities. EPDS total scores were negatively correlated with PSSS total scores and the partner support factor scores, both showing strong associations, and with the significant other support factor scores, showing a moderate association (Table 9).

Table 9. Correlations between EPDS total score and PSSS total score and its two factors.

| Scales | <i>r</i> | [95% BCa CI] | bias | <i>se</i> | <i>p</i> _{boot} |
|----------------------------------|----------|----------------|--------|-----------|--------------------------|
| PSSS total score | -.659 | [-.750, -.539] | .002 | .055 | < .01 |
| Partner support factor | -.588 | [-.714, -.435] | .004 | .067 | < .01 |
| Significant other support factor | -.407 | [-.555, -.245] | < .001 | .079 | < .01 |

Note. PSSS = Perceived Social Support Scale, *r* = Pearson’s product-moment correlation, 95% BCa CI = 95% bootstrap confidence interval using the bias-corrected and accelerated percentile method, bias = difference between bootstrap estimate (mean of the 1000 estimates from the 1000 bootstrap samples) and original sample estimate, *se* = bootstrap standard error, *p*_{boot} = two-sided bootstrap probability for the null hypothesis of zero correlation.

3.8.2. Depression During Pregnancy

As expected, the mean of the group with depression during pregnancy ($m_1 = 15.5$) was higher than the mean of the group without depression during pregnancy ($m_2 = 9.348$) with a mean difference of 6.152, 95% BCa CI [3.962, 8.825].

When comparing the central tendency using the one-tailed Mann-Whitney’s U test, the difference was statistically significant: $U_1 = 1119.5, U_2 = 200.5, U = 200.5, z(U) = -3.672, p$ (left-tailed asymptotic probability) = .00012 < $\alpha = .05, p$ (left-

tailed bootstrap probability) = .0001 < α = .05. The median EPDS total score of the 10 women with depression during pregnancy was 15 and that of the 132 women without depression during pregnancy was 9.

The effect size was medium, according to both Rosenthal's (Rosenthal, 1991) correlation ($r = |z|/\sqrt{n} = .308 > .3$) and Spearman's (Spearman, 1904) rank correlation ($r_s = -.309 < -.3$). Likewise, Kerby's (Kerby, 2014) d , equivalent to Cureton's (Cureton, 1968) rank-biserial correlation, also indicated a medium effect size: $d = r_{rb} = .696$; $CI \in [.691, .788]$. A d value of .696 suggests that in 69.6% of cases, the values in group 1 are higher than those in group 2. In other words, there is a .696 probability that a randomly selected participant from the group with depression during pregnancy (group 1) will score higher on the EPDS than a randomly selected participant from the group without depression during pregnancy (group 2). This value is comparable to the Area Under the ROC Curve (AUC).

3.8.3. Marital Status

When comparing the central tendency in the EPDS total scores among the four marital status groups using the Kruskal-Wallis's test, significant differences were found at a 5% significance level: $H = 8.698$, $p = .034$ (Table 10). Pairwise comparisons using the Dunn's test with Bonferroni's correction showed that single women without partners had a significantly higher mean rank than married and cohabiting women (Table 10 and Table 11). As expected, women without partners experience the highest levels of depression. The size of the effect of marital status on depression level was medium according to the epsilon squared ($.06 < \epsilon^2 = H / (n - 1) = 8.698 / (142 - 1) = .062 < .14$), but small according to the eta squared ($.01 < \eta^2 = (H - k + 1) / (n - k) = (8.698 - 3 + 1) / (142 - 3) = .048 < .06$).

Due to the small number of single women without partner ($n = 4$) and single women with partners ($n = 6$), and the absence of divorced or separated women, this analysis should be interpreted with caution and considered exploratory.

Table 10. Comparison of central tendency in EPDS total scores among the four marital status groups using the Kruskal-Wallis test.

| Marital status | n | m | [95% BCa CI] | MR | H | df | p |
|------------------------|-----|-------|---------------|--------|-------|------|------|
| Married | 110 | 9.39 | [8.50, 10.23] | 69.40 | 8.698 | 3 | .034 |
| Single without partner | 4 | 9.75 | [6, 14] | 73.00 | | | |
| Single with partner | 6 | 16.83 | [12, 20] | 119.75 | | | |
| Cohabiting | 22 | 9.82 | [7.39, 12.90] | 68.59 | | | |
| Total | 142 | 9.78 | [8.97, 10.62] | | | | |
| Married | 110 | 9.39 | [8.50, 10.23] | 69.40 | 8.698 | 3 | .034 |

Note. n = subsample size, m [95% BCa CI] = arithmetic mean with a 95% BCa confidence interval, MR = mean rank, H = value of the test statistic, df = degrees of freedom, $p = P(\chi^2[3] \geq H)$ = probability at the right tail of a chi-square distribution with three degrees of freedom.

Table 11. Pairwise comparisons of mean ranks in EPDS total score across marital status categories using Dunn's test with Bonferroni correction.

| Difference | <i>d</i> | <i>se</i> | <i>z</i> | <i>p</i> _{2 colas} | <i>p</i> _{aj} |
|------------|----------|-----------|----------|-----------------------------|------------------------|
| RM1 - RM2 | -3.605 | 20.890 | -.173 | .863 | 1 |
| RM1 - RM3 | -50.355 | 17.206 | -2.927 | .003 | .021 |
| RM1 - RM4 | .805 | 9.585 | .084 | .933 | 1 |
| RM2 - RM3 | -46.750 | 26.492 | -1.765 | .078 | .466 |
| RM2 - RM4 | 4.409 | 22.308 | .198 | .843 | 1 |
| RM3 - RM4 | 51.159 | 18.902 | 2.707 | .007 | .041 |

Note. *RM* = mean rank of the group: 1 = married women, 2 = single with a partner, 3 = single without a partner and 4 = cohabiting, *d* = difference between mean ranks, *se* = standard error of the difference, $z = d/se$ = standardized value of the difference, $p = 2 \times (1 - P(Z \leq |z|))$ = two-tailed probability value in a standard normal distribution, $p_{aj} = \min(p \times 6, 1)$, where 6 represents the number of tested differences or probability value with the Bonferroni adjustment for type I error (false positives) in multiple comparisons.

4. Discussion

This EPDS validation study, conducted with an incidental sample of 142 Mexican mothers, examines psychometric properties that have not been previously addressed in the literature. The first objective was to analyze the psychometric characteristics of the 10 items comprising the scale. Most items exhibit a descending response pattern, with higher frequencies at 0 (items 1, 2, 7, and 10) or 1 (items 3, 4, and 6), as expected for a scale measuring a psychopathological trait (DeYoung et al., 2022). However, items 5, 8, and 9 display the opposite pattern, with a mode at 3 and a left-skewed distribution, despite being negatively keyed like most items (eight out of ten).

These three items do not define a distinct factor, even when two factors are extracted using an exploratory factor analysis based on Cattell's criterion for determining the number of factors. Factor 1, which includes items 1, 2, 7, 8, 9, and 10, has an eigenvalue of 4.322, while Factor 2, consisting of items 3, 4, 5, and 6, has an eigenvalue of .671. Specifically, factors were extracted using the principal axis factoring method, and the factor matrix was rotated using Promax rotation, although similar results were obtained using other factoring methods. This suggests that these three items reflect common experiences among women during their baby's first year.

As expected, no item exhibits a ceiling effect. However, item 10 shows a clear floor effect, with over 80% of responses concentrated at the lowest value. Items 1, 2, and 7 also display a floor effect when applying the three-quarters distribution criterion. A one-quarter criterion would be inappropriate for these items, as it is more suitable when the total score is expected to follow a normal distribution and the number of response options is at least 5, preferably 7 (Moral, 2006).

An analysis of central tendency measures shows that values of 2 and 3 indicate a high level of depressive symptoms across most items, except for item 3, in which

only a value of 3 reflects a high level. Notably, a response of 1 already suggests elevated symptomatology in items 1, 2, 7, 8, 9, and 10—particularly in items 1, 2, 9, and 10. A value of 0 corresponds to a low level of depression in all items. Item 5 stands out for its distinct distribution, which most closely resembles a uniform pattern.

The items demonstrate discriminating power, as all significantly differentiate between groups with high and low total scores on the scale. In seven of the ten items (items 3 – 7), the mean difference exceeds 1. Item 10, which shows a floor effect, exhibits the weakest discriminative capacity, with a mean difference of less than .63.

The 10 items demonstrate reliability according to three criteria: 1) a significant correlation between an item and the sum of the remaining items, equal to or greater than .5; 2) a decrease in overall internal consistency when the item is removed; and 3) communality, defined as the proportion of variance in the item explained by the linear combination of the other items, equal to or greater than one-quarter. Item 10 has the lowest item-total correlation and the lowest communality. It is also the first item whose removal leads to the smallest decrease in internal consistency when assessed using Green-Yang categorical omega and the second when assessed using McDonald's omega coefficient. Although this item—characterized by a floor effect—is the weakest in terms of both discriminating power and reliability, it meets the minimum criteria and is therefore retained as an indicator of the one-factor model.

The second objective of the study was to test the one-factor model, which is implicitly assumed but had never been empirically verified. With 10 indicators, all factor loadings are significant and exceed .60, including item 10, which emerges as a strong indicator with a loading of .71 and over 50% of its variance explained. The model's fit ranges from acceptable to good and demonstrates convergent validity, with an average variance extracted (*AVE*) above .5 and composite reliability (McDonald's and Green-Yang's omega coefficients) greater than .7 (Cheung et al., 2024). Removing any item does not improve model fit. Thus, all ten items are definitively retained.

Regarding the third objective, overall internal consistency is very high ($\omega > .90$), whether estimated using McDonald's omega coefficient or Green-Yang's categorical omega. These coefficients are appropriate given the non-homogeneous factor loadings. Their values are not excessively high ($< .95$), which would otherwise suggest item redundancy. Moreover, no item has a loading equal to one, which could lead to negative residual variances. The chi-square statistic for model fit exceeds 1; a lower value could indicate an overfitted model. Taken together, these results suggest that the items function as strong, non-redundant indicators, each contributing distinct content.

The overall internal consistency exceeds that reported in previous studies, which have remained below .90, typically ranging from .80 to .87 based on Cronbach's alpha coefficient (Gibson et al., 2009). This discrepancy is largely attributable to the use of omega (ω) coefficients to estimate internal consistency

(Amirrudin et al., 2021). These indices are more appropriate when the assumption of tau-equivalence is violated and when the variables are ordinal (Orçan, 2023)—particularly when the factor model is estimated using a polychoric correlation matrix and methods such as Weighted Least Squares Mean and Variance adjusted (WLSMV) or Diagonally Weighted Least Squares (DWLS). In the present sample, internal consistency estimated using Cronbach's alpha is .874. However, this coefficient assumes tau-equivalence and is better suited to continuous variables than to ordinal ones (Shaw, 2021). When these assumptions are not met, alpha tends to underestimate internal consistency.

Regarding the fourth objective—describing the distribution of the EPDS total score—a deviation from normality is observed, characterized by positive skewness. The relative variability is twice as high as expected under normality (Panichkitkosolkul, 2015), and an outlier is detected in the right tail, as anticipated given the psychopathological nature of the construct being measured (Simms et al., 2022). Consequently, the scale should be standardized using percentile scores (Chien & Yao, 2024). The 70th percentile corresponds to a score of 12, which is the suggested cutoff for use in Mexico (Alvarado-Esquivel et al., 2006; Macías-Cortés et al., 2020). However, considering that the maximum expected prevalence of postpartum depression is approximately 20% (Secretaría de Salud, 2016; Wang et al., 2021a, 2021b), it would be advisable to raise the cutoff to at least 14, corresponding to the 80th percentile in this incidental sample of 142 Mexican mothers who had given birth at least two months prior to being surveyed.

Regarding the fifth objective—assessing the convergent validity of the EPDS—the results are consistent with expectations. As in previous studies (Cho et al., 2022; Hajipoor et al., 2021), the EPDS shows a negative correlation with perceived social support, with a strong association between the total score and partner support, and a moderate association with support from a significant other. These findings reinforce the role of social support as a key protective factor (Agrawal et al., 2022).

As expected, depression during pregnancy is a risk factor for postpartum depression (Agrawal et al., 2022; Flores-Ramos et al., 2025; Quesada & Chinchilla, 2022), with a medium effect size based on rank-biserial correlation analysis. This association may be influenced by both personality-based diathesis and genetic predisposition, as well as by depressive environmental factors that persist beyond pregnancy, such as poor partner relationships, lack of social support, social exclusion, or financial difficulties (Agrawal et al., 2022; Quesada & Chinchilla, 2022).

According to the hypothesis, being a single mother is a risk factor for postpartum depression, with a medium effect size based on the epsilon-squared coefficient. However, studies often report no significant association between EPDS total scores and marital status (Kahveci et al., 2021; Oliveira et al., 2022). This may be due to the common practice of dichotomizing this qualitative variable (e.g., married vs. unmarried, or with a partner vs. without a partner), which fails to distinguish between single women with and without a partner—a distinction that is

particularly relevant in the present study. Single mothers without a partner face greater social and economic challenges and receive less emotional support than those who are partnered. When marital status is treated as a multi-level variable, [Moya et al. \(2023\)](#) found that being single, compared to the reference category (divorced/widowed), functioned as a protective factor (OR = .09, 95% CI [.02, .55], $p = .009$). In another unexpected finding, [Riesco-González et al. \(2022\)](#) reported that being divorced was a protective factor compared to being single or married. These conflicting results—even relative to theoretical expectations—highlight the need for a more nuanced analysis of this variable. Notably, the present sample included no divorced or separated women.

The study has several limitations. The sample was obtained incidentally via social media, and the scale was administered online. Although the participant-to-item ratio (142:10 \approx 14:1) and the participant-to-parameter ratio in the factor model (142:20 \approx 7:1) meet conventional guidelines, the overall sample size remains somewhat limited. Minors were excluded due to the challenges of obtaining parental consent through social media and ensuring their participation in the online questionnaire. However, this sociodemographic group may have included a higher proportion of single mothers without a partner. In addition, mothers within the first four weeks postpartum were excluded to avoid false positives for postpartum depression caused by postpartum blues syndrome ([Manurung & Setyowati, 2021](#)). As a result, the study focused on a more specific subgroup of postpartum depression ([Jadresic, 2017](#)).

The measurement instrument used in this study consists of Likert-type items, which are ordinal in nature. While this limitation does not affect most studies that rely on the dichotomized EPDS total score (case/non-case), it is addressed here by incorporating recent advances and best practices in psychometric research ([Ferrando & Morales-Vives, 2023](#); [Jebb et al., 2021](#); [Kyriazos & Poga-Kyriazou, 2023](#)). The analytical approach is grounded in Classical Test Theory ([Gorham & Randall, 2022](#)), although alternative methodologies, such as Item Response Theory, could also be applied ([Bock & Gibbons, 2021](#)). Classical Test Theory is generally more appropriate for assessing personality traits and emotional states, whereas Item Response Theory is better suited for evaluating abilities—particularly in contexts involving competitive selection ([Lang & Tay, 2021](#)).

Due to the difficulty of recontacting participants, the present study did not assess the temporal reliability of the EPDS, which represents a limitation. However, other studies have addressed this psychometric property. For example, [Kernot et al. \(2015\)](#) reported a high level of test–retest reliability for total scores ($ICC = .92$) in a sample of 118 Australian mothers over an interval of approximately three days. [Alfayumi-Zeadna et al. \(2022\)](#) found a correlation of .40 between scores during pregnancy and postpartum among 332 Bedouin women over a longer time interval (36 – 48 weeks of gestation and 2 – 4 months postpartum). These Israelite authors tested the invariance of a model consisting of a general depression factor and three hierarchical factors — anhedonia (items 1 – 2), anxiety (items 3 – 6),

and sadness (items 7 - 10) — across two paired samples. Their results were favorable in terms of model fit and invariance. Song et al. (2024) examined the invariance of this hierarchical three-factor model in a sample of 1,207 Chinese women assessed at two different time points and found support for metric, scalar, and strict invariance. It is worth noting that this alternative model was not evaluated in the present study due to the good fit of the original single-factor model, which offers greater parsimony.

5. Conclusions

The 10 items comprising the EPDS are suitable for retention. As the scale assesses a psychopathological trait, item distributions exhibit asymmetry. Item 10 presents a distinct floor effect and shows the lowest discriminating power and internal consistency compared to the other items, which perform strongly on these metrics. Nonetheless, item 10 still meets the minimum criteria for discriminability and reliability and functions as a valid indicator within the one-factor model.

The measurement model demonstrates an acceptable to good fit and shows evidence of convergent validity ($AVE > .5$ and $\omega > .80$). The scale's internal consistency is very high. Given the significant deviation from normality due to positive skewness, the total score should be standardized using percentile ranks. In this sample, a score of at least 14 (80th percentile) appears appropriate for identifying cases of postpartum depression, whereas traditional cut-off scores of 10 or 12 may be too low.

The scale also demonstrates concurrent construct validity: depression during pregnancy and being single without a partner emerged as risk factors with medium effect sizes, while perceived social support showed a strong negative association with EPDS scores.

6. Recommendations for Future Research

It is recommended that future studies replicate this psychometric research using a larger, randomly selected sample, including underage mothers and those in the early postpartum period. Administering the EPDS—designed to assess postpartum depression—alongside the Maternal Blues Scale (Manurung & Setyowati, 2021) could help determine whether these two syndromes are truly distinct. Based on the findings of this study, one could hypothesize that single or separated/divorced mothers without a partner are more likely to experience postpartum depression compared to single, separated/divorced, married, or cohabiting women with a partner. Marital status should not be treated as a simplified (i.e., dichotomized) variable; rather, it should be nuanced by considering the presence or absence of a partner as a polychotomous qualitative variable.

It would be valuable to conduct a study estimating the test-retest correlation and examining the invariance or temporal stability of the one-factor model with ten indicators in a random sample of paired data collected approximately three to four weeks postpartum. Similarly, factorial invariance between rural and urban

populations could be assessed. In addition, potential differences in central tendency on the scale could be explored. If such differences exist, they would justify the development of separate interpretative norms (percentile scores) for each population.

Funding

The study was funded by the authors' own resources.

Acknowledgements

We thank the women who participated in the study for kindly and selflessly responding to the questionnaire.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

- Agrawal, I., Mehendale, A. M., & Malhotra, R. (2022). Risk Factors of Postpartum Depression. *Cureus*, *14*, e30898. <https://doi.org/10.7759/cureus.30898>
- Alfayumi-Zeadna, S., O'Rourke, N., Azbarga, Z., Froimovici, M., & Daoud, N. (2022). Temporal Stability of Responses to the Edinburgh Postpartum Depression Scale by Bedouin Mothers in Southern Israel. *International Journal of Environmental Research and Public Health*, *19*, Article 13959. <https://doi.org/10.3390/ijerph192113959>
- Alvarado-Esquivel, C., Sifuentes-Alvarez, A., Salas-Martinez, C., & Martínez-García, S. (2006). Validation of the Edinburgh Postpartum Depression Scale in a Population of Puerperal Women in Mexico. *Clinical Practice and Epidemiology in Mental Health*, *2*, Article No. 33. <https://doi.org/10.1186/1745-0179-2-33>
- American Psychiatric Association (APA) (2013). *Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (DSM-5)*. American Psychiatric Publishing, Inc.
- American Psychological Association (APA) (2017). *Ethical Principles of Psychologists and Code of Conduct*. <https://www.apa.org/ethics/code/index.aspx>
- Amirrudin, M., Nasution, K., & Supahar, S. (2021). Effect of Variability on Cronbach Alpha Reliability in Research Practice. *Jurnal Matematika, Statistika dan Komputasi*, *17*, 223-230. <https://doi.org/10.20956/jmsk.v17i2.11655>
- Anderson, T. W., & Darling, D. A. (1952). Asymptotic Theory of Certain "Goodness of Fit" Criteria Based on Stochastic Processes. *The Annals of Mathematical Statistics*, *23*, 193-212. <https://doi.org/10.1214/aoms/1177729437>
- Anscombe, F. J., & Glynn, W. J. (1983). Distribution of the Kurtosis Statistic B_2 for Normal Samples. *Biometrika*, *70*, 227-234. <https://doi.org/10.1093/biomet/70.1.227>
- Bai, Y., Li, Q., Cheng, K. K., Caine, E. D., Tong, Y., Wu, X. et al. (2023). Prevalence of Postpartum Depression Based on Diagnostic Interviews: A Systematic Review and Meta-analysis. *Depression and Anxiety*, *2023*, Article ID: 8403222. <https://doi.org/10.1155/2023/8403222>
- Beck, C. T., & Gable, R. K. (2000). Postpartum Depression Screening Scale: Development and Psychometric Testing. *Nursing Research*, *49*, 272-282. <https://doi.org/10.1097/00006199-200009000-00006>

- Beck, C. T., & Gable, R. K. (2003). Postpartum Depression Screening Scale: Spanish Version. *Nursing Research*, *52*, 296-306. <https://doi.org/10.1097/00006199-200309000-00004>
- Bock, R. D., & Gibbons, R. D. (2021). *Item Response Theory*. Wiley. <https://doi.org/10.1002/9781119716723>
- Bonferroni, C. E. (1936). *Teoria Statistica delle Classi e Calcolo delle Probabilità [Statistical Class Theory and Probability Calculation]*. Reale Istituto Superiore di Scienze Economiche e Commerciali di Firenze.
- Bowley, A. L. (1901). *Elements of Statistics*. P. S. King.
- Byrne, B. M. (2016). *Structural Equation Modelling with AMOS Basic Concepts, Applications, and Programming* (3rd ed.). Routledge. <https://doi.org/10.4324/9781315757421>
- Cámara de Diputados del H. Congreso de la Unión (2014). *Reglamento de la Ley General de Salud en Materia de Investigación para la Salud*. Diario Oficial de la Federación. https://www.dof.gob.mx/nota_detalle.php?codigo=5339162&fecha=02/04/2014#gsc.tab=0
- Centers for Disease Control and Prevention (CDC) (2008). Prevalence of Self-Reported Postpartum Depressive Symptoms-17 States, 2004-2005, MMWR. *Morbidity and Mortality Weekly Report*, *57*, 361-366. <https://www.cdc.gov/mmwr/index.html>
- Cheung, G. W., Cooper-Thomas, H. D., Lau, R. S., & Wang, L. C. (2024). Reporting Reliability, Convergent and Discriminant Validity with Structural Equation Modeling: A Review and Best-Practice Recommendations. *Asia Pacific Journal of Management*, *41*, 745-783. <https://doi.org/10.1007/s10490-023-09871-y>
- Chien, C., & Yao, G. (2024). Norms, Psychometrics. In F. Maggino (Ed.), *Encyclopedia of Quality of Life and Well-Being Research* (pp. 4723-4725). Springer International Publishing. https://doi.org/10.1007/978-3-031-17299-1_1965
- Cho, H., Lee, K., Choi, E., Cho, H. N., Park, B., Suh, M. et al. (2022). Association between Social Support and Postpartum Depression. *Scientific Reports*, *12*, Article No. 3128. <https://doi.org/10.1038/s41598-022-07248-7>
- Cienfuegos-Martínez, Y. I. (2010). *Violencia en la relación de pareja: Una aproximación desde el modelo ecológico*. Master's Thesis, Universidad Nacional Autónoma de México. <https://repositorio.unam.mx/contenidos/88645>
- Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2nd ed.). Lawrence Erlbaum Associates.
- Conover, W. J. (1999). *Practical Nonparametric Statistics* (3th ed.). John Wiley & Sons.
- Cox, J. L., Holden, J. M., & Sagovsky, R. (1987). Detection of Postnatal Depression: Development of the 10-Item Edinburgh Postnatal Depression Scale. *British Journal of Psychiatry*, *150*, 782-786. <https://doi.org/10.1192/bjp.150.6.782>
- Cureton, E. E. (1968). Rank-biserial Correlation When Ties Are Present. *Educational and Psychological Measurement*, *28*, 77-79. <https://doi.org/10.1177/001316446802800107>
- D'Agostino, R. B. (1970). Transformation to Normality of the Null Distribution of G_1 . *Biometrika*, *57*, 679-681. <https://doi.org/10.1093/biomet/57.3.679>
- D'Agostino, R. B., Belanger, A., & D'agostino, R. B. (1990). A Suggestion for Using Powerful and Informative Tests of Normality. *The American Statistician*, *44*, 316-321. <https://doi.org/10.1080/00031305.1990.10475751>
- de Castro, F., Marie Place, J., Allen-Leigh, B., Rivera-Rivera, L., & Billings, D. (2016). Provider Report of the Existence of Detection and Care of Perinatal Depression: Quantitative Evidence from Public Obstetric Units in Mexico. *Salud Pública de México*, *58*, 468-

471. <https://doi.org/10.21149/spm.v58i4.8028>
- DeYoung, C. G., Chmielewski, M., Clark, L. A., Condon, D. M., Kotov, R., Krueger, R. F. et al. (2022). The Distinction between Symptoms and Traits in the Hierarchical Taxonomy of Psychopathology (HITOP). *Journal of Personality*, *90*, 20-33. <https://doi.org/10.1111/jopy.12593>
- Dunn, O. J. (1964). Multiple Comparisons Using Rank Sums. *Technometrics*, *6*, 241-252. <https://doi.org/10.2307/1266041>
- Efron, B., & Narasimhan, B. (2020). The Automatic Construction of Bootstrap Confidence Intervals. *Journal of Computational and Graphical Statistics*, *29*, 608-619. <https://doi.org/10.1080/10618600.2020.1714633>
- Ferrando, P. J., & Morales-Vives, F. (2023). Is It Quality, Is It Redundancy, or Is Model Inadequacy? Some Strategies for Judging the Appropriateness of High-Discrimination Items. *Anales de Psicología*, *39*, 517-527. <https://doi.org/10.6018/analesps.535781>
- Fisher, R. A. (1930). The Moments of the Distribution for Normal Samples of Measures of Departure from Normality. *Proceedings of the Royal Society of London, Series A: Mathematical, Physical, and Engineering Sciences*, *130*, 16-28.
- Flores-Ramos, G., Castro-Apodaca, F. J., Garay-Vizcarra, L. A., Favela-Heredia, C. E., Acosta-Alfaro, L. F., Canizalez-Román, A., Magaña-Ordorica, D., Terán-Cabanillas, E., León-Sicarios, N. M., Peña-García, G. M., Sandoval-Quíñonez, P. A., & Murillo-Llanes, J. (2025). Depresión Perinatal en Mujeres Mexicanas Diagnosticadas mediante la Escala de Depresión Posnatal de Edimburgo. *Ginecología y Obstetricia de México*, *93*, 6-12. <https://doi.org/10.24245/gom.v93i1.114>
- Furr, R. M. (2021). *Psychometrics: An Introduction*. SAGE Publications.
- García-Esteve, L., Ascaso, C., Ojuel, J., & Navarro, P. (2003). Validation of the Edinburgh Postnatal Depression Scale (EPDS) in Spanish Mothers. *Journal of Affective Disorders*, *75*, 71-76. [https://doi.org/10.1016/s0165-0327\(02\)00020-4](https://doi.org/10.1016/s0165-0327(02)00020-4)
- Gibson, J., McKenzie-McHarg, K., Shakespeare, J., Price, J., & Gray, R. (2009). A Systematic Review of Studies Validating the Edinburgh Postnatal Depression Scale in Antepartum and Postpartum Women. *Acta Psychiatrica Scandinavica*, *119*, 350-364. <https://doi.org/10.1111/j.1600-0447.2009.01363.x>
- Gorham, A., & Randall, J. (2022). *Classical Test Theory*. Routledge. <https://doi.org/10.4324/9781138609877-REE26-1>
- Green, S. B., & Yang, Y. (2009). Reliability of Summed Item Scores Using Structural Equation Modeling: An Alternative to Coefficient Alpha. *Psychometrika*, *74*, 155-167. <https://doi.org/10.1007/s11336-008-9099-3>
- Grenander, U. (1965). Some Direct Estimates of the Mode. *The Annals of Mathematical Statistics*, *36*, 131-138. <https://doi.org/10.1214/aoms/1177700277>
- Grubbs, F. E. (1969). Procedures for Detecting Outlying Observations in Samples. *Technometrics*, *11*, 1-21. <https://doi.org/10.2307/1266761>
- Hajipoor, S., Pakseresht, S., Niknami, M., Atrkar Roshan, Z., & Nikandish, S. (2021). The Relationship between Social Support and Postpartum Depression. *Journal of Holistic Nursing and Midwifery*, *31*, 93-103. <https://doi.org/10.32598/jhnm.31.2.1099>
- Han, H. (2022). The Effectiveness of Weighted Least Squares Means and Variance Adjusted Based Fit Indices in Assessing Local Dependence of the Rasch Model: Comparison with Principal Component Analysis of Residuals. *PLOS ONE*, *17*, e0271992. <https://doi.org/10.1371/journal.pone.0271992>
- Izah, S. C., Sylva, L., & Hait, M. (2023). Cronbach's Alpha: A Cornerstone in Ensuring

- Reliability and Validity in Environmental Health Assessment. *ES Energy & Environment*, 23, Article 1057. <https://doi.org/10.30919/eseel1057>
- Jadresic M., E. (2017). Depresión posparto en el contexto del hospital general. *Revista Médica Clínica Las Condes*, 28, 874-880. <https://doi.org/10.1016/j.rmcl.2017.10.007>
- Jebb, A. T., Ng, V., & Tay, L. (2021). A Review of Key Likert Scale Development Advances: 1995-2019. *Frontiers in Psychology*, 12, Article 637547. <https://doi.org/10.3389/fpsyg.2021.637547>
- Jiménez-Brito, L. (2022). Maternidad y Seguridad Social en México. *Las Notas Técnicas*, 2, 1-57. <https://ciss-bienestar.org/2022/09/06/nota-tecnica-18/>
- Kahveci, G., Kahveci, B., Aslanhan, H., & Bucaktepe, P. G. E. (2021). Evaluation of Prevalence and Risk Factors for Postpartum Depression Using the Edinburgh Postpartum Depression Scale: A Cross-Sectional Analytic Study. *Gynecology Obstetrics & Reproductive Medicine*, 27, 227-233. <https://doi.org/10.21613/gorm.2020.1109>
- Kelley, K. (2023). *The MBESS R Package*. <https://cran.r-project.org/web/packages/MBESS/MBESS.pdf>
- Kelley, T. L. (1923). *Statistical Methods*. The Macmillan Company.
- Kelley, T. L. (1939). The Selection of Upper and Lower Groups for the Validation of Test Items. *Journal of Educational Psychology*, 30, 17-24. <https://doi.org/10.1037/h0057123>
- Kerby, D. S. (2014). The Simple Difference Formula: An Approach to Teaching Nonparametric Correlation. *Comprehensive Psychology*, 3, Article 1. <https://doi.org/10.2466/11.it.3.1>
- Kernot, J., Olds, T., Lewis, L. K., & Maher, C. (2015). Test-Retest Reliability of the English Version of the Edinburgh Postnatal Depression Scale. *Archives of Women's Mental Health*, 18, 255-257. <https://doi.org/10.1007/s00737-014-0461-4>
- Kline, R. B. (2016). *Principles and Practice of Structural Equation Modeling*. The Guilford Press.
- Kolmogorov, A. N. (1933). Sulla Determinazione Empirica di una Legge di Distribuzione [On the Empirical Determination of a Law of Distribution]. *Giornale dell'Istituto Italiano degli Attuari*, 4, 83-91.
- Kruskal, W. H., & Wallis, W. A. (1952). Use of Ranks in One-Criterion Variance Analysis. *Journal of the American Statistical Association*, 47, 583-621. <https://doi.org/10.2307/2280779>
- Kyriazos, T., & Poga-Kyriazou, M. (2023). Applied Psychometrics: Estimator Considerations in Commonly Encountered Conditions in CFA, SEM, and EFA Practice. *Psychology*, 14, 799-828. <https://doi.org/10.4236/psych.2023.145043>
- Lang, J. W. B., & Tay, L. (2021). The Science and Practice of Item Response Theory in Organizations. *Annual Review of Organizational Psychology and Organizational Behavior*, 8, 311-338. <https://doi.org/10.1146/annurev-orgpsych-012420-061705>
- Lara, M. A., Navarrete, L., Navarro, C., & Le, H. (2013). Evaluation of the Psychometric Measures for the Postpartum Depression Screening Scale-Spanish Version for Mexican Women. *Journal of Transcultural Nursing*, 24, 378-386. <https://doi.org/10.1177/1043659613493436>
- Li, C. (2021). Statistical Estimation of Structural Equation Models with a Mixture of Continuous and Categorical Observed Variables. *Behavior Research Methods*, 53, 2191-2213. <https://doi.org/10.3758/s13428-021-01547-z>
- Lilliefors, H. W. (1967). On the Kolmogorov-Smirnov Test for Normality with Mean and Variance Unknown. *Journal of the American Statistical Association*, 62, 399-402.

<https://doi.org/10.1080/01621459.1967.10482916>

- Lyubenova, A., Neupane, D., Levis, B., Wu, Y., Sun, Y., He, C. et al. (2021). Depression Prevalence Based on the Edinburgh Postnatal Depression Scale Compared to Structured Clinical Interview for DSM Disorders Classification: Systematic Review and Individual Participant Data Meta-analysis. *International Journal of Methods in Psychiatric Research*, 30, e1860. <https://doi.org/10.1002/mpr.1860>
- Macías-Cortés, E. d. C., Lima-Gómez, V., & Asbun-Bojalil, J. (2020). Exactitud diagnóstica de la Escala de Depresión Posnatal de Edimburgo: Consecuencias del tamizaje en mujeres mexicanas. *Gaceta Médica de México*, 156, 202-208. <https://doi.org/10.24875/gmm.19005424>
- Malpartida Ampudia, M. K. (2020). Depresión postparto en atención primaria. *Revista Medica Sinergia*, 5, e355. <https://doi.org/10.31434/rms.v5i2.355>
- Mann, H. B., & Whitney, D. R. (1947). On a Test of Whether One of Two Random Variables Is Stochastically Larger than the Other. *The Annals of Mathematical Statistics*, 18, 50-60. <https://doi.org/10.1214/aoms/1177730491>
- Manurung, S., & Setyowati, S. (2021). Development and Validation of the Maternal Blues Scale through Bonding Attachments in Predicting Postpartum Blues. *Malaysian Family Physician*, 16, 64-74. <https://doi.org/10.51866/oa1037>
- McDonald, R. P. (1999). *Test Theory: A Unified Treatment*. Psychology Press. <https://doi.org/10.4324/9781410601087>
- McGraw, K. O., & Wong, S. P. (1992). A Common Language Effect Size Statistic. *Psychological Bulletin*, 111, 361-365. <https://doi.org/10.1037/0033-2909.111.2.361>
- McNeish, D., & Wolf, M. G. (2023). Dynamic Fit Index Cutoffs for Confirmatory Factor Analysis Models. *Psychological Methods*, 28, 61-88. <https://doi.org/10.1037/met0000425>
- Moral, J. (2006). Análisis Factorial y su Aplicación al Desarrollo de Escalas. In R. Landero, & M. T. González (Eds.), *Estadística con SPSS y Metodología de la Investigación* (pp. 387-443). Trillas.
- Moya, E., Mzembe, G., Mwaminga, M., Truwah, Z., Harding, R., Ataide, R. et al. (2023). Prevalence of Early Postpartum Depression and Associated Risk Factors among Selected Women in Southern Malawi: A Nested Observational Study. *BMC Pregnancy and Childbirth*, 23, Article No. 229. <https://doi.org/10.1186/s12884-023-05501-z>
- Oliveira, T. A., Luzetti, G. G. C. M., Rosalém, M. M. A., & Mariani Neto, C. (2022). Screening of Perinatal Depression Using the Edinburgh Postpartum Depression Scale. *Revista Brasileira de Ginecologia e Obstetrícia/RBGO Gynecology and Obstetrics*, 44, 452-457. <https://doi.org/10.1055/s-0042-1743095>
- ORÇAN, F. (2023). Comparison of Cronbach's Alpha and McDonald's Omega for Ordinal Data: Are They Different? *International Journal of Assessment Tools in Education*, 10, 709-722. <https://doi.org/10.21449/ijate.1271693>
- Organización Panamericana de la Salud (2016). *El Protocolo de Vigilancia Epidemiológica de la Mortalidad Materna*. <https://iris.paho.org/handle/10665.2/33712>
- Panichkitkosolkul, W. (2015). Improved Confidence Intervals for a Coefficient of Variation of a Normal Distribution. *Thailand Statistician*, 7, 193-199. <https://ph02.tci-thaijo.org/index.php/thaistat/article/view/34315>
- Quesada, M. V., & Chinchilla, K. V. (2022). Detección temprana de la depresión posparto. *Revista Ciencia y Salud Integrando Conocimientos*, 6, 37-44. <https://doi.org/10.34192/cienciaysalud.v6i5.474>
- Reed, G. M., Maré, K. T., First, M. B., Jaisoorya, T. S., Rao, G. N., Dawson-Squibb, J. et al.

- (2024). The Who Flexible Interview for ICD-11 (FLII-11). *World Psychiatry*, 23, 359-360. <https://doi.org/10.1002/wps.21227>
- Riesco-González, F. J., Antúnez-Calvente, I., Vázquez-Lara, J. M., Rodríguez-Díaz, L., Palomo-Gómez, R., Gómez-Salgado, J. et al. (2022). Body Image Dissatisfaction as a Risk Factor for Postpartum Depression. *Medicina*, 58, Article 752. <https://doi.org/10.3390/medicina58060752>
- Rosenthal, R. (1991). *Meta-Analytic Procedures for Social Research*. SAGE Publications, Inc. <https://doi.org/10.4135/9781412984997>
- Rosseel, Y. (2012). Lavaan: An R Package for Structural Equation Modeling. *Journal of Statistical Software*, 48, 1-36. <https://doi.org/10.18637/jss.v048.i02>
- Rosseel, Y., Jorgensen, T. D., & De Wilde, L. (2024). *Lavaan: Latent Variable Analysis. Version 0.6-19*. <https://doi.org/10.32614/CRAN.package.lavaan>
- Royston, P. (1992). Approximating the Shapiro-Wilk W-Test for Non-Normality. *Statistics and Computing*, 2, 117-119. <https://doi.org/10.1007/bf01891203>
- Saharoy, R., Potdukhe, A., Wanjari, M., & Taksande, A. B. (2023). Postpartum Depression and Maternal Care: Exploring the Complex Effects on Mothers and Infants. *Cureus*, 15, e41381. <https://doi.org/10.7759/cureus.41381>
- Santiago Sanabria L., Ibarra-Gussi P. M., Rendón-Macías M. E., Treviño-Villarreal P., Islas-Tezpa David., Porras-Ibarra G. D., & Van Tienhoven (2023). Depresión Posparto: Prevalencia y Factores de Riesgo Asociados en una Muestra de Población Mexicana. *Ginecología y Obstetricia de México*, 91, 227-240. <https://doi.org/10.24245/gom.v91i4.8456>
- Santiago Sanabria, L., Islas Tezpa, D., & Flores Ramos, M. (2022). Trastornos del estado de ánimo en el postparto. *Acta Médica Grupo Angeles*, 20, 173-177. <https://doi.org/10.35366/104280>
- Secretaría de Salud (2016). *Sufre Depresión Posparto, Una de Cada 10 Mexicanas*. <https://www.gob.mx/salud/prensa/sufre-depresion-posparto-una-de-cada-10-mexicanas-61679#:~:text=Una%20de%20cada%2010%20mujeres,%2C%20doctor%20Francisco%20Morales%20Carmona>
- Secretaría de Salud (2023). 122. *En México, Dos de Cada 10 Mujeres Presentan Depresión Durante el Embarazo o Después del Parto*. <https://www.gob.mx/salud/prensa/122-en-mexico-dos-de-cada-10-mujeres-presentan-depresion-durante-el-embarazo-o-despues-del-parto?idiom=es>
- Shafian, A. K., Mohamed, S., Nasution Raduan, N. J., & Hway Ann, A. Y. (2022). A Systematic Review and Meta-Analysis of Studies Validating Edinburgh Postnatal Depression Scale in Fathers. *Heliyon*, 8, e09441. <https://doi.org/10.1016/j.heliyon.2022.e09441>
- Shapiro, S. S., & Wilk, M. B. (1965). An Analysis of Variance Test for Normality (Complete Samples). *Biometrika*, 52, 591-611. <https://doi.org/10.2307/2333709>
- Shaw, B. P. (2021). Meeting Assumptions in the Estimation of Reliability. *The Stata Journal: Promoting communications on statistics and Stata*, 21, 1021-1027. <https://doi.org/10.1177/1536867x211063407>
- Simms, L. J., Wright, A. G. C., Cicero, D., Kotov, R., Mullins-Sweatt, S. N., Sellbom, M. et al. (2022). Development of Measures for the Hierarchical Taxonomy of Psychopathology (HITOP): A Collaborative Scale Development Project. *Assessment*, 29, 3-16. <https://doi.org/10.1177/10731911211015309>
- Smirnov, N. (1948). Table for Estimating the Goodness of Fit of Empirical Distributions. *The Annals of Mathematical Statistics*, 19, 279-281.

<https://doi.org/10.1214/aoms/1177730256>

- Sociedad Mexicana de Psicología (2010). *Código Ético del Psicólogo* (5th ed.). Editorial Trillas.
- Song, Z., Zhang, D., Yang, L., Zhu, P., Liu, Y., Wang, S. et al. (2024). Factor Structure and Longitudinal Invariance for the Chinese Mainland Version of the Edinburgh Postnatal Depression Scale during Pregnancy. *Midwifery*, *132*, Article ID: 103963. <https://doi.org/10.1016/j.midw.2024.103963>
- Spearman, C. (1904). The Proof and Measurement of Association between Two Things. *The American Journal of Psychology*, *15*, 72-101. <https://doi.org/10.2307/1412159>
- Wang, Z., Liu, J., Shuai, H., Cai, Z., Fu, X., Liu, Y. et al. (2021a). Mapping Global Prevalence of Depression among Postpartum Women. *Translational Psychiatry*, *11*, Article No. 543. <https://doi.org/10.1038/s41398-021-01663-6>
- Wang, Z., Liu, J., Shuai, H., Cai, Z., Fu, X., Liu, Y. et al. (2021b). Correction: Mapping Global Prevalence of Depression among Postpartum Women. *Translational Psychiatry*, *11*, Article No. 640. <https://doi.org/10.1038/s41398-021-01692-1>
- World Health Organization (2024). *CIE-11. International Classification of Diseases and Related Health Problems, 11th Revision. The Global Standard for Diagnostic Health Information*. World Health Organization. <https://icd.who.int/en/>
- Yang, M., Song, B., Jiang, Y., Lin, Y., & Liu, J. (2023). Mindfulness-Based Interventions for Postpartum Depression: A Systematic Review and Meta-Analysis. *Iranian Journal of Public Health*, *52*, 2496-2505. <https://doi.org/10.18502/ijph.v52i12.14311>
- Yates, F. (1934). Contingency Tables Involving Small Numbers and the X² Test. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, *1*, 217-235. <https://doi.org/10.2307/2983604>
- Zheng, S., & Cao, Y. (2022). Correlation Analysis for Different Types of Variables and Relationship between Different Correlation Coefficients. *Biometrics & Biostatistics International Journal*, *11*, 127-129. <https://doi.org/10.15406/bbij.2022.11.00365>
- Zumbo, B. D., Gadermann, A. M., & Zeisser, C. (2007). Ordinal Versions of Coefficients Alpha and Theta for Likert Rating Scales. *Journal of Modern Applied Statistical Methods*, *6*, 21-29. <https://doi.org/10.22237/jmasm/1177992180>

Appendix. Edinburgh Postpartum Depression Scale (EPDS)

Queremos saber cómo se siente si ha tenido un bebé recientemente. Por favor marque la respuesta que más se acerque a cómo se ha sentido en LOS ÚLTIMOS 7 DÍAS, no solamente cómo se sienta hoy.

We want to know how you have been feeling if you have recently had a baby. Please select the answer that best reflects how you have felt in THE LAST 7 DAYS, not just how you feel today.

1. He sido capaz de reír y ver el lado bueno de las cosas

- Tanto como siempre
- No tanto ahora
- Mucho menos
- No, no he podido

2. He mirado el futuro con placer

- Tanto como siempre
- No tanto ahora
- Mucho menos
- No, no he podido

3. Me he culpado sin necesidad cuando las cosas no salían bien

- Sí, la mayoría de las veces
- Sí, algunas veces
- No muy a menudo
- No, nunca

4. He estado ansiosa y preocupada sin motivo

- No, para nada
- Casi nada
- Sí, a veces
- Sí, a menudo

5. He sentido miedo y pánico sin motivo alguno

- Sí, bastante
- Sí, a veces
- No, no mucho
- No, nada

6. Las cosas me oprimen o agobian

- Sí, la mayor parte de las veces
- Sí, a veces
- No, casi nunca
- No, nada

7. Me he sentido tan infeliz que he tenido dificultad para dormir

- Sí, la mayoría de las veces
- Sí, a veces
- No muy a menudo
- No, nada

8. Me he sentido triste y desgraciada

- Sí, casi siempre
- Sí, bastante a menudo
- No muy a menudo
- No, nada

9. He sido tan infeliz que he estado llorando

- Sí, casi siempre
- Sí, bastante a menudo
- Sólo en ocasiones
- No, nunca

10. He pensado en hacerme daño a mí misma

- Sí, bastante a menudo
- A veces
- Casi nunca
- No, nunca