The Spanish Journal of Psychology (2021), **24**, e47, 1–9.

doi:10.1017/SJP.2021.44

CAMBRIDGE UNIVERSITY PRESS

External Validation and Test-Retest Reliability of Postpartum Bonding Questionnaire in Spanish Mothers

Anna Torres-Giménez[1], Alba Roca-Lecumberri[1], Bàrbara Sureda[1], Susana Andrés-Perpiña[1], Bruma Palacios-Hernández[2], Estel Gelabert[3], Borja Farré-Sender[4], Susana Subirà-Álvarez[3] and Lluïsa García-Esteve[1]

[1] *Hospital Clinic of Barcelona, IDIBAPS (Spain)*
[2] *Universidad Autónoma del Estado de Morelos (Mexico)*
[3] *Universitat Autònoma de Barcelona (Spain)*
[4] *Hospital Universitari Dexeus - Grupo Quirónsalud (Spain)*

Abstract. The aim of the present study was to validate the Spanish Postpartum Bonding Questionnaire (PBQ) against external criteria of bonding disorder, as well as to establish its test-retest reliability. One hundred fifty-six postpartum women consecutively recruited from a perinatal mental health outpatient unit completed the PBQ at 4–6 weeks postpartum. Four weeks later, all mothers completed again the PBQ and were interviewed using the Birmingham Interview for Maternal Mental Health to establish the presence of a bonding disorder. Receiver operating characteristic curve analysis revealed an area under the curve (AUC) value for the PBQ total score of 0.93, 95% CI [0.88, 0.98], with the optimal cut-off of 13 for detecting bonding disorders (sensitivity: 92%, specificity: 87%). Optimal cut-off scores for each scale were also obtained. The test-retest reliability coefficients were moderate to good. Our data confirm the validity of PBQ for detecting bonding disorders in Spanish population.

Received 29 April 2021; Revised 13 September 2021; Accepted 16 September 2021.

Keywords: clinical sample, external validation, mother-infant relationship disorder, Postpartum Bonding Questionnaire, test-retest reliability

Postpartum bonding disorders refers to the early disturbances in the feelings and affection of the mother towards her child (Taylor et al., 2005). It includes several clinical conditions, from mild disorders, such as delay, ambivalence or loss of maternal emotional response, to most severe disorders, such as pathological anger or rejection of the child (Brockington, Aucamp, et al. 2006). The prevalence rates range between 2.9%–8.6% (Edhborg et al., 2005; Garcia-Esteve et al., 2016) in the general population; whereas in mothers attending psychiatric services, the prevalence ranges between 39.4%–69.4% (Brockington, Fraser, et al., 2006; Loh & Vostanis, 2004; Siu et al., 2010). Disturbances in the mother-infant bonding can have a negative impact on the development of brain structures, leading to an effect on children's cognitive development and socio-emotional adaptation (Farré-Sender et al., 2018). This underlines the importance of early detecting and treating the mother-infant bonding disturbances.

Brockington, Fraser, et al. (2006) developed diagnostic criteria for bonding disorders, grouping them into: a) Mild disorder, if the mother experience delay in the onset, ambivalence, or loss of the maternal emotional response; b) infant focused anxiety, with two grades, mild and severe; c) pathological anger towards the infant, with three grades, mild, moderate and severe; d) threatened rejection, with a lack of a positive emotional response to the baby and the wish for temporary transfer of care; and e) established rejection, with a desire for permanent relinquishment of care.

Correspondence concerning this article should be addressed to Anna Torres-Giménez. Clínic Barcelona Hospital Universitari-Maternitat. Servei de Psiquiatria i Psicologia. Unitat de Salut Mental Perinatal. Carrer de Sabino Arana, 1. 08028 Barcelona (Spain). Phone: +34–932279948.

E-mail: atorresg@clinic.cat

Conflict of Interest: None.

Funding Statements: This research received no specific grant from any funding agency, commercial or not-for-profit sectors.

Author note: Both Anna Torres-Giménez and Alba Roca-Lecumberri contributed equally to the work and are considered as co-first authors of the present manuscript.

How to cite this article:
Torres-Giménez, A., Roca-Lecumberri, A., Sureda, B., Andrés-Perpiña, S., Palacios-Hernández, B., Gelabert, E., Farré-Sender, B., Subirà-Álvarez, S., & García-Esteve, L. (2021). External validation and test-retest reliability of postpartum bonding questionnaire in spanish mothers. *The Spanish Journal of Psychology, 24*. e47. Doi:10.1017/SJP.2021.44

Clinical interviews designed to explore the perinatal mental health, such as The Birmingham Interview for Maternal Mental Health – Fifth Edition (Brockington, Oates, et al., 2006), or the BIMMH - Sixth Edition (Stafford Interview) (Brockington et al., 2017), include a section on the mother-infant relationship that allows to establish the diagnostic of a bonding disorder. Brockington et al. (2001) also developed the Postpartum Bonding Questionnaire (PBQ) as screening instrument for detecting mothers at risk of bonding disorder. The final version of the instrument was composed by 25 items grouped into four scales: Scale 1, Impaired Bonding (34% of the variance; 12 items); Scale 2, Rejection and anger (8% of the variance; 7 items); Scale 3, Anxiety about care (3.7% of the variance; 4 items); and Scale 4, Risk of abuse (3.4% of the variance; 2 items). Test retest reliability was assessed by means of Pearson's product moment correlation coefficients between the first and the second administration of the instrument with an interval of about one hour (Brockington et al., 2001). The coefficients were .95, .95, .93 and .77 for the four scales.

The PBQ was translated and validated into multiple languages, including Chinese (Siu et al., 2010), Japanese (Ohashi et al., 2016; Suetsugu et al., 2015), Italian (Busonera et al., 2017), German (Reck et al., 2006) or Tamil (Vengadavaradan et al., 2019). Subsequent studies did not confirm the original factorial structure of the PBQ (Reck et al., 2006; Wittkowski et al., 2010). Garcia-Esteve et al. (2016) assessed the factorial validity and internal consistency reliability of the Spanish version of PBQ. A series of Confirmatory Factor Analyses did not replicate neither the original four-factor structure (Brockington et al., 2001) nor alternative structures (Reck et al. 2006; Wittkowski et al. 2010). A subsequent Exploratory Factor Analysis showed a four-factor solution: Factor 1: Impaired bonding; Factor 2: Anxiety about care; Factor 3: Lack of enjoyment and affection for the baby; and Factor 4: Rejection and risk of abuse (Garcia-Esteve et al., 2016). Finally, a Schmid-Leiman transformation was performed to assess the common and specific dimensions of the Spanish PBQ, finding a *general factor* that accounted for 61% of the variance of the PBQ (Garcia-Esteve et al., 2016). These results support the idea that the best solution is to consider one general factor which explains most of the variance of the Spanish PBQ.

Brockington and colleagues conducted two studies validating the PBQ against external criteria of bonding disorder (Brockington, Fraser, et al., 2006; Brockington et al., 2001). In 2001, 104 mothers were recruited from various sources, including general populations from maternity clinics, and mothers with bonding disorders. A subsample was interviewed using as a gold standard the Third edition of the Structured Interview for pregnancy-associated disorders (which later became the BIMMH). The cut-off scores used for each scale were: Scale 1 >12, Scale 2 >16, Scale 3 >9 and Scale 4 >2. Subsequently, in the 2006 study, 125 mothers were recruited from other specialists. The cut-off scores established in the second study for each scale were: Scale 1 >12, Scale 2 >12, Scale 3 >9 and Scale 4 >1. Additionally, a cut off score of 26 in the total score was proposed to identify "any type of bonding disorder" and a cut off score of 40 to identify a severe bonding disorder. The Chinese version of the PBQ was administered to 62 postnatal women recruited from a Perinatal Specialist Outpatient Psychiatrist Clinic (Siu et al., 2010). In this study the sensitivity and specificity in the Chinese sample of the cut-off scores previously established by Brockington were calculated (Siu et al., 2010). The Tamil version also established cut-off points for the PBQ subscales, but their results are not comparable with the other versions because they modified the length of the instrument (Vengadavaradan et al., 2019).

The aim of the present study was to validate the Spanish PBQ against external criteria of bonding disorder, as well as to establish its test-retest reliability.

Method

Participants and Procedure

The research was approved by the institutional review board of the first author's institution. All women signed an informed consent form. This study was conducted according to the Declaration of Helsinki. We report how we determined our sample size, all data exclusions, all data inclusion/exclusion criteria, whether inclusion/exclusion criteria were established prior to data analysis, all measures in the study, and all analyses including all tested models.

The sample used for the validation study was a convenience sample composed by 156 postpartum women, recruited from a perinatal mental health outpatient unit, between February 2014 and January 2018. This sample size is sufficient to estimate confidence intervals of proportions with a precision of less than 11% and assuming that the sensitivity and specificity of the questionnaire ranged between .80 and .90. Inclusion criteria, established prior to data analyses, were women with psychiatric disorder in pregnancy or postpartum who was treated in the perinatal mental health unit. The women who did not understand Spanish, those who had difficulties in filling the PBQ and those who had a dead newborn were excluded from the study.

The mothers completed at 4–6 weeks postpartum an assessment protocol that included sociodemographic variables and the PBQ. Subsequently, at 8–10 weeks postpartum, all mothers were assessed for the presence of bonding disorder by means of the Fifth and Sixth Edition

https://doi.org/10.1017/SJP.2021.44 Published online by Cambridge University Press

of the BIMMH, and in the same day, completed again the PBQ. Four perinatal clinical psychologists conducted the interviews and established the presence and type of bonding disorder. They were blind of PBQ scores.

Instruments

PBQ. The PBQ (Brockington, Fraser, et al., 2006; Brockington et al., 2001) is a self-report instrument that assesses the risk of presenting disorders in the mother-infant relationship during the postpartum period. It includes four subscales with a total of 25 items which are rated by the mother on a verbal frequency scale from 0 = *never* to 5 = *always*, as established in Appendix 1 of the article published by the authors who developed the questionnaire (Brockington, Fraser, et al., 2006). The PBQ has been translated into Spanish in a previous publication, presenting a factorial structure of four factors and a general factor that explained 61% of the variance of the PBQ (Garcia-Esteve et al., 2016). The factorial structure of the Spanish validation differed substantially from that found in the original study (Brockington et al., 2001). Factor 1, called "*impaired bonding*", included Items 1, 2, 3, 5, 6, 17, 22 and 23, which reflected topics such as emotional distance, regret about having the baby or unavailability to take care of the baby ("I feel distant from my baby"). Regarding the original Brockington's *impaired bonding* factor (Brockington et al., 2001), only 4 items coincided. Factor 2, called "*anxiety about care*", included Items 7, 10, 12, 13, 19, 20, 21 and 25, referring to maternal distress ("my baby makes me feel anxious"). This factor included three of the four items that made up Brockington's original *anxiety about care* factor; the rest of the items made up the original Brockington's *impaired bonding* factor. Factor 3, called "*lack of enjoyment and affection for the baby*", included Items 4, 8, 9, 11 and 16 (inversely scored: "I love to cuddle my baby"). This factor included three items that were included in Brockington''s Factor 1 and 2 items that were in Brockington's Factor 2 (*rejection and anger*). Finally, Factor 4, called "*rejection and risk of abuse*", included Items 14, 15, 18 and 24, which described negative feelings towards the baby and behaviors that put the baby's welfare at risk ("I feel angry with my baby"). This factor included the two items that constituted Brockington's *risk of abuse* factor. The internal consistency showed values between .56 and .85 for the scales, and .90 for the total score (Garcia-Esteve et al., 2016). In the present study, internal consistency values were .89 for the total score, .82 for the Factor 1, .74 for the Factor 2, .83 for the Factor 3, and .31 for the Factor 4. For the present study, PBQ scores were computed for the total scale, and for each of the factors described in the previous Spanish validation (Garcia-Esteve et al., 2016).

BIMMH. The *Fifth and Sixth Edition of the BIMMH* (Brockington et al., 2017; Brockington, Chandra, et al., 2006 were used to establish the external criterion of bonding disorder, according to Brockington, Fraser et al. (2006) diagnostic criteria. They were designed to explore the social, psychological and psychiatric course of pregnancy, birth and postpartum. A section on the mother-infant relationship that assess abnormalities in the mother-infant bond was included in both editions. This section explores infant characteristics and maternal involvement in care ("Please, tell me what (name of the baby) is like. Is s/he easy to understand and soothe?"), mother's emotional response to her infant ("How did your feelings for (name of baby) develop after the birth?"), and anger and abuse ("Do you feel angry with your baby?"). A subsample of 45 mothers were assessed blindly by two raters (BP, EG) to establish the inter-rater reliability.

Socio-demographic data. Mothers completed a form which records information on sociodemographic (How would you define your financial situation?), obstetric variables (Have you voluntarily terminated a pregnancy?) and affective disorders throughout life and during pregnancy (At any time in your life have you received medical treatment for depression, nerves, insomnia, or other emotional problems?).

Statistical Analyses

Missing values were explored for each variable. PBQ scores were considered valid if the scale was completed in its entirety. At the evaluation at 4–6 weeks postpartum, the PBQ was incomplete in 31 mothers; while at the evaluation at 8–10 weeks postpartum, all mothers completed the PBQ in its entirety. A Shapiro-Wilk test for normality was run for the PBQ scores, confirming that the distribution in each group was non-normal. Sociodemographic characteristics were summarized by descriptive statistics. Bonding disorder and severe bonding disorder prevalence rates were estimated with 95% confidence intervals (CI). Inter-rater reliability of bonding disorder was established using Cohen's Kappa coefficients in the subsample of 45 mothers assessed by two raters. κ values < .40 indicate poor agreement, .41 < κ < .60 indicate fair agreement, .61 < κ < .80 indicate good agreement, and κ values > 0.80 indicate excellent agreement (Landis & Koch, 1977).

The distribution of the PBQ scores by diagnostic groups was described using box-plot diagrams in which the median, interquartile range (IQR) (25^{th}–75^{th}) and outliers are represented graphically. Comparisons between mean PBQ scores by diagnostic groups were analyzed through Kruskal–Wallis test. Post hoc comparisons were performed using the Mann–Whitney *U* test. Mean differences between the first and second

https://doi.org/10.1017/SJP.2021.44 Published online by Cambridge University Press

Table 1. *Diagnostic Groups of Mother-Infant Bonding in the Study Sample (N = 156)*

Diagnostic group	*N*	%	κ (*n* = 45)	PBQ total score 4–6 weeks (*n* = 125)		PBQ total score 8–10 weeks (*n* = 156)		Post hoc comparisons
				Mdn	IQR	*Mdn*	IQR	
Mothers with normal bond (A)	104	66.7	.85	6	6	6	6	PBQ 4–6 weeks: C > B > A
Mothers with some kind of bonding disorder in the previous weeks but not at the time of the interview (8–10 weeks) (B)	14	9.0	.88	13	10	8.5	10	PBQ 8–10 weeks: C > A, B
Mothers with any type of bonding disorder (C)	38	24.3	.76	24	13	22	8	
Mild bonding disorders (D)	21	13.5	.63	22	15	19	11	PBQ 4–6 weeks: No differences between groups PBQ 8–10 weeks: G > D
Mild bonding disorders with infant-focused anxiety (E)	8	5.1	.79	27	7.5	23	8.5	
Mild bonding disorders with pathological mild anger (F)	3	1.9	1	18	11	18	9	
Threatened Rejection (G)	6	3.8	1	30	2	29	21	
Established Rejection	0	0	-	-	-	-	-	

Note. PBQ = Postpartum Bonding Questionnaire; IQR = interquartile range.

administration of PBQ by diagnostic group were assessed by means of Wilcoxon matched-pairs signed-ranks test.

The validity of the PBQ for detecting bonding disorder was analyzed using receiver operating characteristic (ROC) curves, and the area under the curve (AUC) was calculated with a 95% confidence interval (CI). Sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV) and positive (+LR) and negative (-LR) likelihood ratio were calculated for a range of cut off scores against the presence of external criteria of bonding disorder. Cohen's Kappa coefficients were calculated for assessing the concordance of PBQ cut off scores and diagnostic of bonding disorder. AUC values were compared using DeLong's method (DeLong et al., 1988). Test-retest reliability was analyzed using the intra-class correlation coefficient (ICC). An ICC ≥ .70 indicates good test-retest reliability (Terwee et al., 2007). Analyses were performed using STATA v.11.1, MedCalc v.7.3 and SPSS v.18. Statistical significance was set at $p < .05$.

Results

Characteristics of Participants

The mean age of the mothers was 35.0 years (range: 20–45). Ninety-four per cent were living with a partner, most of them was born in Spain (84.5%), only 1.3% had no primary education, and 83.9% completed high school or a university degree. Half of the mothers were primiparous (54.2%), a total of 20.9% reported financial difficulties, 75.7% were employed, 70.9% had planned the pregnancy, and 21.1% required some form of assisted reproduction. A total of 59.2% of babies were boys, and 12.6% had a low birth weight (< 2,500 gr.). Not breastfeeding after childbirth was reported by 25.9% of the mothers.

Regarding the distribution of principal psychiatric diagnoses in postpartum, 43% of the mothers had a depressive disorder, 36.5% had an anxiety disorder, 8.3% had an obsessive-compulsive disorder, 3.2% had a posttraumatic stress disorder, 2.6% had a bipolar disorder, 1.9% had a psychotic disorder, 1.9% had a personality disorder, 0.6% had an attention deficit hyperactivity disorder, 0.6% had a substance use disorder, and 0.6% had a dissociative disorder. Eighty per cent of the sample had previous personal history of psychiatric disorder, and 49% family history of psychiatric disorder.

Prevalence of Bonding Disorder and Distribution of PBQ Scores

The prevalence of mothers who met the criteria for bonding disorder at the time of the assessment was 24.3%, 95% CI [17.5, 31.2], while the prevalence of severe bonding disorder was 3.8%, 95% CI [0.8%, 6.9%]. Diagnostic concordance varied as a function of specific

https://doi.org/10.1017/SJP.2021.44 Published online by Cambridge University Press

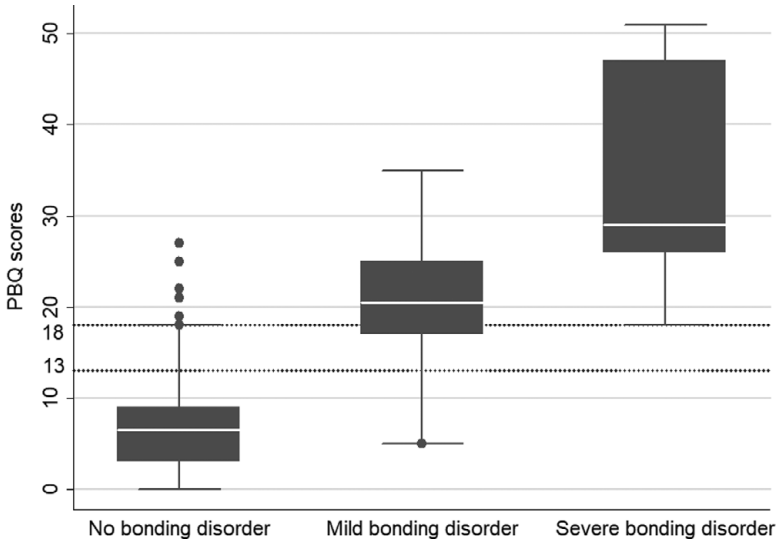

Figure 1. Box Diagram and Distribution of the Postpartum Bonding Questionnaire (PBQ) Scores (Median, Quartiles and Outliers) by Diagnostic Groups

Note. The dotted lines represent the optimal cut-off points for detecting bonding disorder and severe bonding disorder. PBQ = Postpartum Bonding Questionnaire.

diagnostic group, with Kappa values between .63 and 1.00 (Table 1).

In our sample PBQ scores at 8–10 weeks postpartum had a median (range) of 8.5 (0– 51). The Figure 1 shows the distribution of PBQ scores at 8–10 weeks postpartum according to three diagnostic groups: Absence of bonding disorder ($n = 118$), mild bonding disorder ($n = 32$), and severe bonding disorder ($n = 6$). PBQ scores at 8–10 weeks postpartum showed an association with diagnostic group (Kruskal-Wallis $\chi^2 = 63.72$, $df = 2$, $p < .001$). Post hoc analyses revealed that mothers with mild bonding disorder ($p < .001$), and with severe bonding disorder ($p < .001$) scored higher than mothers without bonding disorder; whereas mothers with severe bonding disorder ($p < .05$) scored higher than mothers with mild bonding disorder. Table 1 shows the distribution of PBQ scores at 8–10 weeks postpartum according to normal bond ($n = 104$), bonding disorder in the previous weeks but not at the time of the assessment ($n = 14$), and any type of bonding disorder ($n = 38$) (Kruskal-Wallis $\chi^2 = 64.05$, $df = 2$, $p < .001$). Post hoc analyses revealed that mothers with any type of bonding disorder scored higher than mothers with normal bond ($p < .001$) and mothers with bonding disorder in the previous weeks but not at the time of the assessment ($p < .001$). Regarding specific types of bonding disorder (Table 1), post hoc comparisons revealed that mothers with threatened rejection scored higher on PBQ scores at 8–10 weeks postpartum than mothers with mild bonding disorders (without infant-focused anxiety or pathological mild anger) ($p < .01$). Finally, there were no differences between the two assessments (PBQ at 4–6 weeks and PBQ at 8–10 weeks) within each diagnostic group.

Validation of the PBQ for Detection of Bonding Disorder

The AUC value for PBQ total score was .93, 95% CI [.88, .98], indicating an excellent validity for detecting bonding disorder. For the PBQ total score, 13 was the optimal cut-off for detecting bonding disorder, with a sensitivity of 92.1%, 95% CI [78.6, 98.2], a specificity of 87.3%, 95% CI [79.9, 92.7], a PPV of 70.0% and a NPV of 97.2%. Table 2 shows the AUC values of PBQ total score and PBQ scales for detecting bonding disorder and severe bonding disorder, as well as their optimal cut off points, sensitivities, specificities, positive predictive values, negative predictive values, positive and negative likelihood ratios, and Kappa values with clinical interview diagnosis. Full range of PBQ scores with corresponding values of sensitivity, specificity, predictive values, and likelihood ratios are available as supplementary material (Table S1–Table S10, in the Supplementary_Material_External_Validation_PBQ [SMPBQ]). The comparison of AUC values showed that the AUC values of the PBQ total score and PBQ Scale 1 were not significantly different. PBQ total score, $\chi^2(1, N = 157) = 5.80$, $p = .02$, performed slightly better than PBQ Scale 2 for detecting bonding disorder. PBQ total score, $\chi^2(1, N = 157) = 15.41$, $p < .001$, and PBQ Scale 1, $\chi^2(1, N = 157) = 10.54$, $p = .001$) performed better than PBQ Scale 3. Finally, PBQ Scale 4 performed worse than PBQ total score, $\chi^2(1, N = 157) = 16.71$, $p < .001$, PBQ Scale 1, $\chi^2(1, N = 157) = 9.97$, $p = .002$, and PBQ Scale 2, $\chi^2(1, N = 157) = 5.69$, $p = .017$, (Table 2 and Figure 2).

The AUC value for PBQ total score was .92, 95% CI [.83, 1.00], indicating a good validity for detecting

https://doi.org/10.1017/SJP.2021.44 Published online by Cambridge University Press

Table 2. *Areas under the ROC Curves and Optimal Cut off Points of PBQ Total Score and the Four Scales of Spanish PBQ for Detecting Bonding Disorder and Severe Bonding Disorder*

		95% CI				95% CI			95% CI						
	AUC	*LL*	*UL*	Cut off	Sens.	*LL*	*UL*	Spec.	*LL*	*UL*	PPV	NPV	+LR	-LR	κ
Bonding disorder															
PBQ total score	.93	.88	.98	13	92.1	78.6	98.2	87.3	79.9	92.7	70.0	97.2	7.25	0.09	.72
PBQ F1	.91	.86	.96	4	78.9	62.7	90.4	90.7	83.9	95.2	73.2	93.0	8.47	0.23	.68
PBQ F2	.87	.81	.93	9	78.9	62.7	90.4	82.2	74.1	88.6	58.8	92.4	4.44	0.26	.55
PBQ F3	.77	.68	.86	2	60.5	43.4	75.9	91.5	85.0	95.9	69.7	87.8	7.14	0.43	.54
PBQ F4	.79	.71	.87	1	73.7	56.9	86.6	81.4	73.1	87.9	56.0	90.6	3.95	0.32	.50
Severe bonding disorder															
PBQ total score	.92	.83	1.00	18	100.0	54.1	100.0	78.0	70.5	84.3	15.4	100.0	4.55	0.00	.21
PBQ F1	.97	.93	1.00	6	100.0	54.1	100.0	84.0	77.1	89.5	20.0	100.0	6.25	0.00	.29
PBQ F2	.83	.70	.95	8	100.0	54.1	100.0	60.0	51.7	67.9	9.1	100.0	2.50	0.00	.10
PBQ F3	.93	.87	.99	2	100.0	54.1	100.0	82.0	74.9	87.8	18.2	100.0	5.56	0.00	.26
PBQ F4	.74	.57	.91	1	83.3	36.1	97.2	70.0	62.0	77.2	10.0	99.1	2.78	0.24	.12

Note. ROC = Receiver operating characteristic; PBQ = Postpartum Bonding Questionnaire; AUC = Area under the curve; CI = confidence interval; Sens. = sensitivity; Spec. = specificity; LL = lower limit; UL = upper limit; PPV = positive predictive value; NPV = negative predictive value; +LR = positive likelihood ratio; -LR = negative likelihood ratio; κ = Cohen's Kappa coefficients with clinical interview diagnosis.

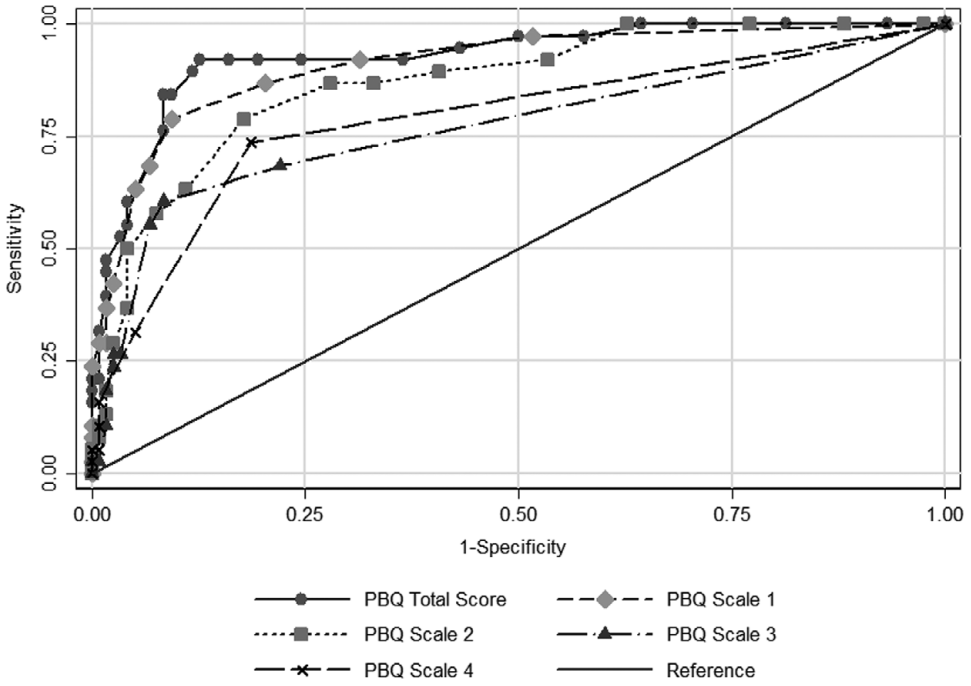

Figure 2. Receiver Operating Characteristic curve of the Postpartum Bonding Questionnaire (PBQ) Total Score and Scales 1, 2, 3 and 4 for Detecting Bonding Disorder
Note. PBQ = Postpartum Bonding Questionnaire.

https://doi.org/10.1017/SJP.2021.44 Published online by Cambridge University Press

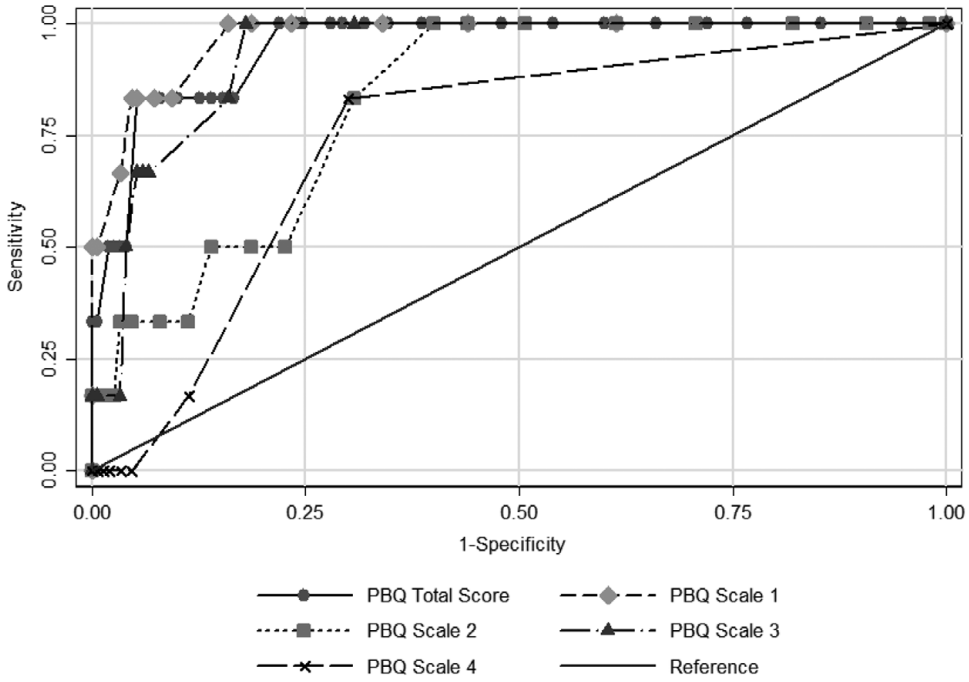

Figure 3. Receiver Operating Characteristic Curve of the Postpartum Bonding Questionnaire (PBQ) Total Score and Scales 1, 2, 3 and 4 for Detecting Severe Bonding Disorder
Note. PBQ = Postpartum Bonding Questionnaire.

severe bonding disorder. For the PBQ total score, 18 was the optimal cut-off for detecting severe bonding disorder, with a sensitivity of 100.0%, 95% CI [54.1, 100.0], a specificity of 78.0%, 95% CI [70.5, 84.3], a PPV of 15.4% and a NPV of 100.0%. The AUC values and optimal cut off points of PBQ scales for detecting severe bonding disorder are shown in Table S2 (SMPBQ). The AUC values of PBQ total score, PBQ Scale 1 and PBQ Scale 3 were not significantly different. PBQ total score, $\chi^2(1, N = 157) = 5.93$, $p = .01$; PBQ Scale 1, $\chi^2(1, N = 157) = 6.28$, $p = .01$; and PBQ Scale 3, $\chi^2(1, N = 157) = 4.67$, $p = .03$, performed better than PBQ Scale 2. PBQ total score, $\chi^2(1, N = 157) = 5.87$, $p = .01$; PBQ Scale 1, $\chi^2(1, N = 157) = 6.48$, $p = .01$; and PBQ Scale 3, $\chi^2(1, N = 157) = 6.31$, $p = .01$, performed better than PBQ Scale 4 (Table 2 and Figure 3).

Test-retest Reliability

The ICC value was .78, 95 % CI [.69, .85]; for the PBQ total score, .75, 95% CI [.64, .82]; for the PBQ Scale 1, .80, 95% CI [.71, .86]; for the PBQ Scale 2, .75, 95% CI [.64, .82]; for the PBQ Scale 3, and .81, 95% CI [.73, .87], for the PBQ Scale 4.

Discussion

This study provides an external validation of the Spanish version of the PBQ for detecting bonding disorders against a gold standard external criterion. It complements a previous work in which the Spanish PBQ showed adequate reliability and a factorial validity (Garcia-Esteve et al., 2016), providing initial support for good diagnostic accuracy and adequate test-retest reliability.

There are limitations to the study that should be taken into account. The factorial structure of the PBQ has not been replicated across the different studies (Garcia-Esteve et al., 2016; Reck et al., 2006; Wittkowski et al., 2010). In the present study, the factorial structure found in the Spanish version has been used (Garcia-Esteve et al., 2016). This structure arises from an exploratory factor analysis after not confirming the previous structures by confirmatory factor analysis. This structure is substantially different from that found by the original author and later studies (Brockington et al., 2001; Reck et al., 2006; Wittkowski et al., 2010). In addition, it has limitations, such as the poor reliability of Factor 4. For these reasons, it would be necessary to replicate this factorial structure in other studies with the Spanish population. All this suggests the use of cut off points for the PBQ total score and discourages the use of PBQ scales; especially it is recommended to avoid using the cut-off points of PBQ Scale 4, due to their low reliability.

Another limitation refers to the differences between the cut-off points found in comparison with previous studies. The optimal cut off scores for PBQ total score obtained in our study are much lower than the scores established in previous studies (Brockington, Fraser,

https://doi.org/10.1017/SJP.2021.44 Published online by Cambridge University Press

et al., 2006; Siu et al., 2010). In our study, 13 was the optimal cut off score for detecting bonding disorder (sensitivity: 92%, specificity: 87%), and 18 was the optimal cut off for detecting severe bonding disorder (sensitivity: 100%, specificity: 78%). In previous studies, cut off scores where 26 for detecting any type of bonding disorder, and 40 for detecting severe bonding disorder (Brockington, Fraser, et al., 2006; Siu et al., 2010). Sociocultural and methodological differences could explain these differences. Our sample had a median PBQ total score of 8.5, whereas in the Brockington, Fraser, et al. study (2006) the median was 30. Although in our study there were fewer cases of bonding disorder, and severe bonding disorder, this does not fully explain the differences in median scores. Therefore, sociocultural factors as well as social desirability bias may partly explain the differences in PBQ scores. Clinically, there was a tendency for some mothers to underestimate PBQ scores, while in the interview they did refer clear difficulties in the mother-infant bonding. Furthermore, to establish cut off scores we use receiver operating characteristic (ROC) curves. ROC curves allow the exploration of the entire range of sensitivities and specificities at each possible cut-off score. This disparity in our cut-off points compared to previous publications suggests the need to replicate these cut-off points in other samples with a Spanish population.

The third limitation refers to the cut-off points for severe bonding disorders. The small sample size (n = 6) as well as the high dispersion of the PBQ scores in these cases leads to cut-off points with wide sensitivity confidence intervals and low positive predictive values, limiting the validity of the cut-off points. It would be necessary to design future validation studies enlarging the sample size of severe bonding disorders, providing a wide spectrum of severe bonding disorders, including rejection, severe anger or severe anxiety. Another limitation is the lack of an instrument of observation of the mother-infant interaction.

In our clinical sample, the prevalence of bonding disorder was 24.3%, whereas the prevalence of severe bonding disorder was 3.8%. These prevalence rates are higher than those found in general population (Brockington, Aucamp, et al. 2006; Edhborg et al., 2005; Garcia-Esteve et al., 2016; Reck et al., 2006; van Bussel et al., 2010), but lower than observed in previous studies from perinatal psychiatric services (Brockington, Fraser, et al., 2006; Loh & Vostanis, 2004; Siu et al., 2010). Although we use the same clinical diagnostic interview, differences in sample characteristics could explain partially our results. Our sample included women treated in a perinatal mental health outpatient unit for a psychiatric disorder either in pregnancy or postpartum, whereas in previous studies the mothers were referred in postpartum. Furthermore, our sample included a wide variety of mental disorder diagnoses, whereas several studies included only mothers with postpartum depression (Loh & Vostanis, 2004; Siu et al., 2010).

The intraclass correlation coefficients were .75–.81, indicating a good test-retest reliability. The stability was higher for the Scales 2 "anxiety about care" and 4 "rejection and risk of abuse". The values were clearly lower than those found by Brockington's study (2001), but the interval between administrations was one hour, while in our study the time interval was between two and four weeks. A very short time interval may have artificially increased the correlation between both administrations. However, the values were slightly lower than obtained by the Japanese validation of PBQ (Ohashi et al., 2016), in which the test-retest reliability was assessed between Day 5 and 1 month after childbirth, indicating a higher stability of the measure during the first four weeks after childbirth.

The present study provides an external validation of the Spanish version of the Postpartum Bonding Questionnaire for detecting bonding disorders against a gold standard external criterion. The PBQ demonstrated adequate diagnostic accuracy for detecting bonding disorders, as well as moderate to good test-retest reliability coefficients. These results confirm the utility of PBQ as screening tool of mother-infant bonding difficulties. This study calls for replication in other Spanish samples.

Supplementary Materials

To view supplementary material for this article, please visit http://dx.doi.org/10.1017/SJP.2021.44.

References

Brockington, I. F., Aucamp, H. M., & Fraser, C. (2006). Severe disorders of the mother-infant relationship: Definitions and frequency. *Archives of Women's Mental Health*, *9*(5), 243–251. https://doi.org/10.1007/s00737-006-0133-0

Brockington, I. F., Fraser, C., & Wilson, D. (2006). The Postpartum Bonding Questionnaire: A validation. *Archives of Women's Mental Health*, *9*(5), 233–242. https://doi.org/10.1007/s00737-006-0132-1

Brockington, I. F., Oates, J., George, S., Turner, D., Vostanis, P., Sullivan, M., Loh, C., & Murdoch, C. (2001). A Screening Questionnaire for mother-infant bonding disorders. *Archives of Women's Mental Health*, *3*(4), 133–140. https://doi.org/10.1007/s007370170010

Brockington, I. F., Chandra, P., Bramante, A., Dubow, H., Fakher, W., Garcia-Esteve, L., Hofberg, K., Moussa, S., Palacios-Hernández, B., Parfitt, Y., & Shieh, P.-L. (2017). The Stafford Interview : A comprehensive interview for mother-infant psychiatry. *Archives of Women's Mental Health*, *20*(1), 107–112. https://doi.org/10.1007/s00737-016-0683-8

Brockington, I. F., Chandra, P., Felice, E., George, S., Grybos, M., Hofberg, K., Kiejna, A., Kitzol, M., Lanczik, M. H., Loh, C., Niemela, P., Rymaszewska, J., Rondon, M., Shi, S., Vostania, P., & Wainscott, G. (2006). *The Birmingham interview for maternal mental health* (5th Ed.). Eyre Press.

Busonera, A., Cataudella, S., Lampis, J., Tommasi, M., & Zavattini, G. C. (2017). Psychometric properties of the Postpartum Bonding Questionnaire and correlates of mother–infant bonding impairment in Italian new mothers. *Midwifery*, *55*, 15–22. https://doi.org/10.1016/j.midw.2017.08.011

DeLong, E. R., DeLong, D. M., & Clarke-Pearson, D. L. (1988). Comparing the areas under two or more correlated receiver operating characteristic curves: A nonparametric approach. *Biometrics*, *44*(3), 837–845. https://doi.org/10.2307/2531595

Edhborg, M., Matthiesen, A.-S., Lundh, W., & Widström, A.-M. (2005). Some early indicators for depressive symptoms and bonding 2 months postpartum-A study of new mothers and fathers. *Archives of Women's Mental Health*, *8*(4), 221–231. https://doi.org/10.1007/s00737-005-0097-5

Farré-Sender, B., Torres, A., Gelabert, E., Andrés, S., Roca, A., Lasheras, G., Valdés, M., & Garcia-Esteve, L. (2018). Mother-infant bonding in the postpartum period: Assessment of the impact of pre-delivery factors in a clinical sample. *Archives of Women's Mental Health*, *21*(3), 287–297. https://doi.org/10.1007/s00737-017-0785-y

Garcia-Esteve, L., Torres, A., Lasheras, G., Palacios-Hernández, B., Farré-Sender, B., Subirà, S., Valdés, M., & Brockington, I. F. (2016). Assessment of psychometric properties of the Postpartum Bonding Questionnaire (PBQ) in Spanish mothers. *Archives of Women's Mental Health*, *19*(2), 385–394. https://doi.org/10.1007/s00737-015-0589-x

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, *33*(1), 159–174. https://doi.org/10.2307/2529310

Loh, C.-C., & Vostanis, P. (2004). Perceived mother-infant relationship difficulties in postnatal depression. *Infant and Child Development*, *13*(2), 159–171. https://doi.org/10.1002/icd.347

Ohashi, Y., Kitamura, T., Sakanashi, K., & Tanaka, T. (2016). Postpartum Bonding Disorder: Factor structure, validity, reliability and a model comparison of the Postnatal Bonding Questionnaire in Japanese mothers of infants. *Healthcare*, *4*(3), Article 50. https://doi.org/10.3390/healthcare4030050

Reck, C., Klier, C. M., Pabst, K., Stehle, E., Steffenelli, U., Struben, K., & Backenstrass, M. (2006). The German version of the Postpartum Bonding Instrument: Psychometric properties and association with postpartum depression. *Archives of Women's Mental Health*, *9*(5), 265–271. http://doi.org/10.1007/s00737-006-0144-x

Siu, B. W.-M., Ip, P., Chow, H. M.-T., Kwok, S. S.-P., Li, O.-L., Koo, M.-L., Cheung, E. F.-C., Yeung, T. M.-H., & Hung, S.-F. (2010). Impairment of mother-infant relationship. *The Journal of Nervous and Mental Disease*, *198*(3), 174–179. https://doi.org/10.1097/NMD.0b013e3181d14154

Suetsugu, Y., Honjo, S., Ikeda, M., & Kamibeppu, K. (2015). The Japanese version of the Postpartum Bonding Questionnaire: Examination of the reliability, validity, and scale structure. *Journal of Psychosomatic Research*, *79*(1), 55–61. https://doi.org/10.1016/j.jpsychores.2015.02.008

Taylor, A., Atkins, R., Kumar, R., Adams, D., & Glover, V. (2005). A new Mother-to-Infant Bonding Scale: Links with early maternal mood. *Archives of Women's Mental Health*, *8*(1), 45–51. https://doi.org/ 10.1007/s00737-005-0074-z

Terwee, C. B., Bot, S. D. M., de Boer, M. R., van der Windt, D. A. W. M., Knol, D. L., Dekker, J., Bouter, L. M., & de Vet, H. C. W. (2007). Quality criteria were proposed for measurement properties of health status questionnaires. *Journal of Clinical Epidemiology*, *60*(1), 34–42. https://doi.org/10.1016/j.jclinepi.2006.03.012

van Bussel, J. C. H., Spitz, B., & Demyttenaere, K. (2010). Three self-report questionnaires of the early mother-to-infant bond: Reliability and validity of the Dutch version of the MPAS, PBQ and MIBS. *Archives of Women's Mental Health*, *13*(5), 373–384. https://doi.org/10.1007/s00737-009-0140-z

Vengadavaradan, A., Bharadwaj, B., Sathynarayanan, G., Durairaj, J., & Rajaa, S. (2019). Translation, validation and factor structure of the Tamil version of the Postpartum Bonding Questionnaire (PBQ-T). *Asian Journal of Psychiatry*, *40*, 62–67. https://doi.org/10.1016/j.ajp.2019.01.018

Wittkowski, A., Williams, J., & Wieck, A. (2010). An examination of the psychometric properties and factor structure of the Postpartum Bonding Questionnaire in a clinical inpatient sample. *The British Journal of Clinical Psychology/The British Psychological Society*, *49*, 163–172. https://doi.org/10.1348/014466509X445589