Factor structure and psychometric properties of the Birth Satisfaction Scale–Revised in Portuguese postpartum women

Helena Moreira[1], Caroline J. Hollins Martin[2,] and Colin R. Martin[3]

[1]University of Coimbra, Faculty of Psychology and Educational Sciences, Center for Research in Neuropsychology and Cognitive-Behavioural Intervention

[2]School of Nursing, Midwifery and Social Care, Edinburgh Napier University

[3]Institute for Health and Wellbeing, University of Suffolk, UK

Author Note

We have no known conflict of interest to disclose.

Correspondence concerning this article should be addressed Helena Moreira, Center for Research in Neuropsychology and Cognitive-Behavioral Intervention, Faculty of Psychology and Educational Sciences, University of Coimbra, Rua do Colégio Novo, 3030-115 Coimbra, Portugal. E-mail: Helena.tcmoreira@gmail.com

Acknowledgements: This paper is part of the COST Action Riseup-PPD CA18138 and was supported by COST under COST Action Riseup-PPD CA18138. COST (European Cooperation in Science and Technology) is a funding agency for research and innovation networks. Our Actions help connect research initiatives across Europe and enable scientists to grow their ideas by sharing them with their peers. This boosts their research, career and innovation. www.cost.eu

Availability of the BSS-R

The BSS-R is free to use for clinical and research purposes but requires permission. Contact Professor Hollins Martin c.hollinsmartin@napier.ac.uk for permission to use and the dedicated dedicated BSS-R website www.bss-r.co.uk for more information on the measure.

Abstract

Objective. This study aims to develop a Portuguese-language version of the Birth Satisfaction Scale-Revised (BSS-R) for clinical and research use in Portugal (PT-BSS-R). The factor structure, its invariance across depression levels, and the internal consistency of the PT-BSS-R were analysed. **Method.** A sample of 1373 Portuguese-speaking postpartum women completed a sociodemographic and clinical form, the PT-BSS-R, and the Edinburgh Postnatal Depression Scale through an online survey tool. Four competing models were tested through confirmatory factor analyses and bifactor model-based psychometric indices were calculated. The internal consistency and the divergent/convergent validity were analysed. **Results.** The psychometric properties of the PT-BSS-R were found to be generally excellent. Both the original correlated three-factor model and the bifactor model exhibited a good fit to the data. Results from the bifactor model support the use of both the BSS-R total score and the subscale scores. Exemplary invariance findings were observed between groups stratified on the basis of a depression screening. The PT-BSS-R has demonstrated both divergent and convergent validity as well as internal consistency. **Conclusion.** The PT-BSS-R is a psychometrically robust measure of birth satisfaction appropriate for clinical and research use in Portugal.

Keywords: Birth satisfaction, BSS-R, confirmatory factor analysis, bifactor, psychometrics

Introduction

Birth satisfaction is a multidimensional construct that has been conceptualised within a model comprising distinct but related domains of stress, maternal traits, and quality of care appraised from the woman's perspective (Hollins Martin & Martin, 2014). To assess this construct, Hollins Martin and Fleming (2011) developed the Birth Satisfaction Scale (BSS). Thirty items were developed to assess three distinct but correlated domains: quality of care provision (QC; assessing birth environment, support, home assessment, and relationships with health care professionals), women's personal attributes (WA; assessing the ability to cope during labour, feeling in control, childbirth preparation, and relationship with baby), and stress experienced during labour (SE; assessing distress, obstetric injuries, receiving sufficient medical care, obstetric intervention, pain, long labour, and baby's health).

A short-form version of the BSS, the Birth Satisfaction Scale-Revised (BSS-R), was later developed by Hollins Martin and Martin (2014). The BSS-R has only 10 items to measure the same three domains and proved to be a robust, valid, and reliable instrument for measuring postnatal women's satisfaction with their birth experience. Although the BSS-R has significantly fewer items than the original BSS, Cronbach's alpha coefficient values were shown to be mostly adequate and the original correlated three-factor model demonstrated a good fit to the data (Hollins Martin & Martin, 2014). Given its psychometric robustness and its easy and quick administration and scoring, the BSS-R has been extensively used in research and clinical settings. In addition, it has recently been recommended as a measure of choice for measuring birth satisfaction by the *International Consortium of Health Outcomes Measurement* (ICHOM), within the *Pregnancy and Childbirth Standard Set* (The International Consortium for Health Outcome Measurement, 2016).

In addition to the subscales scores, it is possible to compute a total score as an overall index of birth satisfaction (Hollins Martin & Martin, 2014). In the original validation study,

the authors established a three-factor correlated model as the measurement model of the BSS-R. However, a second-order model with a second-order factor measuring the experience of childbirth and the three first-order factors corresponding to the three subscales has also shown good fit to the data, although similar to the correlated three-factor model. To further explore whether BSS-R should be used as a multidimensional or unidimensional scale, Martin et al. (2018) analysed a bifactor model. With a bifactor model it is possible to test how much of the variance for all items is explained by a general factor and how much of the remaining variance is explained by domain-specific factors (Chen et al., 2006). In the study of Martin et al. (2018), both the correlated three-factor model and the bifactor model demonstrated an excellent fit to data. The pattern of factor loadings in the bifactor model supported the relative independence of the QC subscale, in contrast to the WA and SE subscales, which were better explained by a general factor of experience of childbirth. Martin et al. (2018) have also tested a two-factor correlated model with a QC factor and a combined factor of WA and SE, which has also revealed an excellent fit to data. Based on these results, the authors concluded that is possible to compute a total score of satisfaction with the experience of childbirth, whereas the BSS-R domain-specific subscales, particularly the QC subscale, can also be used as independent measures of distinct aspects of birth satisfaction.

A critical aspect of any valid and reliable instrument is its performance across different groups of participants distinguished by a particular and relevant clinical characteristic. Ideally, the instrument should measure the same construct in the same way in each group, even within the context of scores being anticipated to be different between groups. Previous studies of the BSS-R have confirmed the tool to have these extremely desirable measurement characteristics. For example, a large study of the US version of the BSS-R found that groups distinguished by having a planned or unplanned pregnancy responded to the measure in an equivalent way in terms of the measurement characteristics of

the BSS-R (Martin et al., 2017). Extrapolating this aspect of measurement equivalence to mental health, specifically depression, is important because postnatal depression is not only common (Lyubenova et al., 2021), but may be related to birth experience (Rosseland et al., 2020). It is thus important to know that women who may experience postnatal depression are responding to assessments of their birth experience in the same way as those who are not depressed to rule out any systemic confound or inherent measurement bias in the instrument as an artefact of affective status. A robust approach to undertake this appraisal is by invariance analysis (Byrne, 2010), although this analysis has not yet been performed with the BSS-R in relation to postpartum depression.

The Present Study

Extensively translated and validated (Barbosa-Leiker et al., 2015; Göncü Serhatlioglu et al., 2018; Jefford et al., 2018; Mortazavi et al., 2021; Nespoli et al., 2021; Romero-Gonzalez et al., 2019; Skvirsky et al., 2020; Vardavaki et al., 2015), the BSS-R has gathered attraction and utility internationally in part due to ease of use and scoring, acceptability to women, and psychometric performance (Alfaro Blazquez et al., 2017). A Portuguese language version of the BSS-R has recently been developed (Ferrari et al., 2022), however it was translated into Brazilian Portuguese because it was intended for usage in Brazil. In addition, the Brazilian Portuguese version of the BSS-R was fundamentally a translation study and the factor structure of the scale was not examined. Therefore, the present study aims to develop and validate a European Portuguese-language version of the BSS-R (PT-BSS-R) for use within Portugal and the context of the Portuguese maternity care system. To validate the PT-BSS-R, the study had the following specific objectives: 1) examine the factor structure of the PT-BSS-R; 2) evaluate its invariance characteristics by comparison of groups screened positive or negative for postpartum depression; 3) evaluate the internal consistency of PT-BSS-R subscales and total score; 4) determine the divergent validity of the PT-BSS-R

utilising the approach taken in the original BSS-R development study (i.e., by examining the correlations between PT-BSS-R and mothers' age); and 5) evaluate its convergent validity by examination of the relationship with a single item assessing women's satisfaction with birth. Significantly positive correlations were predicted between scores on the question assessing birth satisfaction and PT-BSS-R subscales and total score.

Materials and Methods

Participants

Participants in this study were 1373 mothers of infants up to 10 months. The complete sample characterization is presented in Table 1.

Procedure

Translation Process

The first step of the forward-backward translation procedure was translating the BSS-R items from English to Portuguese. Two Portuguese researchers fluent in English and experts in perinatal mental health independently translated the items. After translating the items, the two authors met to compare the translations and analyse any differences found. The second step consisted of an expert panel meeting, which involved the two researchers who translated the questionnaire, a senior clinical psychologist from a maternity department of a public hospital, and two senior researchers in the field of reproductive mental health. The panel agreed with the translation of all items, with the exception of item 1 ("I came through childbirth virtually unscathed") and item 8 ("I felt out of control during my birth experience"). The translation of these two items was thoroughly discussed until the panel reached consensus. The third step consisted in the back translation of the preliminary version of the PT-BSS-R into English by an independent researcher who was fluent in English and who was not familiar with the questionnaire. The last step consisted in the comparison of the

original and back-translated versions of the scale to confirm that no item in the Portuguese version had changed the original meaning.

Data collection

Participants were eligible for inclusion in the study if they were 18 years or older, Portuguese speaking, and had a child between 0 and 10 months. The sample was collected online, through an online survey tool (LimeSurvey®). The link to the survey was shared on social networks, both through unpaid cross-posting and through paid boosting campaigns. The survey included a sociodemographic and obstetric form and self-report questionnaires. The first page of the survey included information on the study objectives and the ethical aspects underlying the study. In the second page of the survey, participants were asked to provide their informed consent by selecting the option "I understand and accept the conditions of the study". After providing consent, participants were granted access to the survey. Authorization for the sample collection was obtained from the Ethics Committee of the Faculty of Psychology and Educational Sciences of the University of Coimbra.

Measures

Birth Satisfaction Scale-Revised

The Birth Satisfaction Scale-Revised (BSS-R; Hollins Martin & Martin, 2014) is a 10-item scale designed to measure birth satisfaction. The BSS-R has three subscales: (1) Stress Experienced During Labour (e.g., "I found giving birth a distressing experience"); (2) Quality of Care Provision (e.g., "The staff communicated well with me during labour"); and (3) Women's Personal Attributes (e.g., "I felt very anxious during my labour and birth"). Items are answered on a 5-point Likert scale that ranges from 0 (*strongly disagree*) to 4 (*strongly agree*). After reverse coding negative items, the total score and the subscale scores can be computed by summing all the responses or the responses to each of the three subscales, so that higher scores indicate higher levels of birth satisfaction.

Edinburgh Postnatal Depression Scale

Depression was assessed using the Portuguese version (Areias et al., 1996) of the Edinburgh Postnatal Depression Scale (EPDS; Cox et al., 1987). The EPDS is a 10-item screening tool designed to measure the presence of depressive symptoms (e.g., sadness, tearfulness) in the previous seven days (e.g., "I have been able to laugh and see the funny side of things"). Items are answered on a 4-point Likert scale (from 0 to 3). The total score can be computed by summing all the responses, so that higher scores indicate higher levels of depressive symptoms. According to the scoring recommendations of the Portuguese EPDS, a score of 10 or above indicates the probable presence of clinically relevant symptoms. In the present study, Cronbach's alpha value was .88.

Global Birth Satisfaction

A single question 'How satisfied are you with the delivery you had?', measured on a 0 to 10 scale anchored 0 = *totally unsatisfied*, 5 = *moderately satisfied*, and 10 = *very satisfied*, was used to assess the convergent validity of PT-BSS-R.

Statistical Analysis

The data analyses were conducted using the Statistical Package for the Social Sciences (SPSS, version 27.0; IBM SPSS, Chicago, IL) and the R software. Confirmatory factor analysis (CFA) was used to evaluate the factor structure of the BSS-R. The parametric assumptions which underlie this approach to model evaluation (Brown, 2015) are contingent on normally distributed scale items and an absence of outliers. BSS-R scale items were therefore scrutinized for excessive skew and kurtosis. Multivariate outliers were also identified and removed (Kline, 2000). Models evaluated were (i) a single-factor model (Models 1 and 5); (ii) the tri-dimensional measurement model of the BSS-R (Hollins Martin & Martin, 2014) comprising three correlated factors (SE, WA, and QC; Models 2 and 6); (iii)

a two-factor correlated model comprising two correlated factors (QC and SE/WA; Models 3 and 7); and (iv) a bifactor model comprising uncorrelated SE, WA, and QC and a general factor of birth experience (Martin et al., 2018; Models 4 and 8). Each model was tested with and without outliers. Maximum-likelihood estimation (Brown, 2015; Kline, 2011) was used to evaluate models. Model fit adequacy was determined using the CFI Bentler, 1990), the RMSEA Steiger & Lind, 1980), and the SRMR Hu & Bentler, 1999). Cut-off values of >0.90 (CFI), < 0.08 (RMSEA) and < 0.06 (SRMR) were used as reference threshold values for model acceptability.

Bifactor model-based psychometric indices were computed: the explained common variance (ECV; Sijtsma, 2009; Ten Berge & Sočan, 2004), the percentage of uncontaminated correlations (PUC; Bonifay et al., 2015; Reise, Scheines, et al., 2013), and the omega (ω) reliability coefficients (McDonald, 1999; Reise, 2012; Zinbarg et al., 2005). The ECV is an index of the degree of unidimensionality and assesses the relative strength of the general factor (Rodriguez et al., 2016a, 2016b), with higher values indicating a strong general factor and suggesting unidimensionality. Higher values of PUC suggest that the parameter estimates in a unidimensional model are less likely to be biased. According to Rodriguez et al. (2016a), "when ECV is > .70 and PUC > .70 relative bias will be slight and the common variance can be regarded as essentially unidimensional" (p. 232). However, when PUC values are < .80 but general ECV values are > .60 and the omega hierarchical (ωH) for the total score is > .70, it can be considered that the presence of multidimensionality in not severe enough to disqualify the interpretation of the measure as essentially unidimensional (Reise, Scheines, et al., 2013).

The ω index is a factor-analytic model-based index of internal reliability, with higher values suggesting a highly reliable multidimensional composite (Rodriguez et al., 2016a). While ω provides an estimate of the amount of the score variance due to all common factors, ωH and the omega hierarchical subscale (ωHS) estimate the amount of the score variance due

to a single common factor (i.e., the general or specific factor). Higher values of ωH indicate that the general factor is the dominant source of systematic variance, with an ωH greater than .50 being suggestive of factor strength (Hancock & Mueller, 2001; Reise, Bonifay, et al., 2013). According to Rodriguez et al. (2016a), an ωH of .80 or above indicates that total scores can be considered essentially unidimensional.

The three-factor measurement model was explored and submitted to measurement invariance analysis based on EPDS screen status. The EPDS cut-off score of 10 was used for group stratification to distinguish between non-depressed and depressed caseness classification. To establish good fit to the measurement model through CFA, a series of increasingly restrictive models were evaluated to determine measurement equivalence at each level between groups. CFI values were used for model comparison, with differences values of ≤ 0.01 indicating measurement invariance between models (Cheung & Rensvold, 2002).

Internal consistency of the PT-BSS-R subscales and total scale was determined using Cronbach's alpha (Cronbach, 1951). Alpha of .70 or greater is regarded as acceptable (Kline, 2000). The two item WA subscale was also assessed using the inter-item correlation (Pearson's *r*) with threshold values of .15 to .50 deemed acceptable (Clark & Watson, 1995).

Consistent with the original UK instrument development study (Hollins Martin & Martin, 2014) Pearson's *r* correlation coefficients were calculated between PT-BSS-R subscales and participant age. Given that medium to large sample sizes, as in the current study, can result in statistically significant *p* values when absolute values of *r* are small (Aggarwal & Ranganathan, 2016), it is predicted that effect size values for *r* will be small using Cohen's (1988) convention of .10. Convergent validity was assessed by correlating PT-BSS-R subscale and total scores with EPDS scores and with the single question formulated to evaluate women's global birth satisfaction. Using Cohen's criterion (1988) it is anticipated that the effect size for convergent validity testing will be large (.50).

Results

Preliminary Analyses

One thousand, three hundred and forty nine women took part in the study. One participant had 10% missing PT-BSS-R data and was excluded. Distances from the centroid was calculated on PT-BSS-R data and Mahalanobis distances revealed 23 outliers which were removed and thus provided a final dataset for full psychometric analysis of 1373 participants. The descriptive and distributional characteristics of PT-BSS-R items, subscales and total scores are summarized in Table 2 and reveal no evidence of excessive skew or kurtosis.

Confirmatory Factor Analysis

As presented in Table 3, the single-factor models (1 and 5) revealed a poor fit to data. The three-factor models (2 and 6) demonstrated a very good fit to data. However, it was noted that the correlation between SE and WA factors was high (~1) for both models; therefore, two-factor models (3 and 7) were run with SE and WA items specified within a single factor. These models also revealed a very good fit to data, with the chi-square differences test revealing no significant difference in fit between three-factor and two-factor models (model 2 *vs*. model 3, $\Delta\chi^2 = 0.30$, $\Delta df = 2$, $p = 0.86$; model 6 *vs*. model 7, $\Delta\chi^2 = 2.49$, $\Delta df = 2$, $p = 0.29$). Due to the potential for Heywood cases in view of the almost unity correlation between SE and WA factors, the bifactor models were run based on the two-factor models above, specifying a general factor and two uncorrelated factors of QC and a combined SE/WA items factor. These bifactor models (4 and 8) revealed excellent fit to data across all fit indices. Model 8 was examined further in terms of ancillary psychometric indices (Rodriguez et al., 2016c) (see Table 4).

The ECV of the general factor value of .69 was insufficient to determine that the BSS-R was sufficiently unidimensional (Reise et al., 2010), since threshold values of >.85 are required (Stucky & Edelen, 2015). However, using the criterion of PUC <.80, ECV >.60 and

(ωH) >.70 (Rodriguez et al., 2016a), values of .53, .69, and .76 respectively, suggested that despite multidimensionality, the measure may still be reliably interpreted as a total score construct. Additionally, item-level ECV examination revealed an average ECV of .73 (range .44 - .99) thus indicating both a multidimensional model in concert with a substantive general factor. Scrutiny of the item-factor loadings of the bifactor model (Figure 1) also support the role of a general factor in explaining substantive model variance while the specific factors also contribute to unique variance within the model.

Invariance Analysis

Using the EPDS cut-off score threshold of ten, 516 participants screened positive (37.6%). The tri-dimensional measurement model of the BSS-R was run separately in each group and were both found to offer an excellent fit to data. The configural model was then run and found to offer an excellent fit to data. Using the CFI difference criteria outlined earlier, metric invariance was also found. The scalar model again revealed no evidence of non-equivalance between groups and, finally, the strict invariance model was run and invariance was found at this greatest level of measurement model restriction (Table 5).

PT-BSS-R Subscale and Total Score Correlations

Correlations between all subscales and subscale-total score combinations were all statistically significant ($p < .01$) with the common variance explained ranging from 24% to 83%. Except for the relationship between the WA subscale and the total BSS-R scale score, the correlational comparison method of Diedenhofen and Musch (2015) revealed correlations to be statistically significantly higher ($p < .05$) compared to the original UK BSS-R development study (Table 6).

Internal Consistency

Cronbach alpha's greater than .70 were observed for the PT-BSS-R total scale and the WA and QC subscales (Table 7). Internal consistency of the WA subscale was .69. The inter-

item correlation of the two WA items was $r = .52$ (95%CI, 0.48 - 0.56), $p < .001$. Cronbach alpha's for the SE subscale and the total score were observed to be significantly higher than those of the original UK version (Hollins Martin & Martin, 2014) when compared using the method of Diedenhofen and Musch (2016).

Divergent Validity

Statistically significant correlations were observed between WA and QC subscales, the PT-BSS-R total score and participant age (respectively $r = .07$, $p = .007$; $r = .08$, $p = .002$; and $r = .07$, $p = .007$). No significant correlations were observed between the SE subscale and participant age ($r = .04$, $p = .13$). All r values were observed to be $< .10$.

Convergent Validity

Correlations between PT-BSS-R total and subscale scores and the single item of global birth satisfaction were all statistically significant ($p < .001$) with r values indicative of large effect sizes ($> .50$) (see Table 8).

Discussion

The current investigation has yielded robust findings for a valid and reliable translation of the BSS-R into Portuguese-language version of the tool, faithful to the conceptual and measurement framework of the original. Appraisal of individual item distributional characteristics of the measure reveals satisfactory characteristics in terms of data normality, a finding consistent with many previous psychometric studies of the BSS-R (Jefford et al., 2018; Romero-Gonzalez et al., 2019; Skvirsky et al., 2020). Consistent with previous investigations, maximum-likelihood's model estimation was able to be used to undertake the CFAs without modification, an observation giving confidence to the exemplary distributional aspects of the tool.

It was noted that the tri-dimensional measurement model of the BSS-R offered an excellent fit to the data although the correlation between SE and WA factors was very high. A two-factor correlated model combining SE and WA items into a single factor also revealed an excellent fit to data but was not statistically superior to the three-factor model. A pertinent question then is should SE and WA items be combined into a single scale? A compelling argument against this could be made on the basis of the theoretical grounds underpinning the tri-dimensional measurement model of the tool, thus the measurement model itself being originally informed by themes identified from an extensive review of the literature prior to scale development (Hollins Martin and Martin, 2014). However, inspection of the magnitude of the correlation (.74) between SE and WA subscales offers additional insights. This degree of correlation indicates that these subscales share approximately 55% of common variance, which means that nearly half of the variance is unique to each scale, providing a statistically convincing rationale for the distinctiveness of each subscale. Given the theoretical premise, the excellent fit of the tri-dimensional measurement model and each subscale's unique variance, the preference for the use of the measure as comprising three subscales is convincing.

Tangential to this are the findings from the bifactor model evaluation. This model was found to offer an excellent fit to data and necessarily had to be run as a model comprising a general factor and two specific factors (SE and WA combined items and QC items) due to the potential of Heywood cases in relation to the high correlation between SE and WA factors. The additional psychometric indices revealed the BSS-R to be multidimensional but with a strong general factor. This should not be surprising as in the original BSS-R development study (Hollins Martin & Martin, 2014), evidence was found for good fit to a hierarchical model with a second-order factor of childbearing, a similar finding observed in the US version (Barbosa-Leiker et al., 2015). Our findings are consistent with this notion of the

general factor representing a broad domain of childbearing satisfaction within the context of a fundamentally multidimensional tool. The most parsimonious interpretation of the findings from the bifactor analysis would thus be consistent with Martin et al. (2018), that there co-exists sufficient unidimensionality and multidimensionality within the measure to support the use of both the total score, as recommended in international guidelines (The International Consortium for Health Outcome Measurement, 2016), and the subscale scores which offer the opportunity to explore additional theoretically-relevant research findings (e.g., Burduli et al., 2017; Martin et al., 2017; Romero-Gonzalez et al., 2019).

This inherent flexibility in scoring approaches engenders the BSS-R with applicability for fundamental outcome monitoring where a robust metric of birth experience is required, and a multidimensional scaled tool to investigate more complex clinical research questions. Further, the exemplar invariance characteristics found in the current study extend previous findings (Martin et al., 2017) regarding the measurement equivalence of the tool to an important area of mental health (depression). These results afford confidence in the robustness of BSS-R score findings across important subgroups of women against potential systemic bias as function of affective status. It is also important to note that across all models the presence of outliers made little difference in both model fit characteristics and interpretation compared to non-outlier data. This is a particularly reassuring finding in terms of a tool designed for clinical use, where an outlier will inevitably and periodically be part of the participant population.

Correlations between PT-BSS-R subscales and the total score were all highly statistically significant, a finding consistent with previous studies (Jefford et al., 2018; Romero-Gonzalez et al., 2019). Comparison with the original UK BSS-R development study revealed the majority (5 of 6) of correlation combinations to be significantly greater in terms of absolute *r* values. Clearly, in relation to significant correlations between subscales, this

observation is entirely consistent and to be anticipated in relation to the tri-dimensional measurement model of the BSS-R. The internal consistency observations were also observed to be acceptable and generally very good, with Cronbach's alpha of the SE subscale and the total score being significantly higher than in the original UK BSS-R development study (Hollins Martin & Martin, 2014). The two-item WA subscale was noted to be slightly above the criteria of Clark and Watson (1995) but we note that the shared common variance of these items at a inter-item correlation of 0.52 is modest at 27%, suggesting each item uniquely contributes to this subscale.

The divergent validity findings were as predicted with observations of correlations between PT-BSS-R subscale and total scores and participant age having *r* values < .10 and generally negligible to small effect sizes. Convergent validity findings in contrast and as predicted for correlations between PT-BSS-R subscale and total scores and the satisfaction question yielded *r* values all greater than the large effect size threshold.

Strengths and Limitations

This study has some limitations that should be noted. First, because the sample was recruited online, its representativeness may be compromised, as online recruitment is frequently associated with self-selection bias (i.e., mothers who participate in an online study tend to be more interested in the study themes and to be more motivated to complete questionnaires). Second, most mothers were living with the baby's father, had a college or graduate degree, had a term pregnancy and an assisted/intervention delivery in a public hospital. Therefore, caution is needed in generalizing the results to all mothers, particularly to those with lower educational levels and in different family circumstances, as well as to those with a preterm pregnancy and with different birth contexts. Third, because there was only one evaluation moment, the test-retest reliability of the PT-BSS-R could not be established. Finally, we did not examine the issue of any influence of time since birth on factor structure.

We note that there is a large variability in the time since birth in BSS-R validation studies for example, less than 10 days postpartum (Hollins Martin and Martin, 2014) to up to five years postpartum (Emmens et al., 2021), though the measurement model appears remarkably consistent between studies. However, we would suggest future studies evaluate any impact of time since birth on the BSS-R measurement model to determine unequivocally any influence of this temporal domain.

Despite these limitations, this study makes a significant addition to the analysis of the factor structure and psychometric properties of the BSS-R. Although a previous Brazilian Portuguese version has been developed (Ferrari et al., 2022), its factor structure was not examined. Additionally, because Brazilian Portuguese and European Portuguese differ in several ways, the PT-BSS-R produced a version that was quite different from the Brazilian one. Benefitting from a large sample size, the current study developed and validated a version of the BSS-R specifically contextualised for use in Portugal. The PT-BSS-R was found to have excellent psychometric qualities across all parameters desirable for the determination of a valid and reliable measure. In summary, the current study has furnished a valid and reliable Portuguese version of the BSS-R appropriate for clinical and research use.

Declaration of interest statement

The authors report no conflict of interest.

References

Aggarwal, R., & Ranganathan, P. (2016). Common pitfalls in statistical analysis: The use of correlation techniques. *Perspectives in Clinical Research*, *7*(4), 187-190. https://doi.org/10.4103/2229-3485.192046

Alfaro Blazquez, R., Corchon, S., & Ferrer Ferrandiz, E. (2017). Validity of instruments for measuring the satisfaction of a woman and her partner with care received during labour and childbirth: Systematic review. *Midwifery*, *55*, 103-112. https://doi.org/10.1016/j.midw.2017.09.014

Areias, M. E., Kumar, R., Barros, H., & Figueiredo, E. (1996). Comparative incidence of depression in women and men, during pregnancy and after childbirth. Validation of the Edinburgh Postnatal Depression Scale in Portuguese mothers. *British Journal of Psychiatry*, *169*(1), 30-35. https://doi.org/10.1192/bjp.169.1.30

Barbosa-Leiker, C., Fleming, S., Hollins Martin, C. J., & Martin, C. R. (2015). Psychometric properties of the Birth Satisfaction Scale-Revised (BSS-R) for US mothers. *Journal of Reproductive and Infant Psychology*, *33*(5), 504-511. https://doi.org/10.1080/02646838.2015.1024211

Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin*, *107*(2), 238-246. http://www.ncbi.nlm.nih.gov/pubmed/2320703

Bonifay, W. E., Reise, S. P., Scheines, R., & Meijer, R. R. (2015). When are multidimensional data unidimensional enough for structural equation modeling? An evaluation of the DETECT multidimensionality index. *Structural Equation Modeling*, *22*(4), 504-516. https://doi.org/10.1080/10705511.2014.938596

Brown, T. (2015). *Confirmatory Factor Analysis for Applied Research* (2nd ed.). Guilford Press.

Burduli, E., Barbosa-Leiker, C., Fleming, S., Hollins Martin, C. J., & Martin, C. R. (2017). Cross-cultural invariance of the Birth Satisfaction Scale-Revised (BSS-R): Comparing UK and US samples. *Journal of Reproductive and Infant Psychology*, *35*(3), 248-260. https://doi.org/10.1080/02646838.2017.1310374

Byrne, B. M. (2010). *Structural Equation Modeling with AMOS: Basic Concepts, Applications and Programming* (2nd ed.). Routledge/Taylor and Francis Group.

Chen, F. F., West, S. G., & Sousa, K. H. (2006). A comparison of bifactor and second-order models of quality of life. *Multivariate Behavioral Research*, *41*(2), 189–225.

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, *9*(2), 233-255. https://doi.org/10.1207/S15328007SEM0902_5

Clark, L. A., & Watson, D. (1995). Constructing validity: Basic issues in objective scale development. *Psychological Assessment*, *7*(3), 309-319.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum Associates.

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://www.ncbi.nlm.nih.gov/pubmed/3651732

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, *16*(3), 297–334.

Diedenhofen, B., & Musch, J. (2015). cocor: A comprehensive solution for the statistical comparison of correlations. *PloS One*, *10*(4), e0121945. https://doi.org/ doi:10.1371/journal.pone.0121945

Diedenhofen, B., & Musch, J. (2016). cocron: A web interface and R package for the statistical comparison of Cronbach's alpha coefficients. *International Journal of Internet Science*, *11*(1), 51–60.

Emmens, B., Hollins Martin, C.J. & Martin, C.R. (2021). Translation and validation of the Dutch version of the Birth Satisfaction Scale-Revised (BSS-R). *Journal of Reproductive and Infant Psychology,* DOI: 10.1080/02646838.2021.1979200

Ferrari, R. B., Martin, C., Hollins Martin, C., de Souza, F. G., Clini, J. V., Onofre, L. B. O., & Diniz Zanetti, M. R. (2022). Translation of the UK-Birth-Satisfaction-Scale-Revised (BSS-R) into Brazilian (Portuguese) and description of initial measurement properties. *Journal of Maternal-Fetal and Neonatal Medicine*, *35*(25), 6373-6379. https://doi.org/10.1080/14767058.2021.1913579

Göncü Serhatlioglu, S., Karahan, N., Hollins Martin, C. J., & Martin, C. R. (2018). Construct and content validity of the Turkish Birth Satisfaction Scale - Revised (T-BSS-R). *Journal of Reproductive and Infant Psychology*, *36*(3), 235-245. https://doi.org/10.1080/02646838.2018.1443322

Hancock, G. R., & Mueller, R. O. (2001). Rethinking construct reliability within latent variable systems. In R. Cudeck, S. du Toit, & D. Soerbom (Eds.), *Structural equation modeling: Present and future. A festschrift in honor of Karl Jöreskog*. Scientific Software International.

Hollins Martin, C., & Fleming, V. (2011). The birth satisfaction scale. *International Journal of Health Care Quality Assurance*, *24*(2), 124-135. https://doi.org/10.1108/09526861111105086

Hollins Martin, C. J., & Martin, C. R. (2014). Development and psychometric properties of the Birth Satisfaction Scale-Revised (BSS-R). *Midwifery*, *30*(6), 610-619. https://doi.org/10.1016/j.midw.2013.10.006

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, *6*, 1-55.

Jefford, E., Hollins Martin, C. J., & Martin, C. R. (2018). Development and validation of the Australian version of the Birth Satisfaction Scale-Revised (BSS-R). *Journal of Reproductive and Infant Psychology*, *36*(1), 42-58. https://doi.org/10.1080/02646838.2017.1396302

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

Kline, R. B. (2011). *Principles and Practice of Structural Equation Modeling* (3rd ed.). Guilford Press.

Lyubenova, A., Neupane, D., Levis, B., Wu, Y., Sun, Y., He, C., Krishnan, A., Bhandari, P. M., Negeri, Z., Imran, M., Rice, D. B., Azar, M., Chiovitti, M. J., Saadat, N., Riehm, K. E., Boruff, J. T., Ioannidis, J. P. A., Cuijpers, P., Gilbody, S., … Thombs, B. D. (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*(1), e1860. https://doi.org/10.1002/mpr.1860

Martin, C. R., Hollins Martin, C. J., Burduli, E., Barbosa-Leiker, C., Donovan-Batson, C., & Fleming, S. E. (2017). Measurement and structural invariance of the US version of the Birth Satisfaction Scale-Revised (BSS-R) in a large sample. *Women and Birth*, *30*(4), e172-e178. https://doi.org/10.1016/j.wombi.2016.11.006

Martin, C. R., Hollins Martin, C. J., Burduli, E., Barbosa-Leiker, C., Donovan-Batson, C., & Fleming, S. E. (2018). The Birth Satisfaction Scale - Revised (BSS-R): Should the subscale scores or the total score be used? *Journal of Reproductive and Infant Psychology*, *36*(5), 530-535. https://doi.org/10.1080/02646838.2018.1490498

McDonald, R. P. (1999). *Test theory: A unified approach*. Erlbaum.

Mortazavi, F., Mehrabadi, M., Hollins Martin, C. J., & Martin, C. R. (2021). Psychometric properties of the birth satisfaction scale-revised (BSS-R) in a sample of postpartum Iranian women. *Health Care for Women International*, *42*(4-6), 836-851. https://doi.org/10.1080/07399332.2020.1802464

Nespoli, A., Colciago, E., Fumagalli, S., Locatelli, A., Hollins Martin, C. J., & Martin, C. R. (2021). Validation and factor structure of the Italian version of the Birth Satisfaction Scale-Revised (BSS-R). *Journal of Reproductive and Infant Psychology*, *39*(5),516-531. https://doi.org/10.1080/02646838.2020.1836333

Reise, S. P. (2012). The rediscovery of bifactor measurement models. *Multivariate Behavioral Research*, *47*, 667–696. https://doi.org/10.1080/00273171.2012.715555

Reise, S. P., Bonifay, W. E., & Haviland, M. G. (2013). Scoring and modeling psychological measures in the presence of multidimensionality. *Journal of Personality Assessment*, *95*(2), 129-140. https://doi.org/10.1080/00223891.2012.725437

Reise, S. P., Moore, T. M., & Haviland, M. G. (2010). Bifactor models and rotations: Exploring the extent to which multidimensional data yield univocal scale scores. *Journal of Personality Assessment*, *92*(6), 544-559. https://doi.org/10.1080/00223891.2010.496477

Reise, S. P., Scheines, R., Widaman, K. F., & Haviland, M. G. (2013). Multidimensionality and structural coefficient bias in structural equation modeling: A bifactor perspective. *Educational and Psychological Measurement*, *73*(1), 5-26. https://doi.org/10.1177/0013164412449831

Rodriguez, A., Reise, S. P., & Haviland, M. G. (2016a). Applying bifactor statistical indices in the evaluation of psychological measures. *Journal of Personality Assessment*, *98*(3), 223-237. https://doi.org/10.1080/00223891.2015.1089249

Rodriguez, A., Reise, S. P., & Haviland, M. G. (2016b). Evaluating bifactor models: Calculating and interpreting statistical indices. *Psychological Methods*, *21*(2), 137-150. https://doi.org/10.1037/met0000045

Rodriguez, A., Reise, S. P., & Haviland, M. G. (2016c). Evaluating bifactor models: Calculating and interpreting statistical indices. *Psychological Methods*, *21*(2), 137-150. https://doi.org/10.1037/met0000045

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

Rosseland, L. A., Reme, S. E., Simonsen, T. B., Thoresen, M., Nielsen, C. S., & Eberhard-Gran, M. (2020). Are labor pain and birth experience associated with persistent pain and postpartum depression? A prospective cohort study. *Scandinavian Journal of Pain*, *20*(3), 591-602. https://doi.org/10.1515/sjpain-2020-0025

Sijtsma, K. (2009). On the use, the misuse, and the very limited usefulness of Cronbach's alpha. *Psychometrika*, *74*(1), 107-120. https://doi.org/10.1007/s11336-008-9101-0

Skvirsky, V., Taubman-Ben-Ari, O., Hollins Martin, C. J., & Martin, C. R. (2020). Validation of the Hebrew Birth Satisfaction Scale - Revised (BSS-R) and its relationship to perceived traumatic labour. *Journal of Reproductive and Infant Psychology*, 38(2), 214-220. https://doi.org/10.1080/02646838.2019.1600666

Steiger, J. H., & Lind, J. (1980). *Statistically-based tests for the number of common factors* Annual Spring Meeting of the Psychometric Society, Iowa City, USA.

Stucky, B. D., & Edelen, M. O. (2015). Using hierarchical IRT models to create unidimensional measures from multidimensional data. In S. P. Reise & D. A. Revicki

(Eds.), *Handbook of Item Response Theory Modeling: Applications to Typical Performance Assessment* (pp. 183-206). Routledge.

Ten Berge, J. M. F., & Sočan, G. (2004). The greatest lower bound to the reliability of a test and the hypothesis of unidimensionality. *Psychometrika*, *69*(4), 613-625. https://doi.org/10.1007/bf02289858

The International Consortium for Health Outcome Measurement. (2016). *Pregnancy and Childbirth Standard Set and Reference Guide*. http://www.ichom.org/medical-conditions/pregnancy-and-childbirth/

Vardavaki, Z., Hollins Martin, C. J., & Martin, C. R. (2015). Construct and content validity of the Greek version of the Birth Satisfaction Scale (G-BSS). *Journal of Reproductive and Infant Psychology*, *33*(5), 488-503. https://doi.org/10.1080/02646838.2015.1035235

Zinbarg, R. E., Revelle, W., Yovel, I., & Li, W. (2005). Cronbach's, Revelle's, and McDonald's H: Their relations with each other and two alternative conceptualizations of reliability. *Psychometrika*, *70*, 123–133. https://doi.org/10.1007/s11336-003-0974-7

Table 1

Sociodemographic and Obstetric Characteristics of the Sample

	N = 1373
Women's age (years) *M*(*SD*); range	31.83 (4.44); 18-44
Infants' age (months) *M*(*SD*); range	2.58 (1.46); 0-10
Infants' gender *n*(%)	
Male	682 (49.7)
Female	691 (50.3)
Living with the baby's father *n*(%)	
Yes	1344 (98)
No	27 (2)
Education *n*(%)	
≤ High school	379 (27.6)
≥ College or graduate degree	994 (72.4)
Monthly family income (post-tax) *n*(%)	
< 1000€	115 (8.4)
1001€-2000€	725 (52.8)
2001€-3000€	340 (24.8)
>3000€	193 (14.1)
Parental leave *n*(%)	
Yes, currently on parental leave	1201 (87.5)
No	172 (12.5)
Conception *n*(%)	
Spontaneous	1313 (95.6)
Medically induced (fertility treatment)	60 (4.4)
Gestational weeks *M*(*SD*); range	38.91 (1.58); 29-43
Preterm vs term pregnancy	

Preterm (< 37 weeks)	76 (5.5)
Term (≥ 37 weeks)	1294 (94.5)
Delivery type *n*(%)	
Unassisted vaginal delivery	538 (39.2)
Assisted/intervention delivery[1]	834 (60.8)
Birth place	
Public hospital	955 (69.6)
Private hospital	405 (29.5)
At home	11 (0.8)
On the way to the hospital	1 (0.1)

Note. [1]Assisted/intervention delivery included elective caesarean section, emergency caesarean section, and use of ventouse and forceps.

Table 2

Mean, Standard Deviation and Distributional Characteristics of Individual Portuguese BSS-R items, Subscale Totals and the Total Portuguese BSS-R Score.

Item	Item content	Domain*	Mean	SD	Min-Max	Skew	Kurtosis	*SE*
BSS-R 1	I came through childbirth virtually unscathed.	SE	2.17	1.31	0 - 4	-0.15	-1.23	0.04
BSS-R 2	I thought my labour was excessively long.	SE	2.69	1.32	0 - 4	-0.72	-0.68	0.04
BSS-R 3	The delivery room staff encouraged me to make decisions about how I wanted my birth to progress.	QC	1.94	1.35	0 - 4	-0.01	-1.22	0.04
BSS-R 4	I felt very anxious during my labour and birth.	WA	2.03	1.36	0 - 4	0.04	-1.34	0.04
BSS-R 5	I felt well supported by staff during my labour and birth.	QC	3.04	1.10	0 - 4	-1.20	0.77	0.03
BSS-R 6	The staff communicated well with me during labour.	QC	3.00	1.10	0 - 4	-1.15	0.59	0.03
BSS-R 7	I found giving birth a distressing experience.	SE	2.73	1.25	0 - 4	-0.71	-0.63	0.03
BSS-R 8	I felt out of control during my birth experience.	WA	2.31	1.33	0 - 4	-0.27	-1.16	0.04
BSS-R 9	I was not distressed at all during labour.	SE	2.16	1.31	0 - 4	-0.07	-1.25	0.04
BSS-R 10	The delivery room was clean and hygienic.	QC	3.67	0.54	1 - 4	-1.42	1.21	0.01
Stress	Subscale total		9.76	4.02	0 - 16	-0.34	-0.65	0.11
Attributes	Subscale total		4.35	2.35	0 - 8	-0.10	-1.04	0.06

Quality	Subscale total	11.64	3.18	1 - 16	-0.78	0.24	0.09
Total	Total score	25.74	8.16	1 - 40	-0.41	-0.33	0.22

Note. *Domain of the Portuguese-BSS-R. SE = Stress Experienced During Labour, WA = Women's Personal Attributes, QC = Quality of Care. *SE* = standard error of kurtosis.

Table 3

Confirmatory Factor Analysis of the Portuguese BSS-R.

Model	χ^2	df	RMSEA	SRMR	CFI
1. Single-factor model (outliers)	1248.69	35	0.158	0.076	0.797
2. Three-factor measurement model (outliers)	166.70	32	0.055	0.031	0.977
3. Two factor (SE+WA combined) (outliers)	166.99	34	0.053	0.031	0.978
4. Bifactor (SE+WA combined) (outliers)	108.74	25	0.049	0.021	0.986
5. Single-factor model (no outliers)	1368.74	35	0.167	0.077	0.784
6. Three-factor measurement model (no outliers)	185.74	32	0.059	0.034	0.975
7. Two factor (SE+WA combined) (no outliers)	188.22	34	0.057	0.034	0.975
8. Bifactor (SE+WA combined) (no outliers)	115.30	25	0.051	0.022	0.985

Table 4

Bifactor Indices for the Bifactor Model

Bifactor model	ECV	Omega/ OmegaS	OmegaH/ OmegaHS	Relative omega
General Factor	.694	.894	.762	.852
Attributes & Stress	.181	.866	.121	.140
Quality of Care	.506	.782	.376	.481

Note. OmegaS and OmegaHS are omega indices for the subscales. ECV = explained common variance.

Table 5

Invariance Analysis of the Portuguese BSS-R Based on Groups Stratified by EPDS Screen Score Cut-Off 10

Model	χ^2	df	MC	$\Delta\chi^2$	Δdf	*p*	RMSEA	SRMR	CFI	ΔCFI
1. Negative screen	145.48	32	NA				0.06	0.04	0.970	
2. Positive screen	72.715	32	NA				0.05	0.04	0.981	
3. Configural	218.20	64	NA				0.06	0.03	0.974	
4. Metric	233.32	71	3	15.13	7	0.03	0.06	0.04	0.973	0.001
5. Scalar	262.93	78	4	29.61	7	< 0.01	0.06	0.04	0.969	0.004
6. Strict	293.59	88	5	30.66	10	< 0.01	0.06	0.05	0.966	0.003

Note. MC = model comparison

Table 6

Correlations of Portuguese BSS-R Subscales and Total Score and Comparison with Original UK BSS-R Validation Study (Hollins Martin & Martin, 2014)

Scale combination	Current study *r*	UK study *r*	Z	95% CI	*p*	Cohen's q	Effect size
Stress-Attributes	.74	.57	4.21	(0.08 – 0.27)	< .001	0.30	Medium
Stress-Quality	.53	.26	4.50	(0.15 – 0.40)	< .001	0.32	Medium
Attributes-Quality	.49	.35	2.37	(0.02 – 0.27)	.02	0.17	Small
Total score-Stress	.91	.86	3.26	(0.02 – 0.09)	.001	0.23	Small
Total score-Attributes	.84	.80	1.70	(-0.005 – 0.09)	.09	0.12	Small
Total score-Quality	.79	.63	4.59	(0.08 – 0.25)	< .001	0.33	Medium

Table 7

Cronbach's Alpha of Portuguese BSS-R Subscales and Total Score and Comparison with Original UK BSS-R Validation Study (Martin & Hollins Martin, 2014).

Subscale	Current study	UK study	χ^2	*p*
Stress	.78	.71	4.74	0.03
Attributes	.69	.64	0.75	0.39
Quality	.74[1]	.74	0.02	0.89
Total score	.86	.79	14.37	< .001

Note. [1]Calculation taken to 4 decimal points to discriminate between studies. Degrees of freedom = 1.

Table 8

Correlation Coefficients Between PT-BSS-R Subscale Scores and Global Birth Satisfaction.

	BSS-R Stress	BSS-R Attributes	BSS-R Quality	BSS-R total score	Global Birth satisfaction
BSS-R Stress	-				
BSS-R Attributes	.74*	-			
BSS-R Quality	.53*	.49*	-		
BSS-R total score	.91*	.84*	.79*	-	
Global Birth satisfaction	.68*	.57*	.63*	.74*	-

* $p < .001$

Figure 1

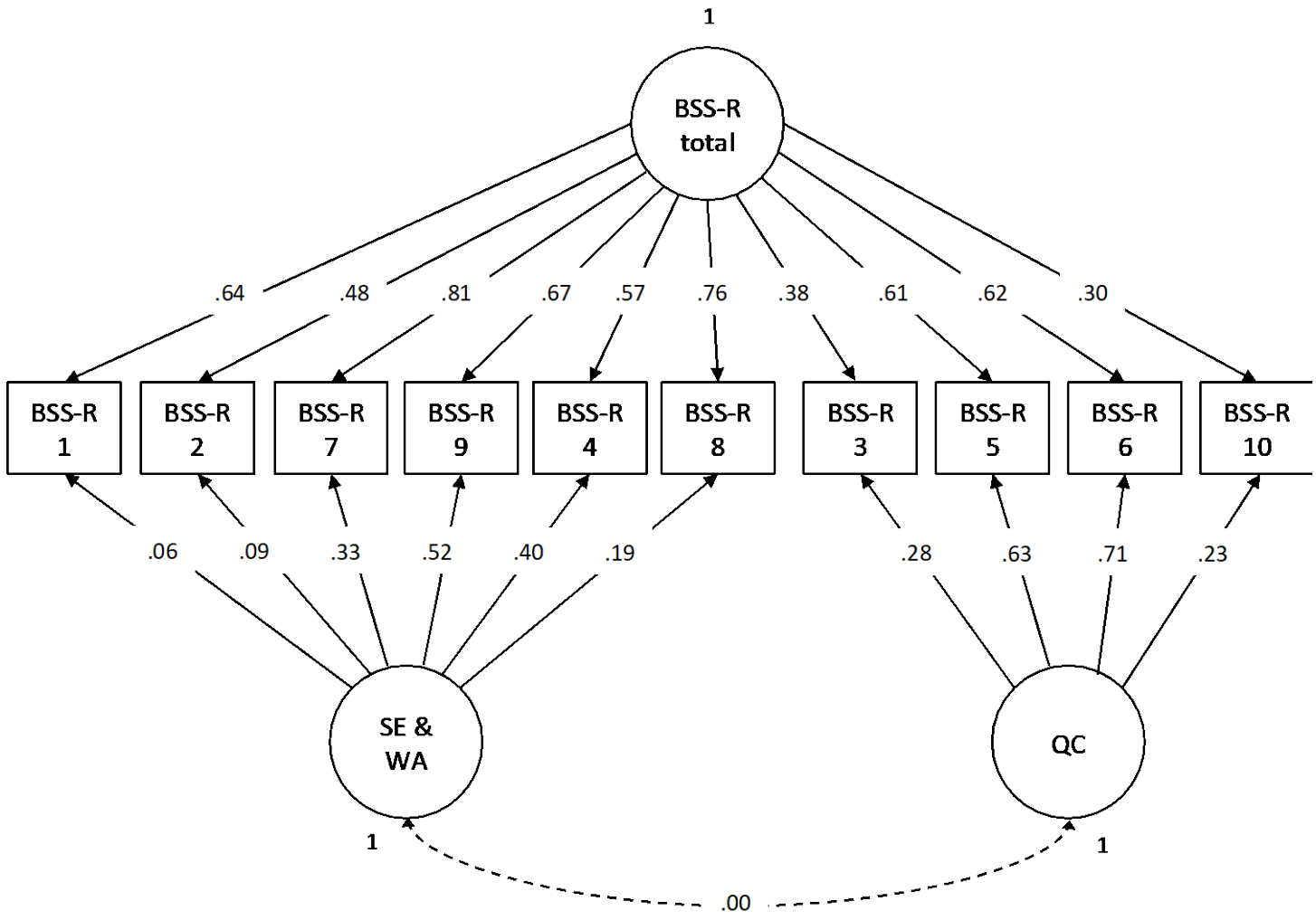
1
BSS-R total
.64
.48
.81
.67
.57
.76
.38
.61
.62
.30
BSS-R 1
BSS-R 2
BSS-R 7
BSS-R 9
BSS-R 4
BSS-R 8
BSS-R 3
BSS-R 5
BSS-R 6
BSS-R 10
.06
.09
.33
.52
.40
.19
.28
.63
.71
.23
SE & WA
QC
1
1
.00

Figure 1 Caption: *Bifactor Model of the Portuguese BSS-R Based on Collapsed SE and WA Factors*

Figure 1 Alt Text: image representing the bifactor model of the Portuguese BSS-R based on collapsed Stress Experienced During Labour and Women's Personal Attributes factors. A general factor of birth satisfaction is represented as well as two specific factors: 1) Stress Experienced During Labour and Women's Personal Attributes and 2) Quality of Care Provision.