Abstract

Postpartum post-traumatic stress disorder (PP-PTSD) affects 3.1-6.3% of women after childbirth. The City Birth Trauma Scale (City-BiTS) is a questionnaire designed to evaluate and diagnose this disorder, according to the DSM-5 criteria, including the following groups of symptoms characteristic of post-traumatic stress: re-experiencing, avoidance, negative cognitions and mood, and hyperarousal. The aim of the present study was to evaluate the psychometric properties of the Spanish-language version of this questionnaire (City-BiTS-S), based on a community sample of Spanish women. A total of 207 mothers, recruited at three health centres in southern Spain, completed the City-BiTS-S questionnaire and provided sociodemographic and obstetric data. Exploratory factor analysis (EFA) of the data replicated the two-factor structure reported in previous studies that explained 47.9% of the variance: Factor 1 of general symptoms, and Factor 2 of birth-related symptoms. Both City-BiTS-S (Cronbach's $\alpha = 0.90$) and the two factors (Cronbach's α for Factor 1 = 0.89; Cronbach's α for Factor 2 = 0.82) presented high internal consistency. Rasch analysis confirmed the unidimensionality of the two factors as valid subscales of the PP-PTSD. It also confirmed that the item about *'Not being able to remember details of the birth'* should be reviewed. Rasch analysis also suggested the item *'Flashbacks to the birth and/or reliving the experience*' should be reworded in the Spanish version. In conclusion, the City-BiTS-S presents appropriate psychometric properties to measure symptoms of postpartum_PTSD. Nevertheless, further research is recommended to confirm its validity in a clinical population and in different medical approaches to the birth process.

Key words: *PTSD, Postpartum, Questionnaire, Validation, Spanish.*

Introduction

Each year, 134 million births take place, worldwide. However, although motherhood is assumed to be a joyous experience, for many women pregnancy and childbirth are stressful, traumatic experiences (Ayers & Pickering, 2001; Caparros-Gonzalez et al., 2019a; Dekel, Stuebe & Dishy, 2017). Studies have shown that traumatic childbirth can provoke severe psychological distress in mothers, some of whom may consequently develop postpartum post-traumatic stress disorder (PP-PTSD) (Ayers, Bond, Bertullies & Wijma, 2016). According to recent reviews and meta-analyses, 3.1-6.3% of women suffer from this disorder after childbirth, a figure that rises to 15.7-18.9% among high risk women, such as those with a history of psychiatric disorder and/or complications during childbirth (Dekel et al., 2017; Grekin & O'Hara, 2014; Yildiz, Ayers & Phillips, 2017a).

The symptoms that define post-traumatic stress disorder are re-experiencing the traumatic event, negative cognitions and mood, avoidance of stimuli associated with the trauma, and a high level of stress and physiological activation (American Psychiatric Association [APA], 2013; McKenzie-McHarg et al., 2015). In addition, PP-PTSD is also characterised by sensations of threat (towards the mother or the baby), dissociation and emotional numbness (Ayers, McKenzie-McHarg & Eagle, 2007; James, 2015). PP-PTSD also presents high rates of comorbidity with prepartum or postpartum depression, anxiety and prior episodes of post-traumatic stress (Agius, Xuereb, Carrick-Sen, Sultana & Rankin, 2016; Grekin & O'Hara, 2014).

Studies have highlighted the negative impact of PP-PTSD on women, a major aspect of which is fear of childbirth, which can provoke severe anxiety in future pregnancies and even discourage mothers from having more children (Ayers, McKenzie-McHarg & Slade, 2015). This disorder may also impact on interpersonal relationships and the mother-child bond, especially if the mother associates the baby with the traumatic experience of childbirth (Hairston, Handelzalts, Assis & Kovo, 2018; Parfitt & Ayers, 2009). Moreover, it has been suggested that PP-PTSD may be related to an increased risk of premature birth (Yonkers et al., 2014), poorer cognitive development of the infant (Parfitt, Pike & Ayers, 2014), behavioural problems and alteration of the child's social and emotional regulation (Cook, Ayers & Horsch, 2018; Garthus-Niegel, Ayers, Martini, von Soest & Eberhard-Gran, 2017).

Despite its prevalence and impact, PP-PTSD continues to receive insufficient attention in maternity services, and patients with this condition are rarely assessed and treated (Ayers,

Wright & Thornton, 2018). One reason for this unsatisfactory situation may be the lack of appropriate, validated assessment instruments with which to assess the disorder. To our knowledge only three specific assessment instruments have been developed for PP-PTSD: the Traumatic Event Scale (TES; Wijma, Söderquist & Wijma, 1997), the Perinatal PTSD Questionnaire (PPQ; Quinnell & Hynan, 1999; Callahan, Borja & Hynan, 2006) and the City Birth Trauma Scale (City-BiTS; Ayers et al., 2018). Although both TES and PPQ have been widely used in perinatal research, neither enables the evaluation and diagnosis of PP-PTSD according to current DSM-5 criteria.

The original version of City-BiTS had good internal consistency (Cronbach's $\alpha = 0.92$), with a two-factor structure of PTSD symptoms related to birth and general PTSD symptoms that explained 56.3% of the total variance. To date, the City-BiTS has been adapted and translated to 14 foreign-language versions, not including Spanish. Apart from the original version, the published versions include Hebrew and Croatian, both of which perform well in terms of reliability (Cronbach's $\alpha > 0.90$) and validity (Handelzalts, Hairston & Matatyahu, 2018; Nakic Radoš et al., 2019). However, no version of City-BiTS has been validated for use with a Spanish-speaking population.

The Rasch model is highly recommended for evaluating the measurement properties of scales (Cano & Hobart, 2011). Rasch analysis addresses essential assumptions of measurement such as unidimensionality, invariance along the construct and the stability of items across different feature groups (Wright & Stone, 1979). Moreover, Rasch analysis is increasingly used to validate assessment instruments (Aryadoust et al., 2019). However, as City-BiTS (Ayers et al., 2018) has only recently been published, no Rasch analysis has yet been applied to data from any sample in which this questionnaire was used.

In summary, the aim of the present study was to determine the psychometric properties of the City-BiTS questionnaire using Rasch analysis in a sample of Spanish postpartum women.

Method

Participants

A total of 236 women were initially recruited to the study. Subsequently, 12 decided not to participate (due to lack of time), 16 did not fill out the questionnaires and one woman's baby was stillborn. The final sample therefore consisted of 207 women, recruited at three public

health centres in southern Spain: Roquetas de Mar Health Centre (Almería), Antequera General Hospital (Málaga) and San Cecilio Clinical Hospital (Granada).

The inclusion criteria were that women were aged at least 18 years, the ability to read and write in Spanish, low-risk pregnancy and pregnancy care within the Spanish public health system. The exclusion criterion was the presence of any obstetric complication such as gestational diabetes or fetal anomaly during pregnancy. In order to obtain an accurate measure of PTSD in accordance with the DSM-5 criteria (APA, 2013), all participants completed the questionnaires after the first month after birth.

This study was approved by the Human Research Ethics Committee of the University of Granada, the Andalusian Biomedical Research Ethics Coordinating Committee and the Ethics Committee at each of the health centres involved. The study was carried out according to the guidelines of the Declaration of Helsinki (World Medical Association, 2013) and the European Union directive on good clinical practices (Directive 2005/28/EC). Participation was voluntary and each participant signed the informed consent form after reading the detailed Participant Information Sheet.

Instruments

The City Birth Trauma Scale (City-BiTS; Ayers et al., 2018) is a 29-item questionnaire based on the DSM-5 diagnostic criteria for PTSD (APA, 2013). The four main groups of symptoms (re-experiencing, avoidance, negative cognitions and mood, and hyperarousal) are measured via 20 items, scored on a Likert scale ranging from 0 (“never”), 1 ("once"), 2 (2-4 times) to 3 (“5 or more times”). The higher the score, the greater the severity of the symptoms. Three dichotomous response items (Yes/No) evaluate stress criteria in accordance with DSM-5 (belief of serious harm or death). Three additional items measure the level of distress (Yes/No/Sometimes), disability (Yes/No/Sometimes) and possible physical causes (Yes/No/Perhaps). Two items assess the onset of symptoms (before birth/first 6 months after birth/>6 months after birth) and their duration (<1 month/1-3 months/>3 months). The remaining two items are for diagnostic purposes to specify subtype of dissociative PTSD. The original version of City-BiTS had high internal consistency (Cronbach’s $\alpha = 0.92$) (Ayers et al., 2018).

The sociodemographic and obstetric data analysed in this study were obtained from the Health Document for Pregnant Women, published by the regional government of Andalusia (Consejería de Salud de la Junta de Andalucía, 2017).

Translation of the questionnaire

The questionnaire was translated in accordance with international guidelines for the intercultural adaptation of questionnaires (Martin & Savage-McGlynn, 2013). A team of experts in perinatal health (all with excellent Spanish and English) first translated City-BiTS into Spanish (Version 1). A professional translator-editor then back-translated the text into English (Version 2). The experts checked that the back-translated Version 2 was grammatically and semantically equivalent to the original questionnaire, and that there was no significant variability between the two versions. Finally, a pilot sample of 10 participants was asked to complete the Spanish-language questionnaire, to assess its understandability and to confirm that the items included addressed their concerns after childbirth. After this pilot study, no further changes were made to the measure. This procedure has been used successfully in previous translations and adaptations of instruments for measuring stress during pregnancy and satisfaction with childbirth (Caparros-Gonzalez et al., 2019b; Romero-Gonzalez et al., 2019).

Procedure

Participants were recruited by three clinical midwives (January 2018 - September 2019) while attending a postnatal appointment at Roquetas de Mar Health Center (Almería), Antequera Hospital (Málaga) or San Cecilio Hospital (Granada) in the south of Spain. Participants who agreed to take part in the study read and signed the informed consent form. Participants took the questionnaires to complete at home and returned them to the midwife at the following postnatal appointment.

Data analysis

Rasch analysis was conducted to determine construct validity and reliability of the scale. The Rasch model (Rasch, 1980) is a probabilistic model of measurement based on the Item Response Theory. The analysis transforms ordinal raw scores into linear measures -the interval unit of measurement is called *logit* (Tesio, 2003). This allows a location or calibration of all items according to their inherent difficulty to be endorsed. Difficulty of item gives

information about severity and frequency of the specific symptom in the latent construct. Rasch analysis determines construct validity of scales by evaluating items fit and location of each item in the latent construct (Linacre, 2002).

Rasch analysis based on the Partial Credit Model (Masters, 1982) was performed on the City-BiTS-S data using RUMM2020 software (Andrich et al., 2003). Analysis determined unidimensionality, overall fit to the Rasch model, individual item fit, targeting of participants, functioning of response categories and differential item functioning (DIF). Extended information about the protocol for Rasch analysis performance and reporting can be found elsewhere (Hagquist, Bruce & Gustavsson, 2009; Tennant & Conaghan, 2007). The sample size of the study must be such as to guarantee the stability of item location estimation at 99% confidence (Linacre, 1994).

Results

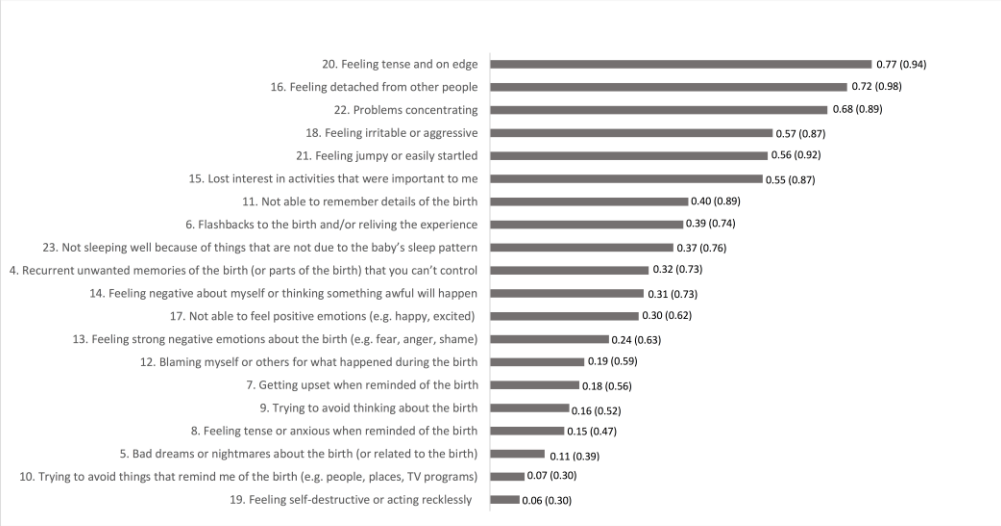
Characteristics of the sample population

The participants were aged between 19 and 45 years (M = 31.94; SD = 5.32). All had given birth to a healthy, full-term infant, with a one-minute Apgar test score of 9-10 (M = 9.17; SD = 0.83) and a five-minute score also of 9-10 (M = 9.27; SD = 0.87). The duration of the pregnancy was 35-42 weeks (M = 39.67 weeks; SD = 1.33), and the majority of the participants were primiparous (N = 123; 59.4%).

Most of the women had a spontaneous vaginal birth (N = 148; 71.2%). A further 29 births (13.9%) were instrumental vaginal births and 30 women (14.5%) required a caesarean section. The majority received epidural pain management during birth (N = 122; 58.7%), while 13 (6.3%) used alternative methods of pain relief (such as music or hydrotherapy) and 72 (34.6%) did not receive any pain relief during labour and birth. Total PTSD scores have a possible range of 0-60 and an actual range in our sample of 0-48 (M = 8.02; SD = 9.13). Most of the women in the sample (76.9%) reported at least one symptom of PTSD.

Figure 1 shows the mean score and standard deviation obtained for each of the City-BiTS-S items.

Figure 1. **Mean and standard deviation (in parentheses) of the scores obtained for each item**



Exploratory factor analysis

An exploratory factor analysis was performed of the principal components, with varimax rotation of the City-BiTS-S items, to test the structure of the instrument. The factors were selected by applying Kaiser's rule (retain factors with eigenvalue >1) and by analysing the sedimentation graph. This analysis showed that two factors accounted for 49.3% of the total variance of the data, and so the two subscales were grouped accordingly (Table 1).

The factor model was assessed using Bartlett's test of sphericity ($\chi 2$ (1) = 2280.4; $p <$ .001) and the Kaiser-Meyer-Olkin (KMO) test (0.83). In both cases, the results obtained were satisfactory. Factor 1 included items related to general symptoms and Factor 2 included items related to birth.

Table 1. Exploratory factor analysis outcome

Item	Factor 1	Factor 2
1. Recurrent unwanted memories of the birth (or parts of the birth) that you can't control	.124	**.784**
2. Bad dreams or nightmares about the birth (or related to the birth)	.082	**.633**
3. Flashbacks to the birth and/or reliving the experience	.139	**.376**
4. Getting upset when reminded of the birth	.165	**.796**
5. Feeling tense or anxious when reminded of the birth	.189	**.757**
6. Trying to avoid thinking about the birth	.074	**.842**
7. Trying to avoid things that remind me of the birth (e.g. people, places, TV programs)	-.099	**.692**
8. Not able to remember details of the birth	.105	.288
9. Blaming myself or others for what happened during the birth	.190	**.489**
10. Feeling strong negative emotions about the birth (e.g. fear, anger, shame)	.187	**.775**
11. Feeling negative about myself or thinking something awful will happen	**.668**	.159
12. Lost interest in activities that were important to me	**.776**	.108
13. Feeling detached from other people	**.807**	.119
14. Not able to feel positive emotions (e.g. happy, excited)	**.656**	.049
15. Feeling irritable or aggressive	**.769**	.160
16. Feeling self-destructive or acting recklessly	**.380**	.211
17. Feeling tense and on edge	**.772**	.107
18. Feeling jumpy or easily startled	**.768**	.168
19. Problems concentrating	**.758**	.073
20. Not sleeping well because of things that are not due to the baby's sleep pattern	**.614**	.168
% of variance	32.9	16.4

Internal consistency

Taking into account the 20 items that measure post-traumatic stress according to the DSM-5 criteria, the City-BiTS-S questionnaire obtained a Cronbach's $\alpha = .90$. Factor 1 (general symptoms) was generated with Cronbach's $\alpha = .89$; Factor 2 (birth-related symptoms) with Cronbach's $\alpha = .82$. The change in Cronbach's α when an item was removed from the scale and the inter-correlation between the items were both satisfactory, with internal consistency values of 0.87 to 0.89 (see Supplementary online material).

Rasch analysis of the 20 items in the two City-BiTS-S factors

Despite the EFA showed two factors, an initial analysis focused on data from all the 20 items of the four main groups of symptoms in the City-BiTS-S to examine the presence of factors or dimensions. Fit to the Rasch model was determined according to the statistics

normally used in this respect (Elder et al., 2017; Pallant & Tennant, 2007). First, we calculated the chi-square of the item-trait interaction, which is statistically non-significant when the items fit the Rasch model i.e. when all these items contribute to the construct and measure it invariantly. That initial Rasch analysis showed that the scale as a whole did not fit the model ($\chi^2 = 177.81$, $p < .001$) (see analysis 1 in Table 2). Sixteen of the items presented disordered thresholds, reflecting a malfunction of the four response categories. Many participants could not properly distinguish between categories, especially "Once" and "2-4 times". Before a second analysis, the thresholds were ordered for 15 items by merging the second and third response categories (options "Once" and "2-4 times"). Nevertheless, this approach failed to remedy the disordered thresholds of item #8, which had to be transformed into a dichotomous item by merging the response option "Not at all" with "Once", and that of "2-4 times" with "5 or more times".

The subsequent analysis also showed a lack of fit to the Rasch model (see analysis 2 in Table 2). Values of the residuals of each item were reviewed to determine their individual fit to the scale. Adequate fit values are typically accepted as the mean of the residual within the range ±2.50 and a non-significant individual chi-square value after Bonferroni adjustment. Two items showed both kinds of misfit (#3 *Flashbacks to the birth and/or reliving the experience*: +2.76, p < .00001; #8 *Not able to remember details of the birth*: +3.08, p < .00001). A mean residual value exceeding +2.5 suggests that the concept underlying the item does not belong to the construct of the whole set of items, and therefore the item should be removed.

After removing items #3 and #8, a third Rasch analysis indicated that the remaining 18 items fitted the model (see analysis 3 in Table 2). To determine the presence of underlying factors in the set of 18 items, dimensionality was examined by conducting a principal component analysis of the person residuals, which is the strictest proof of unidimensionality in a set of items. Details of this three-step analysis can be found in Tennant and Conaghan (2007). This finding confirms that a factorial structure of the construct for the PTSD symptoms underlies items in the City-BiTS-S. Accordingly, the following Rasch analyses were performed separately for the birth-related and general symptoms subscales found in the previous EFA and proposed by the authors of the questionnaire (Ayers et al., 2018). The percentage of t-tests outside the CI at 5% was larger than the criterion of 5% (Tennant & Conaghan, 2007), and therefore unidimensionality of the items could not be assumed (see analysis 3 in Table 2).

Table 2. *Hierarchy of the Items in each dimension of the City BiTS. Items are ordered starting with where it is more likely to endorse for participants*

	Location	Standard Error	Fit Residual	*p* of $\chi 2$
GENERAL SYMPTOMS				
17. Feeling tense and on edge	-1.00	0.11	-0.52	0.17
13. Feeling detached from other people	-0.96	0.11	-0.77	0.45
19. Problems concentrating	-0.61	0.12	0.65	0.33
18. Feeling jumpy or easily startled	-0.55	0.11	-0.75	0.26
15. Feeling irritable or aggressive	-0.43	0.15	-0.41	0.33
12. Lost interest in activities that were important to me	-0.42	0.12	-0.71	0.87
20. Not sleeping well because of things that are not due to the baby's sleep pattern	0.32	0.17	0.99	0.60
14. Not able to feel positive emotions (e.g., happy, excited)	0.61	0.14	0.87	0.55
11. Feeling negative about myself or thinking something awful will happen	0.81	0.18	-0.60	0.38
16. Feeling self-destructive or acting recklessly	2.24	0.32	-0.53	0.67
BIRTH-RELATED SYMPTOMS				
1. Recurrent unwanted memories of the birth (or parts of the birth) that you can't control.	-1.40	0.22	0.32	0.11
10. Feeling strong negative emotions about the birth (e.g., fear, anger, shame)	-1.12	0.23	-0.31	0.02
9. Blaming myself or others for what happened during the birth	-1.10	0.22	0.86	0.13
4. Getting upset when reminded of the birth	-0.55	0.25	-1.15	0.32
6. Trying to avoid thinking about the birth	-0.04	0.27	-0.72	0.46
5. Feeling tense or anxious when reminded of the birth	0.03	0.27	-0.28	0.16

2. Bad dreams or nightmares about the birth (or related to the birth)	1.81	0.31	1.18	0.53
7. Trying to avoid things that remind me of the birth (e.g., people, places, TV programs)	2.37	0.36	-0.85	0.53
3. Flashbacks to the birth and/or reliving the experience	Misfit			
8. Not able to remember details of the birth	Misfit			

Fit and unidimensionality of the subscales

The fourth Rasch analysis was conducted of data referring to the 10 items of the birth-related symptoms factor, obtaining the following results. A significant chi-square value for item-trait interaction indicated a lack of fit of the whole set of items to the model (see analysis 4 in Table 2). Item #3 revealed disordered thresholds, a situation that was remedied by merging the third and fourth categories ("2-4 times" and "5 or more times"). Despite this rescoring, the next analysis also revealed a lack of fit to the model (see analysis 5 in Table 2). Thus, item #8 (*Not able to remember details of the birth*) presented a misfit (residual = +2.73, p = .00003) and was removed from the following analysis. Subsequently, item #3 (*Flashbacks to the birth and/or reliving the experience*) also indicated a misfit (residual = +2.98, p = .003) and was removed. The seventh Rasch analysis indicated that the remaining eight birth-related symptoms fitted the model and constituted a unidimensional factor.

The eighth Rasch analysis showed that the 10 items of the general symptoms factor fitted the model. Regarding the overall item and person fit was reasonable in both subscales, with mean values for the residuals close to 0 and SD values close to 1 (Bond & Fox, 2015).

Reliability

In the Rasch model, the equivalent to the Cronbach's alpha coefficient of reliability is the Person Separation Index (PSI). In our analysis, the PSI was .84 for the birth-related symptoms subscale and .91 for the general symptoms subscale. These values indicate that the first subscale is able to classify persons into three groups according to their level of symptoms (e.g., low, moderate and high), and the second subscale, into four groups (Fisher, 1992).

Differential item functioning (DIF)

In a valid measurement tool, individuals with the same score are expected to have the same probability of endorsing an item, irrespective of any group membership. DIF occurs when groups of people respond differently to an individual item, despite having the same level of the latent construct (Tennant et al., 2004). DIF analysis is a way of determining whether items function differently across subgroups of people after controlling for ability level (Aryadoust et al., 2019).

City-BiTS-S subscales displayed no evidence of DIF within two groups based on the following factors: age (≤33, >33 years); length of gestation (<40, ≥40 weeks); type of birth (normal vaginal birth, instrumental delivery or C-section); pain relief (none, epidural), number of previous children (primiparity yes/no). No evidence of DIF was also found for groups based on the answers to the rest of items of the City-BiTS-S that were symptoms: belief that the mother or baby would be seriously injured (yes, no); or die (yes, no); duration of symptoms (≤1 month, >1 month) and symptoms impeding or preventing otherwise normal activities (yes, no).

Response category functioning

Rasch analysis determines whether participants are able to consistently discriminate between the successive responses categories defined in a rating scale (Andrich & Marais, 2019). In the City-BiTS-S, the response thresholds between categories were disordered in most of the items. Many participants did not endorse the response categories as expected regarding their level in the latent construct, especially “Once” and “2-4 times”. That indicated the four established response categories did not capture consistently the frequency of symptoms in the last week.

Item location or hierarchy

Rasch analysis sorts items of each factor on a logit scale according to respondents’ difficulty in choosing high values from among the response options. This difficulty represents the level of the latent construct expressed by the item. Table 3 shows the hierarchy of the items that fit the Rasch model in each subscale. The unit of the scale is the logit. The larger the logit value, the greater the level of construct that underlies the item. The hierarchy is always centered

on zero logits, representing the item of average difficulty for the scale (Tennant and Conaghan, 2007). The first items, i.e. those with the most negative logit values, are more easily endorsed with higher category responses (on the response scale from “None” to “5 or more times”).

Targeting of the subscales

Targeting refers to degree of matching between the level of the latent construct in the items and the respondents to the scale. Rasch analysis locates in the same *logit* scale both the items and the persons (see Figures 2 and 3). Therefore, targeting could be assessed by comparing the spreading and mean location of persons (top of figures) and item (bottom). The negative mean person location indicated that the average level of general (-3.758) and birth-related symptoms (-2.390) in the sample was lower than the average of construct underlying each subscale (Tennant and Conaghan, 2007). Both subscales poorly targeted the sample because participants had, on average, very low levels of PP-PTSD symptoms. Therefore, floor effects were found in both subcales, particularly birth-related symptoms. The existence of a floor effect is illustrated by the number of persons shown in the two first bars on the left of the logit scale in Figures 2 and 3.

Table 3. *Fit Statistics, Reliability and Unidimensionality Indices for the 20 Scalable Items on the City BiTS (N=207)*

Analysis number & Items included		Item-Trait Interaction $\chi 2$ (*p*)	Item disordered thresholds	Problematic items	Ítem fit residual Mean (*SD*)	Person fit residual Mean (*SD*)	Reliability PSI	Unidimensionality: % of significant *T*-tests 95% CI
1st	All 20 items	147.75 (<.001)	All except 3, 12, 13, 14, 17, 18, 19	N/A	N/A	N/A	N/A	N/A
2nd	All 20 items	109.64 (<.001)	None	#3 #8 (misfit)	N/A	N/A	N/A	N/A
3rd	18 items: #3 and #8 removed	36.39 (.97)	None	None	-.48 (.81)	-.39 (1.05)	.90	14.77% [11.3 to 18.3]
4th	10 Birth-symptoms items	71.61 (<.001)	#3	#8 (misfit)	N/A	N/A	N/A	N/A
5th	The 10 Birth-symptoms items	67.66 (<.001)	None	#8 (misfit)	N/A	N/A	N/A	N/A
6th	9 Birth-symptoms items (#8 removed)	37.93 (.003)	None	#3 (misfit)	N/A	N/A	N/A	N/A
7th	8 Birth-symptoms items (#3 removed)	25.70 (.06)	None	None	-.12 (.84)	-.17 (.76)	.84	0.0% [-.5 to .5]
8th	The 10 General-symptoms items	17.24 (.64)	None	None	-.18 (.72)	-.21 (.79)	.91	3.52% [-.8 to 6.5)

Figure 2. Item and person estimates on the latent continuum: 8 items of the City-BiTS-S birth-related symptoms (bottom) and the 207 participants of the sample (top).

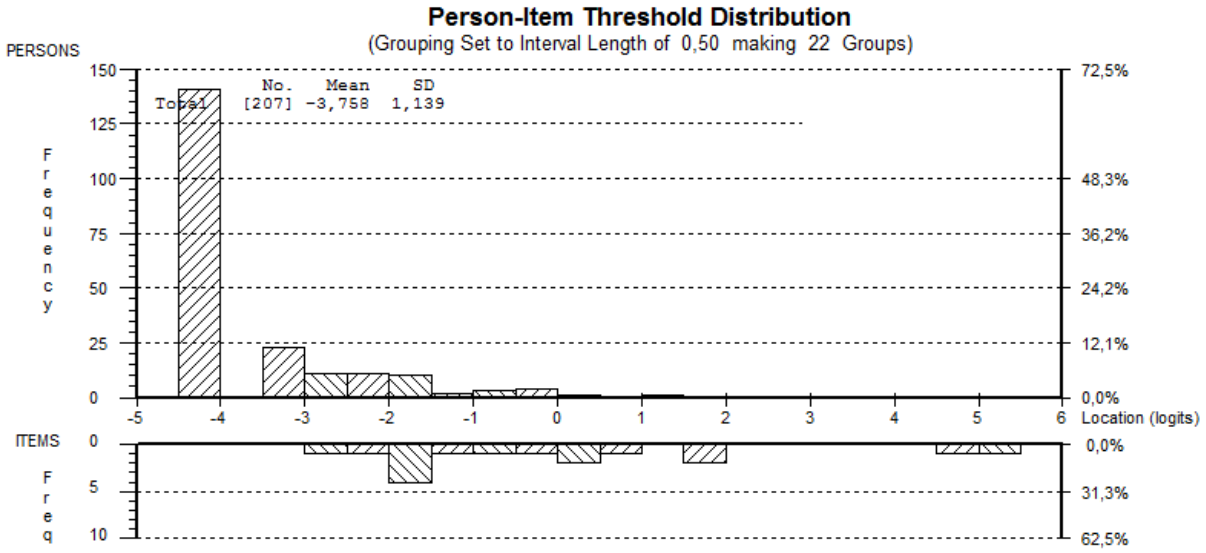
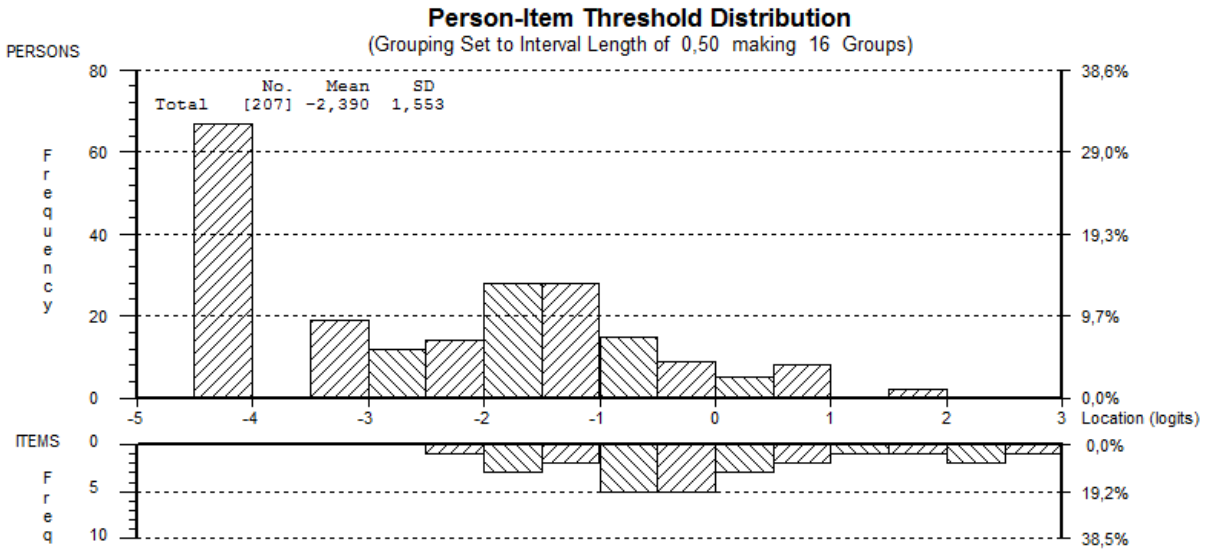


Figure 3. Item and person estimates on the latent continuum: 10 items of the City-BiTS-S general symptoms (bottom) and the 207 participants of the sample (top).



Discussion

The aim of this study was to determine the psychometric properties of the Spanish version of the City-BiTS questionnaire and add to knowledge about the scale's properties through Rasch analysis. Consistent with previous studies of the City-BiTS (Ayers et al., 2018; Handelzalts et al., 2018; Nakić Radoš et al., 2020), factor analysis of the Spanish-language version (City-BiTS-S) showed it has two factors of general symptoms and birth-related symptoms which account for almost 50% of the variance. As in previous studies, the total scale and subscales also had good internal consistency.

To our knowledge, this is the first study in which the Rasch measurement model has been applied to the City-BiTS using both sets of items as appropriate unidimensional subscales. Rasch analyses showed that after deleting two items both subscales were valid measurement instruments. This finding enables us to use the overall score of each subscale as lineal measure in parametric analysis. However, results highlight three changes which might improve the psychometric properties of the City BiTS. These were: (i) removing the item '*Not able to remember details of the birth';* (ii) removing or rewording the item '*Flashbacks to the birth and/or reliving the experience'*; and (iii) changing the response scale so participants can more clearly differentiate between response options, particularly in relation to experiencing symptoms 1 or 2-4 times in the last week.

Removing the item '*Not able to remember details of the birth'* means this DSM-5 symptom would not be assessed. Research on clinical samples of people with PTSD has argued that memories for traumatic events are fragmented or disorganised (Brewin 2014), although this has been contested (e.g. Berntsen, Willert & Rubin, 2003). The evidence for memories of birth being fragmented or disorganised is inconsistent. Some studies have found women with PTSD after birth report less coherent and more disorganised memories (Briddon, Slade, Isaac & Wrench, 2011; Foley, Crawley, Wilkie & Ayers 2014). However, other studies have found that women who experience birth trauma report highly coherent memories of birth (Crawley, Wilkie, Gamble, Creedy, Fenwick, Cockburn & Ayers, 2018) and women with severe PTSD have more coherent narratives than women with mild or no PTSD (Ayers Nakic Rados & Balouch 2015). Research using the City BiTS consistently finds that '*Not able to remember details of the* birth' loads poorly on the birth-related symptoms factor (i.e. a factor loading of <0.30) (Ayers et al., 2018; Handelzalts et al., 2018; Nakić Radoš et al., 2020).

This inconsistency may be due to sampling or to the nature of birth itself, as a highly salient and central life event to women. It may be that this symptom is specific to certain subgroups, such as women who have severe PTSD, or who dissociate during birth. It is also posible that memories for birth differ to other potentially traumatic events because the birth of a baby is a highly important life event that can involve positive and negative emotions, and that women and their partners are likely to remember. This needs further exploration and examination but it is possible this symptom is not common in women experiencing PP-PTSD. If this is the case and the poor loading is consistently replicated in other studies it might therefore be worth removing it from the scale.

The finding in relation to the item '*Flashbacks to the birth and/or reliving the experience'* is not consistently supported by previous studies. This item had good factor loadings for birth-related symptoms in the English and Croatian versions of the scale (Ayers et al., 2018; Nakić Radoš et al., 2020) but a poorer loading in the Hebrew version (Handelzalts et al., 2018). It is therefore possible this item works differently in different languages and cultures. Linguistically the term '*Flashbacks'* was kept in the Spanish version because it is an Anglicism commonly used by Spanish speakers. However, it might be that this term is not as widely understood or the nuances lost when using a non-native word. This item might also be more variable across samples because it is the only item in the scale that does not specify a negative direction. '*Flashbacks to the birth and/or reliving the experience'* does not imply negative feelings or thoughts and a sudden and very clear memory of birth could be positive, negative, or both. Finally, the item might be sensitive to PTSD severity in the sample. The current sample had low levels of PTSD symptoms with very few women endorsing this item. It is possible that in samples with greater levels of PTSD the item loads more coherently. Further research is therefore needed to understand more about how this item works across cultures, languages and samples.

Finally, Rasch analysis showed that the response options were not consistently interpreted, with 16 of 20 symptom items having disordered thresholds. This was particularly in relation to experiencing symptoms 1 or 2-4 times in the last week. In the current study merging these response options made thresholds ordered for all but one of these items. DSM diagnostic criteria do not specify the time frame within which people have to had experienced symptoms other than the symptoms have to be experienced for longer than one month. Other measures of PTSD use a variety of response options to indicate frequency of symptoms within the last week or month. These range from 2-point options (e.g. yes/no; Prins, Bovin, Kimerling,

Kaloupek, Marx, Pless Kaiser & Schnurr, 2015) to 5-point response options (e.g. 0 not at all - 4 extremely; Weathers, Litz, Keane, Palmieri, Marx & Schnurr, 2013). Results of the current study suggest a 3-point response option may be preferable. The response scale could therefore be changed to options such as *never*, *sometimes, often*. Alternatively, the time frame could be incorporated into the response options, such as *daily, weekly, monthly.* The performance of different response option scales therefore needs to be evaluated in future research. The diagnostic accuracy of modifications to items and response scales also needs to be established against the gold standard of structured clinical interviews.

Although our translation and validation of City-BiTS-S confirms it has good psychometric properties, certain limitations to the current study should be acknowledged. The City-BiTS questionnaire is designed to detect postpartum PTSD but our study was a non-clinical sample with low symptoms of PTSD and did not include women diagnosed with this disorder. This may have influenced our results, particularly in relation to the Rasch analysis where the number of women endorsing certain items was low. In future research it would be useful to examine the use of the scale in high-risk women who suffered a complicated or traumatic childbirth and may therefore be especially vulnerable to PP-PTSD. Furthermore, the participants in our study were recruited at three public health centres in southern Spain and were excluded if they had existing pathology. Further research is therefore needed to examine the City-BiTS-S performance with women from public and private health centres, different Spanish regions, and with existing pathology.

Implications and future directions for research

This research replicates the factor structure found in previous studies of the City-BiTS implying it is a robust scale with similar properties observed in different samples. The findings also have a number of potential methodological and theoretical implications. The first is the need to further examine the role of memory fragmentation/coherence in postpartum PTSD, particularly in clinical and non-clinical samples, or those who dissociate during birth to see if this symptom is a characteristic of PTSD following birth. The second is the need to recognise that responses to the item on flashbacks may be interpreted differently in different samples and cultures so this needs careful consideration in how best to interpret and measure it. Finally, the response scale options need to be examined to see if a 3-point scale improves the scales measurement and thresholds. The diagnostic accuracy of the City-BiTS (with and without modifications) also needs to be established.

In conclusion, the City-BiTS-S questionnaire has two main subscales that measure birth-related and general symptoms following childbirth. The construct validity and relativity of each subscale is appropriate. Several improvements to this instrument are suggested to improve its measurement properties. To date, the City-BiTS is the only validated instrument available for evaluating postpartum PTSD symptoms in accordance with current diagnostic criteria (DSM-5). It is therefore a useful resource both for researchers and for healthcare professionals.

References

Agius, A., Xuereb, R. B., Carrick-Sen, D., Sultana, R. & Rankin, J. (2016). The co-existence of depression, anxiety and post-traumatic stress symptoms in the perinatal period: a systematic review. *Midwifery, 36*, 70-79. doi:10.1016/j.midw.2016.02.013.

American Psychiatric Association (1994). *Diagnostic and Statistical Manual of Mental Disorders* (4th ed., DSM-IV). Washington, DC: American Psychiatric Publishing.

American Psychiatric Association (2013). *Diagnostic and Statistical Manual of Mental Disorders* (5th ed., DSM-5). Arlington, VA: American Psychiatric Publishing.

Andrich, D., & Marais, I. (2019). Constructing Instruments to Achieve Measurement. En D. Andrich & I. Marais (Eds.), *A Course in Rasch Measurement Theory: Measuring in the Educational, Social and Health Sciences* (pp. 13-28). Springer. https://doi.org/10.1007/978-981-13-7496-8_2

Aryadoust, V., Tan, H. A. H. & Ng, L. Y. (2019). A scientometric review of Rasch measurement: the rise and progress of a specialty. *Frontiers in Psychology*, *10*. https://doi.org/10.3389/fpsyg.2019.02197

Ayers, S., Bond, R., Bertullies, S. & Wijma, K. (2016). The aetiology of post-traumatic stress following childbirth: a meta-analysis and theoretical framework. *Psychological Medicine, 46*(6), 1121-1134. doi:10.1017/s0033291715002706.

Ayers, S., McKenzie-McHarg, K. & Eagle, A. (2007). Cognitive behavioural therapy for postnatal post-traumatic stress disorder: case studies. *Journal of Psychosomatic Obstetrics and Gynaecology, 28*(3), 177-184. doi:10.1080/01674820601142957.

Ayers, S., McKenzie-McHarg, K. & Slade, P. (2015). Post-traumatic stress disorder after birth. *Journal of Reproductive and Infant Psychology, 33*(3), 215-218. doi:10.1080/02646838.2015.1030250.

Ayers, S. & Pickering, A. D. (2001). Do women get posttraumatic stress disorder as a result of childbirth? A prospective study of incidence. *Birth, 28*(2), 111-118. doi:10.1046/j.1523-536x.2001.00111.x.

Ayers, S., Radoš, S.N. and Balouch, S. (2015). Narratives of traumatic birth: Quality and changes over time. *Psychological Trauma: Theory, Research, Practice, and Policy*, 7(3), doi: 10.1037/a0039044

Ayers, S., Wright, D. B. & Thornton, A. (2018). Development of a measure of postpartum PTSD: the City Birth Trauma Scale. *Frontiers in Psychiatry, 9*(409), 1-8. doi:10.3389/fpsyt.2018.00409.

Berntsen, D., Willert, M., & Rubin, D. C. (2003). Splintered memories or vivid landmarks? Qualities and organization of traumatic memories with and without PTSD. Applied Cognitive Psychology, 17, 675-693. doi:10.1002/acp.894

Bobes J., Calcedo-Barba, A., García, M., François, M., Rico-Villademoros, F., González, M. P., Bascarán, M. T., Bousoño, M. & Grupo Español de Trabajo para el Estudio del Trastorno por Estrés Postraumático (2000). Evaluación de las propiedades psicométricas de la versión española de cinco cuestionarios para la evaluación del trastorno de estrés postraumático. *Actas Españolas de Psiquiatría, 28*(4), 207-218. Retrieved from https://www.actaspsiquiatria.es

Brewin, C. R. (2014). Episodic memory, perceptual memory, and their interaction: Foundations for a theory of posttraumatic stress disorder. Psychological Bulletin, 140, 69-97. doi:10.1037/a0033722

Briddon, E., Slade, P., Isaac, C., & Wrench, I. (2011). How do memory processes relate to the development of posttraumatic stress symptoms following childbirth? Journal of Anxiety Disorders, 25, 1001-1007. doi:10.1016/j.janxdis.2011.06.008

Callahan, J. L., Borja, S. E. & Hynan, M. T. (2006). Modification of the perinatal PTSD questionnaire to enhance clinical utility. *Journal of Perinatology, 26*(9), 533-539. doi:10.1038/sj.jp.7211562.

Cano, S. J. & Hobart, J. C. (2011). The problem with health measurement. *Patient Preference and Adherence, 5*, 279–290. https://doi.org/10.2147/PPA.S14399

Caparros-Gonzalez, R. A., Perra, O., Alderdice, F., Lynn, F., Lobel, M., García-García, I. & Peralta-Ramírez, M. I. (2019). Psychometric validation of the Prenatal Distress Questionnaire (PDQ) in pregnant women in Spain. *Women & Health*, 1-16. doi:10.1080/03630242.2019.1584143.

Caparros-Gonzalez, R. A., Romero-Gonzalez, B., Quesada-Soto, J. M., Gonzalez-Perez, R., Marinas-Lirola, J. C. & Peralta-Ramírez, M. I. (2019). Maternal hair cortisol levels affect neonatal development among women conceiving with assisted reproductive technology. *Journal of Reproductive and Infant Psychology*, *37*(5), 480-498.

Carlson, K. D. & Herdman, A. O. (2012). Understanding the impact of convergent validity on research results. *Organizational Research Methods, 15*(1), 17-32. doi:10.1177/1094428110392383.

Consejería de Salud de la Junta de Andalucía (2017). *Documento de Salud de la Embarazada*. Retrieved on 24 May 2017, from https://www.juntadeandalucia.es/export/drupaljda/DSE2017JUNIO_completo.pdf

Davidson, J. R. T., Book, S. W., Colket, J. T., Tupler, L. A., Roth, S., David, D., Hertzberg, M., Mellman, T., Beckham, J. C., Smith, R. d., Davidson, R. M., Katz, R. & Feldman, M. E. (1997). Assessment of a new self-rating scale for post-traumatic stress disorder. *Psychological Medicine, 27*(1), 153-160. doi:10.1017/s0033291796004229.

Dekel, S., Stuebe, C. & Dishy, G. (2017). Childbirth induced posttraumatic stress syndrome: a systematic review of prevalence and risk factors. *Frontiers in Psychology, 8*(560), 1-10. doi:10.3389/fpsyg.2017.00560.

Elder, H., Czuba, K., Kersten, P., Caracuel, A. & McPherson, K. (2017). Te Waka Kuaka, Rasch analysis of a cultural assessment tool in traumatic brain injury in Māori. *F1000Research, 6*, 1034. https://doi.org/10.12688/f1000research.11500.1

Field, A. (2009). *Discovering statistics using SPSS*. London: Sage publications.

Fisher, W. Jr. (1992). Reliability statistics. *Rasch Measurement Transactions, 6*(3), 238.

Foley, S., Crawley, R. A., Wilkie, S., & Ayers, S. (2014). The Birth Memories and Recall Questionnaire (BirthMARQ): development and evaluation. BMC: Pregnancy and Childbirth, 14, 211. doi:10.1186/1471-2393-14-211

Garthus-Niegel, S., Ayers, S., Martini, J., von Soest, T. & Eberhard-Gran, M. (2017). The impact of postpartum post-traumatic stress disorder symptoms on child development: a population-based, 2-year follow-up study. *Psychological Medicine, 47*(1), 161-170. doi:10.1017/s003329171600235x.

Grekin, R. & O'Hara, M. W. (2014). Prevalence and risk factors of postpartum posttraumatic stress disorder: a meta-analysis. *Clinical Psychology Review, 34*(5), 389-401. doi:10.1016/j.cpr.2014.05.003.

Hairston, I. S., Handelzalts, J. E., Assis, C. & Kovo, M. (2018). Postpartum bonding difficulties and adult attachment styles: the mediating role of postpartum depression and childbirth-related PTSD. *Infant Mental Health Journal, 39*(2), 198-208. doi:10.1002/imhj.21695.

Handelzalts, J. E., Hairston, I. S. & Matatyahu, A. (2018). Construct validity and psychometric properties of the Hebrew version of the City Birth Trauma Scale. *Frontiers in Psychology, 9*(1726), 1-10. doi:10.3389/fpsyg.2018.01726.

Horesh, D., Lowe, S. R., Galea, S., Aiello, A. E., Uddin, M. & Koenen, K. C. (2017). An in-depth look into PTSD-depression comorbidity: A longitudinal study of chronically-exposed Detroit residents. *Journal of Affective Disorders, 208,* 653-661. doi:10.1016/j.jad.2016.08.053.

James, S. (2015). Women's experiences of symptoms of posttraumatic stress disorder (PTSD) after traumatic childbirth: a review and critical appraisal. *Archives of Women's Mental Health, 18*(6), 761-771. doi:10.1007/s00737-015-0560-x.

Kline, P. (2000). *A Psychometrics Primer*. Free Association Books, London: Free Association Books.

Lloret-Segura, S., Ferreres-Traver, A., Hernández-Baeza, A. & Tomás-Marco, I. (2014). El análisis factorial exploratorio de los ítems: una guía práctica, revisada y actualizada. *Anales de Psicología, 30*(3), 1151-1169. doi:10.6018/analesps.30.3.199361.

Martin, C. R. & Savage-McGlynn, E. (2013). A 'good practice' guide for the reporting of design and analysis for psychometric evaluation. *Journal of Reproductive and Infant Psychology, 31*(5), 449-455. doi:10.1080/02646838.2013.835036.

McKenzie-McHarg, K., Ayers, S., Ford, E., Horsch, A., Jomeen, J., Sawyer, A., Thomson, G. & Slade, P. (2015). Post-traumatic stress disorder following childbirth: an update of current issues and recommendations for future research. *Journal of Reproductive and Infant Psychology, 3*(3), 219-237. doi:10.1080/02646838.2015.1031646.

Nakić Radoš, S., Matijaš, M., Kuhar, L., Anđelinović, M. and Ayers, S. ORCID: 0000-0002-6153-2460 (2019). Measuring and conceptualizing PTSD following childbirth: Validation of the City Birth Trauma Scale. Psychological Trauma: Theory, Research, Practice, and Policy, doi: 10.1037/tra0000501.

Parfitt, Y. M. & Ayers, S. (2009). The effect of post-natal symptoms of posttraumatic stress and depression on the couple's relationship and parent-baby bond. *Journal of Reproductive and Infant Psychology, 27*(2), 127-142. doi:10.1080/02646830802350831.

Parfitt, Y. M., Pike, A. & Ayers, S. (2014). Infant developmental outcomes: A family systems perspective. *Infant and Child Development, 23*(4), 353-373. doi:10.1002/icd.1830.

Quinnell, F. A. & Hynan, M. T. (1999). Convergent and discriminant validity of the perinatal PTSD questionnaire (PPQ): a preliminary study. *Journal of Traumatic Stress, 12*(1), 193-199. doi:10.1023/a:1024714903950.

Prins, A., Bovin, M. J., Kimerling, R., Kaloupek, D. G., Marx, B. P., Pless Kaiser, A., & Schnurr, P. P. (2015). *The Primary Care PTSD Screen for DSM-5 (PC-PTSD-5).* [Measurement instrument].

Romero-González, B., Peralta-Ramírez, M. I., Caparrós-González, R. A., Cambil-Ledesma, A., Martin, C. J. H. & Martin, C. R. (2019). Spanish validation and factor structure of the Birth Satisfaction Scale-Revised (BSS-R). *Midwifery, 70,* 31-37. doi:10.1016/j.midw.2018.12.009.

Tennant, A., & Conaghan, P. G. (2007). The Rasch measurement model in rheumatology: What is it and why use it? When should it be applied, and what should one look for in

a Rasch paper? *Arthritis & Rheumatism, 57*(8), 1358-1362. https://doi.org/10.1002/art.23108

Tennant, A., McKenna, S., & Hagell, P. (2004). Application of Rasch Analysis in the Development and Application of Quality of Life Instruments. *Value in Health, 7*(s1), S22-S26. https://doi.org/10.1111/j.1524-4733.2004.7s106.x

Terol-Cantero, M. C., Cabrera-Perona, V. & Martín-Aragón, M. (2015). Revisión de estudios de la Escala de Ansiedad y Depresión Hospitalaria (HAD) en muestras españolas. *Anales de Psicología, 31*(2), 494-503. doi:10.6018/analesps.31.2.172701.

Weiss, D. & Marmar, C. (1997). The Impact of Event Scale – Revised. In J. P. Wilson & T. M. Keane (Eds), *Assessing Psychological Trauma and PTSD* (pp. 399-411). New York: Guilford Press.

Wijma, K., Söderquist, J. & Wijma, B. (1997). Posttraumatic stress disorder after childbirth: a cross sectional study. *Journal of Anxiety Disorders, 11*(6), 587-597. doi:10.1016/s0887-6185(97)00041-8.

World Medical Association (2013). World Medical Association Declaration of Helsinki. Ethical Principles for Medical Research Involving Human Subjects. *Journal of American Medical Association, 310*(20), 2191-2194. doi:10.1001/jama.2013.281053.

Weathers, F.W., Litz, B.T., Keane, T.M., Palmieri, P.A., Marx, B.P., & Schnurr, P.P. (2013). The PTSD Checklist for *DSM-5* (PCL-5). Scale available from the National Center for PTSD at www.ptsd.va.gov.

Yildiz, P. D., Ayers, S. & Phillips, L. (2017a). The prevalence of posttraumatic stress disorder in pregnancy and after birth: a systematic review and metaanalysis. *Journal of Affective Disorders, 208*, 634-645. doi:10.1016/j.jad.2016.10.009.

Yildiz, P. D., Ayers, S. & Phillips, L. (2017b). Factors associated with post-traumatic stress symptoms (PTSS) 4–6 weeks and 6 months after birth: A longitudinal population-based study. *Journal of Affective Disorders, 221*, 238-245. doi:10.1016/j.jad.2017.06.049.

Yonkers, K. A., Smith, M. V., Forray, A., Epperson, C. N., Costello, D., Lin, H. Q. & Belanger, K. (2014). Pregnant women with posttraumatic stress disorder and risk of

preterm birth. *JAMA Psychiatry*, *71*(8), 897-904. doi:10.1001/jamapsychiatry.2014.558.

Zigmond, A. S. & Snaith, R. P. (1983). The hospital anxiety and depression scale. *Acta Psychiatrica Scandinavica*, *67*(6), 361-370. doi:10.1037/t03589-000