Martin, C.R., Hollins Martin, C., Burduli, E., Barbosa-Leiker, C., Donovan-Batson, C., Fleming, S. (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 (2017) e172–e178.

Measurement and structural invariance of the US version of the Birth Satisfaction Scale-Revised (BSS-R) in a large sample

Colin R. Martin*[1], Caroline J. Hollins Martin[2], Ekaterina Burduli[3], Celestina Barbosa-Leiker[4], Colleen Donovan-Batson[5] and Susan E. Fleming[6]

[*1] Professor of Mental Health, Faculty of Society and Health, Buckinghamshire New University, Uxbridge, UK, UB8 1NA. E-mail: colin.martin@bucks.ac.uk

[2] Professor in Maternal Health, School of Nursing, Midwifery and Social Care
Edinburgh Napier University, Scotland, UK. EH11 4BN.
E-mail: C.HollinsMartin@napier.ac.uk

[3] Post-Doctoral Fellow, Sleep and Performance Research Center, Elson S. Floyd College of Medicine, Washington State University, Spokane, Washington State, USA.
E-mail: eburduli@wsu.edu

[4] Associate Professor, Washington State University College of Nursing
Washington State University, Spokane, Washington State, USA. E-mail: celestina@wsu.edu

[5] Midwives Alliance of North America, Washington State, USA.
E-mail: colleendb.midwife@gmail.com

[6] Assistant Professor, Seattle University College of Nursing
Seattle, Washington State, USA.
E-mail: fleminsu@seattleu.edu

Measurement and structural invariance of the US version of the Birth Satisfaction Scale-Revised (BSS-R) in a large sample

Abstract

Background: The 10-item Birth Satisfaction Scale-Revised (BSS-R) is being increasingly used internationally. The use of the measure and the concept has gathered traction in the United States following the development of a US version of the tool. A limitation of previous studies of the measurement characteristics of the BSS-R is modest sample size. Unplanned pregnancy is recognised as being associated with a range of negative birth outcomes, but the relationship to birth satisfaction has received little attention, despite the importance of birth satisfaction to a range of postnatal outcomes.

Aim: The current investigation sought to evaluate the measurement characteristics of the BSS-R in a large postpartum sample.

Methods: Multiple Groups Confirmatory Factor Analysis (MGCFA) was used to evaluate a series of measurement and structural models of the BSS-R to evaluate fundamental invariance characteristics using planned/unplanned pregnancy status to differentiate groups.

Findings: Complete data from N=2116 women revealed that the US version of the BSS-R offers an excellent fit to data and demonstrates full measurement and structural invariance. Little difference was observed between women on the basis of planned/unplanned pregnancy stratification on measures of birth satisfaction.

Discussion: The established relationship between unplanned pregnancy and negative perinatal outcomes was not found to extend to birth satisfaction in the current study. The BSS-R demonstrated exemplary measurement and structural invariance characteristics.

Conclusion: The current study strongly supports the use of the US version of the BSS-R to compare birth satisfaction across different groups of women with theoretical and measurement confidence.

Key Words: Birth Satisfaction Scale-Revised (BSS-R), childbearing women, United States translation, measurement invariance, measurement equivalence.

Introduction

Statement of significance

Problem or Issue: Accurate comparisons between groups of interest on key maternal health concepts assessed by questionnaire requires the measure to be free of bias (measurement non-invariance), however, this is seldom evaluated.

What is Already Known: Methodological approaches to the determination of measurement invariance have been developed and are readily applicable to measures used in maternal health.

What this Paper Adds: Empirical confidence in unbiased comparisons between groups differentiated by planned/unplanned pregnancy status on a key index of birth satisfaction.

Birth satisfaction represents a complex construct of implicit and profound relevance to the woman's perceived birth experience (1). A broad variety of assessment tools have been used to measure birth satisfaction (2-5), though many of the available tools have been criticised for their distal relationship to an underlying theoretical construct (1).

The Birth Satisfaction Scale (BSS) (6) represented a departure from the established instrument pool by developing the measure from a thematic review of the literature. A short-form version was developed by Hollins Martin and Martin (7) comprising the 10 best performing items based on psychometric characteristics and measurement coherence to the thematic structure underpinning the BSS. Consistent with the BSS, the BSS-R assesses three domains (i) stress experienced during childbearing, (ii) women's attributes and, (iii) quality of care, using a self-report Likert format. This instrument, the Birth Satisfaction Scale-Revised (BSS-R) has become increasingly used internationally, with translation and validation studies being published (8-10) or underway (communications to instrument developers).

It is noteworthy, that although the psychometric profile of the BSS-R is impressive from validation studies irrespective of language version (7, 8, 10), the sample size of all of these studies are modest (N=162–N=228). A potential limitation of these studies is that although affirmation of the underlying tri-dimensional factor structure of the instrument is forthcoming, the stability of the underlying structure between groups of interest (for example, parity, or type of birth) cannot be attained, since comparison between such groups from a factor structure measurement perspective requires each group of interest to be of a significant size (N>100; 11)[1]. Consequently, the validation studies conducted to date on the BSS-R have looked at group differences exclusively by comparison with mean scores. This represents an appropriate approach to determine known-groups discriminate validity of the tool. However, the underlying stability of the tool across groups cannot be determined and may thus represent a source of measurement error (12). Given the penetration of the BSS-R into the contemporary birth satisfaction literature and the potential for use of the measure as a key performance indicator for maternity service care delivery (13), the underlying stability of the measurement model of the BSS-R is important if differences observed between groups can be confirmed to be true differences rather than an artefact of measurement error due to groups responding to the measure in a characteristically different way (12, 14). Martin and colleagues (15) conducted a secondary analysis on the original BSS-R validation dataset (7) and the Greek-language validation dataset

[1] Extrapolated from minimum sample size recommendations for exploratory factor analysis.

(10) and were able to confirm that the instrument was generally equivalent between the two versions. The implications of this observation is that scores and data on the tool could be directly compared and any differences between groups being representative of true differences rather than measurement bias (14).

The secondary approach taken by Martin and colleagues (15) was to determine the measurement invariance characteristics of the BSS-R across two BSS-R datasets using multiple groups confirmatory factor analysis (MGCFA) within a structural equation modelling (SEM) framework. This process of measurement invariance evaluation being an established and rigorous approach to determining the equivalence, or otherwise, of a measure between groups or across time points (12). An instrument which fails to demonstrate measurement invariance suggests that any comparisons made and conclusions drawn could be confounded by fundamental response bias issues and thus its findings would be unreliable. Determining measurement invariance therefore goes beyond the assertion of Werneke and colleagues (16) that the measurement characteristics of an instrument should be confirmed in each group of interest before comparisons between groups can be made directly and actually be comparable in a meaningful way and without systemic measurement error.

Unplanned pregnancy: A characteristic of choice for invariance evaluation

It is of note that the term ‘unplanned pregnancy’ comprises two distinct categories of pregnancy intentions, these being mistimed pregnancies that would otherwise be planned for a later date and unwanted pregnancies that are not wanted or desired at a later date (17). While approximations of the percentages of unplanned pregnancy differ, research proposes that in westernised countries 37% to 48% of pregnancies are unintended (18), which encompasses 5% to 23% of the total number of live births (19). There is a considerable amount of research that suggests that unplanned pregnancy is associated with potential adverse outcomes (18). Some of these include lower rates of attending for pre-natal care (20), post-partum depression (20, 21), premature birth (22), low birth weight (22) and poorer quality of parent/child relationship (23). Such findings imply a high cost of unplanned pregnancy for both the woman and society (18).

Evidence supports the perspective that unwanted pregnancy is associated with a comparatively more negative effect than untimed pregnancy (24). Pregnancy intention itself has a variety of effects on both mother and infant outcomes. For example, a woman faced with an unplanned pregnancy is less likely to attend for preconception care (25) and early antenatal care, which can bring costs in terms of reducing vigilance at detecting medical problems or complications that could be remedied. One issue bearing, is that organogenesis and early system development has already taken place, with limited opportunity to influence fetal development in the first trimester. Topics addressed during preconception care involve monitoring of diet (26), maternal weight assessment (27), smoking, substance misuse, and current medication (28), avoidance and treatment of infections (e.g., toxoplasmosis and cytomegalovirus;(29), and sexually transmitted diseases (e.g., chlamydia, gonorrhea, herpes simplex virus, syphilis, & HIV;(30). In addition, preconception care helps perfect management of prior medical conditions, such as diabetes (31). An unplanned pregnancy can also inhibit the woman from taking the fullest advantage of human genetics. The health and social risks associated with potential complications yields greater chance of the woman having a premature birth, caesarian section, high intervention birth, with associated adverse maternal and fetal outcome.

The aim of the current study was to address the shortcomings due to modest sample size of contemporary BSS-R validation studies through evaluation of key measurement properties of the tool in a large N dataset.

The objectives of the study were to:

1. Confirm the adequacy of fit of the tri-dimensional factor structure.
2. Determine the measurement invariance characteristics between groups differentiated on the basis of whether the pregnancy was planned or unplanned.
3. Evaluate the correspondence of adapted items to original items.

Method

A cross-sectional design employing a convenience sample and using the United States validated version of the BSS-R (8) distributed to participants using the Qualtrics (32) survey system via electronic linkages. Differentiation into planned pregnancy status was made on the basis of the single item survey question ‘Was your recent pregnancy planned?’ presented with a dichotomous ‘Yes/No’ response format. Informed consent for study participation was embedded in the survey. Inclusion criteria were women over 18 years of age who had initially planned to give birth either at home or in birth centres in the United States. The study was reviewed and deemed exempt by Seattle University Internal Review Board (IRB) in compliance with 45CFR46.101(b):2 of the United States Department of Health and Human Sciences research guidelines.

Participants

A convenience sample of 2229 women participated in the study. Extensive details of the characteristics of the full sample are described in Fleming et al. (33).

Statistical analysis

Confirmatory Factor Analysis (CFA) was conducted using maximum-likelihood (ML) estimation (12, 34, 35), with this approach justified by the generally normal distribution of BSS-R items observed in the Hollins Martin and Martin (7) study. Two three-factor models from Hollins Martin and Martin's original validation study were evaluated: (i) three-factor correlated model of *stress experienced during labour*, *quality of care provision,* and *women's personal attributes* factors, and (ii) a hierarchical model based on (i), but with a higher order factor of *experience of childbearing*. To determine any issues related to the adaptation of original BSS-R items within the USA version of the scale, these two models will be evaluated with the original UK BSS-R item 'I came through childbirth virtually unscathed' and with the US-specific item 'I came through childbirth virtually unharmed'. Consequently, a total of four models will be evaluated (i. USA three-factor, ii. USA hierarchical, iii. UK three-factor, and iv. UK hierarchical). Model invariance evaluation will first be conducted on the established three-factor models and following this, the hierarchical models will be tested based upon the optimal level of measurement invariance observed, based upon the three-factor model evaluation. Model fit was evaluated by a battery of fit indices (36) including the comparative fit index (CFI;(37), the root mean squared error of approximation (RMSEA) and the standardised root mean residual (SRMR). CFI values > 0.90 indicates an acceptable fit (38) more stringent CFI ≥ 0.95 indicating a good fit to the data (39). RMSEA values ≤ 0.08 indicate acceptable model fit (40), and values of ≤0.05 indicative of good fit (41). SRMR values ≤ 0.08 indicate acceptable model fit (39).

The best-fitting of the two models will then be evaluated for measurement invariance characteristics as a function of the dataset split between participants who either had a planned or unplanned pregnancy. Increasingly restrictive versions of the underlying measurement model are tested to determine measurement invariance following determination of the most appropriate measurement model (12, 14, 15). There remains some debate over the use of an initial omnibus model free of constraints between groups (42) prior to proceeding to increasingly restricted models. An omnibus baseline model of all BSS-R data without group differentiation is conducted to ensure acceptable fit and consistency with observations from previous studies, essentially, this is the best-fit CFA model. A configural invariance model is then evaluated to determine if the factor model and pattern of loadings is equivalent across groups. A metric invariance model is then tested, where item-factor loadings are restricted to be the same across groups and assuming configural invariance. Metric invariance is a requirement to confirm that the measurement model constructs defined by the measurement model have consistency of meaning across groups (43). A further restriction to the model, assuming metric invariance, is scalar invariance evaluation where item intercepts are restricted to be equal across groups. Establishing measurement invariance between groups at the configural, metric, and scalar levels indicates measurement invariance of the tool in this context. It is possible that some items will be invariant across groups, while others won’t be, and this situation is described as partial invariance (12) contextually defined by the level of invariance testing at which a non-invariant item is identified. Recognising that models may be *partially* invariant at each level, the non-invariant component of the model, for example a single item mean or item-factor loading can be identified (12). In the event of a non-invariant model component being identified, the invariance evaluation would normally stop at that particular level, which is essentially, the best-fitting partially invariant model (12, 14, 44). A further level of model constraint is to evaluate item error variance invariance in the event of demonstrable scalar invariance. Strict invariance, though not required for scores to be compared across groups, does offer an additional insight in terms of both demonstrating that the explained variance for each of the items assessed is the same across groups and by implication, the underlying factors (BSS-R sub-scale domains) are the same in terms of item measurement across groups. Beyond the invariance evaluation of the BSS-R at the measurement level, it is also possible to evaluate the *structural* invariance of the tool (45-48). Testing for structural invariance is unusual in a clinically-applied instrument, however, evaluating the structural invariance of a measure can be extremely useful in extrapolating theoretical aspects of the measure to participant's responses to the tool. Structural invariance, though rarely evaluated in terms of MGCFA, focuses exclusively on the underlying latent variables and is only conducted in the event of the demonstration of strict measurement invariance. The structural invariance component of the model is also evaluated by testing increasingly constrained versions of the model, starting with the strict measurement invariance model as a new 'baseline' model. Firstly, factor means are constrained to be equal and if this level of structural invariance is satisfied, a model evaluating factor means and variances constrained to be equal across groups is tested. Finally, in the event of means and variances being observed to be invariant between groups, factor covariances are then constrained to be equal. The order of structural invariance is unimportant, but it is contingent on measurement invariance being established (49).

The criteria to determine if a nested model is significantly different or not from the previous model is to use the χ^2 difference test (12). However, χ^2 is inflated by sample size (50), which represents a particular limitation for large N studies. A more robust approach has been to use the CFI to compare models, with values of ≤ 0.01 indicating measurement invariance between models (51). Similarly, the fit criteria outlined earlier for CFA model acceptability applies to the evaluation of models under measurement and structural invariance testing, thus in the event of determining measurement invariance or structural invariance, irrespective of level, the model is still required to be of acceptable fit.

Statistical comparison of the two BSS-R1 items (UK/US) was made using the paired-sample *t*-test. Finally comparisons will be conducted to determine if there are group differences as a function of planned baby status (planned/unplanned) on the BSS-R (US version) total and sub-scale scores using the between-subjects t-test. Effect sizes will be estimated for each between-subjects comparison using Hedges *g,* which in contrast to Cohen's *d* is better suited for group comparisons of unequal sample sizes (52). Cohen's *d* (53) by contrast will be used for the within-subject comparison.

Statistical analysis was conducted using the R programming language (54).

Results

The dataset was screened for missing BSS-R data from the N=2229, revealing a minimal percentage missing (<1%, N=12). These cases were removed, which left a dataset of N=2217. Detection of multivariate outliers was accomplished by calculating Mahalanobis distances (43, 55) and revealed N=101 (<5%) multivariate outliers, which were subsequently excluded. The requirements of non-missing BSS-R data, and absence of multivariate outliers, yielded a useable sample size of N=2116 for MGCFA, which represented 95% of the pre-screened dataset. Stratifying by planned pregnancy status revealed N=1600 (76%) mothers had planned their baby, compared with N=516 (24%) unplanned babies. Mean BSS-R total and sub-scale mean scores as a function of planned pregnancy status are summarised in Table 1. The between-subjects *t*-test revealed a significant difference between groups ($p < 0.05$) on the BSS-R quality of care sub-scale, with the planned pregnancy group reporting better birth satisfaction on this domain compared to the unplanned pregnancy group. Examination of the effect size reveals, however, this difference to be negligible according to Cohen's (1988) criteria. No other statistically significant between-subjects differences were observed and effect sizes were all negligible.

TABLE 1. ABOUT HERE

The findings of the measurement and invariance testing are summarised in Table 2. The USA version of the BSS-R will be examined first. The overall model (all data model 1a.) was found to offer an excellent fit to the data. Examining each group (planned/unplanned) separately (models 1b. & 1c.) revealed an excellent fit to data. The configural model fit (model 2.) was found to offer a good fit to data. No significant difference (ΔCFI ≤1) was observed between model 2 and model 3, which confirms metric invariance. Similarly, no significant difference was observed between model 4 and model 3, thus confirming scalar invariance. The final element of the measurement model, evaluating model 5 against model 4 confirmed invariance at the strict level. A comparison of this model with the USA hierarchical strict invariance version revealed the three-factor model to offer a descriptively marginal better fit to data. Structural invariance testing revealed factor means invariance (model 6 versus model 5), factor means and variances invariance (model 7 versus model 8), and finally, factor means, variances and covariances invariance (model 8 versus model 9.). Evaluation of the UK version of the BSS-R (models 9a to model 16) revealed a consistently similar pattern of model fit to the USA version that is identical in interpretation. Essentially, measurement and structural invariance and the three-factor strict invariance measurement model demonstrates descriptively marginal better fit to the UK hierarchical strict invariance measurement model.

TABLE 2. ABOUT HERE

A statistically significant difference ($t_{(2115)}$ = 16.12, $p < 0.001$, $d = 0.35$) was observed between the original BSS-R 1 item ‘‘I came through childbirth virtually unscathed’ (M = 3.03, SD = 1.13) and the US version ‘I came through childbirth virtually unharmed’ (M = 3.30, SD = 1.01). Using Cohen’s (53) criteria, the effect size would be classified as small.

Discussion

The current study offers a unique insight into the measurement and structural qualities of the BSS-R, with this being the first study to investigate both measurement and structural invariance on the birth satisfaction measure. Also, this is the first paper that has looked at clinically pertinent domain for equivalence evaluation, i.e., planned pregnancy status. Prior to an examination of the psychometric findings in detail, the direct between-groups comparisons on BSS-R and BSS-R sub-scale scores will be discussed.

Contrary to the prevailing literature on the impact of unplanned pregnancy on relatively deleterious outcomes (24), little evidence was found in the current study for any impact on birth satisfaction. It should be noted however, that an inherent limitation within the study is that unplanned pregnancy categorisation was determined by a dichotomous 'Yes/No' response to a single question regarding planned pregnancy. It has been highlighted that unplanned pregnancy is associated with more negative outcomes than mistimed pregnancy (17), thus the current study design inherently lacked the sensitivity to differentiate between these sub-groups. Given the potential salience of this differentiation to clinical outcomes and potentially, to birth satisfaction, it is suggested that future studies differentiate these two sub-categories of unplanned pregnancy.

Clearly, women who had planned their baby reported significantly higher BSS-R quality of care sub-scale scores, but scrutiny of the mean scores reveals the absolute difference to be small. Indeed, examination of the effect size indicates the difference is negligible. It is acknowledged that sample size contributes to an arbitrary value of statistical significance, and thus even trivial differences in mean scores can lead to statistically significant differences between groups with a sufficiently large sample size (56). This group difference observation should, therefore, not be overstated or over-interpreted at this stage in view of absolute magnitude. Although it is conceded, that should this observation be consistent in other populations evaluated in future studies, further investigation of this phenomenon is warranted. The absence of any significant differences on the BSS-R total score, BSS-R ‘stress experienced during labour’, and the BSS-R ‘women’s attributes’ sub-scale would indicate that the groups are comparable in levels of birth satisfaction.

A possible explanation for this observation may be the participant population, which represents a self-selected group with an engendered desire to have their babies either at home or a birth centre. Therefore, these women may have different attitudes, expectations, and resources that mitigates in the unplanned pregnancy group any negative impact on birth outcomes as assessed by birth satisfaction. To determine the plausibility of such an explanatory account would require a further study, where women representing the spectrum of birthing choices and services could be represented. The attributes of the current participant population may also have impacted on the intriguing finding of a statistically significant difference between both versions (US and UK) of BSS-R item 1. A fascinating juxtaposition was observed whereas, in contrast to the previous US BSS-R study of Barbosa-Leiker, Fleming (8), where participants reported a significantly higher score on the UK version of the item ‘unharmed’. In the current study this was reversed, with the UK version ‘unscathed’ scoring higher, though the effect size was small. Fundamentally the different sampling procedures between Barbosa-Leiker et al. (8) and the current study are likely to define uniquely different populations, and therefore may contribute to the difference observed. Irrespective of origin of influence, our findings would concur entirely with Barbosa-Leiker et al. (8) in advocating the use of the US version of the tool in US populations, and supporting the rationale for the original development and validation of the US version of the BSS-R.

The evaluation of invariance characteristics of the BSS-R as a function of planned pregnancy status represents a valuable contribution to the literature on the psychometric properties of the tool. Importantly, it was observed that the fit to data, prior to invariance evaluation was excellent, both overall and when examined at the planned pregnancy status group level for the three-factor model of the BSS-R. Indeed, this model fit excellence was observed irrespective of whether the US or UK version of the tool was specified within the CFA model. Indeed, comparison of the CFA models of the current study are entirely consistent with the original validation model of the BSS-R (7) across fit indices. Evaluation of measurement invariance revealed both versions of the BSS-R to be invariant to the optimal measurement level of strict invariance. This demonstration of robust measurement invariance goes beyond the accepted criteria generally agreed for meaningful comparisons between groups (12, 14, 49), and demonstrates that comparisons on all domains of birth satisfaction between the groups specified in the current study can be made with confidence. Thus, observations of differences between planned/unplanned pregnancy groups can be made with confidence and without concern of confound due to group level measurement bias or error. A further observation was that the strict-fit measurement model, when re-specified as a hierarchical model, was a slightly poorer, but still acceptable, fit in comparison to the three-factor correlated model (irrespective of UK or USA versions of the measure). Since these differences between hierarchical and three-factor models are relatively small, and some fit measures have inherent bias in relation to parsimony (14, 57), and a hierarchical model represents a complex model, there is insufficient evidence to conclude one model structure is superior to the other. The practical conclusion to this is, consistent with the observation of the previous US-based BSS-R study (8), that the three sub-scale scores and the total score all have significant utility in the assessment of birth satisfaction.

The finding of structural invariance represents the first instance, as far as the authors are aware of structural equivalence within the BSS-R. It has been asserted that the observation of structural invariance within a MGCFA is of mainly theoretical interest (49), particularly given that structural invariance is not a requirement for comparison of the measure between different groups or populations. Since strict measurement invariance is a requirement prior to evaluating structural invariance within a MGCFA, and that instances of strict measurement invariance within the perinatal and reproductive psychology measurement literature are rare, this also precipitates a context of near absence of structural invariance evaluation within the field. However, the observation of structural invariance is important since it demonstrates the conceptual stability of the measure and robustness of its theoretical underpinnings. The BSS-R is a short measure for a multi-dimensional measure, with minimal item redundancy. Therefore, the exemplar measurement and structural invariance qualities highlight the theoretical integrity of the process of development of the original birth satisfaction scale by Hollins Martin and Fleming (6). Moreover, the veracity of best-item selection based on rigorous psychometric criteria for the development of the BSS-R (7). It is noteworthy that this process of developing an instrument directly from a theoretical framework, the mapping of items to that framework, and the development of a short version using a systematic psychometric review and assessment of the measurement characteristics of individual items and their relationship to factors is rare in the perinatal field. The exemplar measurement and structural invariance characteristics demonstrated in the current study are therefore likely to be influenced by the BSS-R instrument development heritage. Confirmation of this perspective may be inferred by the finding that both UK and US versions of the instrument demonstrated full measurement and structural invariance. A strength of the current study is the large sample size, the limitations of small sample sizes being highlighted by other researchers using, developing, adapting or evaluating the BSS-R (8, 9, 15).

The current study did have a number of limitations which may be addressed by further research work on the measure. Firstly, the participant population may be representative of a very specific type of mother. That is, a childbearing woman who has a strong desire for a home birth or birth centre delivery in contrast to a medically-orientated model of care. It is not known how representative this population is of the population of US mothers who experience a limited variety of birthing choices, evidenced by their high elective caesarean section rate, and therefore replication or comparison studies in the wider population of mothers is to be encouraged. A further limitation is the use of online data capture to facilitate a large sample size. Online data collection is considered a legitimate method of data capture, assuming careful design of method and participant recruitment process (58). The online data capture method used in the current study used a network of midwives to facilitate promotion of the internet site within the target population, and in itself this represents an important safeguard to the integrity of the study. However, replication of the study using data capture within a direct face-to-face context, perhaps as part of a large clinical follow-up study would be invaluable in corroborating the findings from the current study. Finally, the current study evaluated a single dimension of invariance, and that is planned pregnancy status. Evidence of measurement and indeed structural invariance using group differentiation factors, such as delivery type (vaginal, instrument, Caesarean section) would offer valuable additional evidence for the veracity of the US version of the BSS-R.

The current study has found additional evidence for the measurement robustness and structural integrity of the US version of the BSS-R using systematic equivalence evaluation and benefitting from a large sample size. In relation to midwifery practice, the validated BBS-R could be used by maternity care professionals to audit and improve standards of intranatal care provision. Firstly, the instrument could be used to discover aspects of birth dissatisfaction that could be remedied, adjusted, or resolved through adapting the labour environment or midwifery approach. Secondly, midwives could use the BSS-R in conjunction with other validated measures to study relationships between aspects of birth satisfaction and, for example, depression, locus of control, or infant attachment. In essence, finding out more about what affects birth satisfaction could help midwives improve standards of intranatal care provision at both a quantitative and qualitative level. The BSS-R offers midwives, other health professionals and researchers a robust measure to quantify childbearing women's satisfaction of their birthing experiences. In addition, the tool may enable midwifery practice, by generating robust and reliable woman-centred and relevant birth satisfaction information to inform policy makers and the wider medical community who share their interest in providing optimal and comprehensive care for childbearing women.

Conclusion

The BSS-R has demonstrated itself to be a theoretically anchored and psychometrically robust measure of the important concept of birth satisfaction.

Importantly, in terms of birth satisfaction, it was found that there was little difference in birth satisfaction between women who planned or did not plan their pregnancy, which suggests minimal impact of planned pregnancy status on birth satisfaction. This finding challenges the almost universal negative perspective ascribed to unplanned pregnancy, with “unplanned” not necessarily equating to “unwanted”. Confidence in the reliability of these observations is forthcoming from the exemplary invariance characteristics of the tool.

References

1. Sawyer A, Ayers S, Abbott J, Gyte G, Rabe H, Duley L. Measures of satisfaction with care during labour and birth: a comparative review. BMC Pregnancy Childbirth. 2013;13:108.

2. Goodman P, Mackey MC, Tavakoli AS. Factors related to childbirth satisfaction. J Adv Nurs. 2004;46(2):212-9.

3. Harvey S, Rach D, Stainton M, Jarrell J, Brant R. Evaluation of satisfaction with midwifery care. Midwifery. 2002;18:260-7.

4. Hodnett ED, Simmons-Tropea D. The labour agentry scale: psychometric properties of an instrument measuring control during childbirth. Research in Nursing and Health. 1987;10:301-10.

5. Redshaw M, Martin CR. Validation of a perceptions of care adjective checklist. J Eval Clin Pract. 2009;15(2):281-8.

6. Hollins Martin CJ, Fleming V. The birth satisfaction scale. Int J Health Care Qual Assur. 2011;24(2):124-35.

7. Hollins Martin CJ, Martin CR. Development and psychometric properties of the Birth Satisfaction Scale-Revised (BSS-R). Midwifery. 2014;30(6):610-9.

8. Barbosa-Leiker C, Fleming S, Hollins Martin CJ, Martin CR. Psychometric properties of the Birth Satisfaction Scale-Revised (BSS-R) for US mothers. Journal of Reproductive and Infant Psychology. 2015;33(5):504-11.

9. Cetin FC, Sezer A, Merih YD. The birth satisfaction scale: Turkish adaptation, validation and reliability study. Northern Clinics of Istanbul. 2015;2(2):142-50.

10. Vardavaki Z, Hollins Martin CJ, Martin CR. Construct and content validity of the Greek version of the Birth Satisfaction Scale (G-BSS). Journal of Reproductive and Infant Psychology. 2015;33(5):488-503.

11. Gorsuch RL. Factor Analysis. 2nd ed. Hillsdale, NJ: Erlbaum; 1983.

12. Byrne BM. Structural Equation Modeling with AMOS: Basic Concepts, Applications and Programming. 2nd ed. New York: Routledge/Taylor and Francis Group; 2010.

13. Martin CR, Hollins Martin CJ, Redshaw M. Development and measurement characteristics of the Birth Satisfaction Scale-Revised Indicator (BSS-RI). Society of Reproductive and Infant Psychology Annual Conference, 13th-14th September, 2016; 13th-14th September, 2016; Leeds, UK2016.

14. Brown T. Confirmatory Factor Analysis for Applied Research. 2nd ed. New York: Guilford Press; 2015.

15. Martin CR, Vardavaki Z, Hollins Martin CJ. Measurement equivalence of the Birth Satisfaction Scale-Revised (BSS-R): further evidence of construct validity. Journal of Reproductive and Infant Psychology. 2016:1-9.

16. Werneke U, Goldberg DP, Yalcin I, Ustun BT. The stability of the factor structure of the General Health Questionnaire. Psychol Med. 2000;30(4):823-9.

17. Goodwin MM, Gazmararian JA, Johnson CH, Gilbert BC, Saltzman LE. Pregnancy intendedness and physical abuse around the time of pregnancy: findings from the pregnancy risk assessment monitoring system, 1996-1997. PRAMS Working Group. Pregnancy Risk Assessment Monitoring System. Matern Child Health J. 2000;4(2):85-92.

18. Boden JM, Fergusson DM, Horwood LJ. Outcomes for Children and Families Following Unplanned Pregnancy: Findings from a Longitudinal Birth Cohort. Child Indicators Research. 2015;8(2):389-402.

19. Singh S, Sedgh G, Hussain R. Unintended pregnancy: worldwide levels, trends, and outcomes. Stud Fam Plann. 2010;41(4):241-50.

20. Cheng D, Schwarz EB, Douglas E, Horon I. Unintended pregnancy and associated maternal preconception, prenatal and postpartum behaviors. Contraception. 2009;79(3):194-8.

21. McCrory C, McNally S. The effect of pregnancy intention on maternal prenatal behaviours and parent and child health: results of an irish cohort study. Paediatr Perinat Epidemiol. 2013;27(2):208-15.

22. Shah PS, Balkhair T, Ohlsson A, Beyene J, Scott F, Frick C. Intention to become pregnant and low birth weight and preterm birth: a systematic review. Matern Child Health J. 2011;15(2):205-16.

23. Nelson JA, O'Brien M. Does an Unplanned Pregnancy Have Long-Term Implications for Mother–Child Relationships? Journal of Family Issues. 2012;33(4):506-26.

24. Mohllajee AP, Curtis KM, Morrow B, Marchbanks PA. Pregnancy intention and its relationship to birth and maternal outcomes. Obstet Gynecol. 2007;109(3):678-86.

25. Jack BW, Atrash H, Coonrod DV, Moos MK, O'Donnell J, Johnson K. The clinical content of preconception care: an overview and preparation of this supplement. Am J Obstet Gynecol. 2008;199(6 Suppl 2):S266-79.

26. Gardiner PM, Nelson L, Shellhaas CS, Dunlop AL, Long R, Andrist S, et al. The clinical content of preconception care: nutrition and dietary supplements. Am J Obstet Gynecol. 2008;199(6 Suppl 2):S345-56.

27. Papachatzi E, Dimitriou G, Dimitropoulos K, Vantarakis A. Pre-pregnancy obesity: maternal, neonatal and childhood outcomes. J Neonatal Perinatal Med. 2013;6(3):203-16.

28. Thorpe PG, Gilboa SM, Hernandez-Diaz S, Lind J, Cragan JD, Briggs G, et al. Medications in the first trimester of pregnancy: most common exposures and critical gaps in understanding fetal risk. Pharmacoepidemiol Drug Saf. 2013;22(9):1013-8.

29. Coonrod DV, Jack BW, Stubblefield PG, Hollier LM, Boggess KA, Cefalo R, et al. The clinical content of preconception care: infectious diseases in preconception care. Am J Obstet Gynecol. 2008;199(6 Suppl 2):S296-309.

30. Workowski KA, Berman S, Centers for Disease C, Prevention. Sexually transmitted diseases treatment guidelines, 2010. MMWR Recomm Rep. 2010;59(RR-12):1-110.

31. Guerin A, Nisenbaum R, Ray JG. Use of maternal GHb concentration to estimate the risk of congenital anomalies in the offspring of women with prepregnancy diabetes. Diabetes Care. 2007;30(7):1920-5.

32. Qualtrics. Provo, Utah, UT: Qualtrics; 2015.

33. Fleming SE, Donovan-Batson C, Burduli E, Barbosa-Leiker C, Hollins Martin CJ, Martin CR. Birth Satisfaction Scale/Birth Satisfaction Scale-Revised (BSS/BSS-R): A large scale United States planned home birth and birth centre survey. Midwifery. 2016;41:9-15.

34. Kline P. The Handbook of Psychological Testing. London: Routledge; 1993.

35. Kline P. A Psychometrics Primer. London: Free Association Books; 2000.

36. Bentler PM, Bonett DG. Significance tests and goodness of fit in the evaluation of covariance structures. Psychological Bulletin. 1980;88:588-606.

37. Bentler PM. Comparative fit indexes in structural models. Psychol Bull. 1990;107(2):238-46.

38. Hu LT, Bentler PM. Evaluating model fit. In: Hoyle RH, editor. Structural Equation Modelling: Concepts, Issues and Applications. Thousand Oaks, CA: Sage; 1995.

39. Hu LT, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. Structural Equation Modeling. 1999;6:1-55.

40. Browne MW, Cudeck R. Alternate ways of assessing model fit. In: Bollen KA, Long JS, editors. Testing Structural Equation Models1993.

41. Schumacker RE, Lomax RG. A Beginner's Guide to Structural Equation Modelling. 3rd ed. New York: Routledge/Taylor and Francis Group; 2010.

42. Steenkamp J-BEM, Baumgartner H. Assessing measurement invariance in cross-national consumer research. Journal of Consumer Research. 1998;25(1):78-90.

43. Kline RB. Principles and Practice of Structural Equation Modeling. 3rd ed. London: Guilford Press; 2011.

44. Byrne BM, Shavelson RJ, Muthen B. Testing for the equivalence of factor covariance and mean structures: The issue of partial measurement invariance. Psychol Bull. 1989;105:456-66.

45. Burns GL, Walsh JA, Gomez R, Hafetz N. Measurement and structural invariance of parent ratings of ADHD and ODD symptoms across gender for American and Malaysian children. Psychol Assess. 2006;18(4):452-7.

46. Cyders MA. Impulsivity and the sexes: measurement and structural invariance of the UPPS-P Impulsive Behavior Scale. Assessment. 2013;20(1):86-97.

47. Dong L, Wu H, Waldman ID. Measurement and structural invariance of the antisocial process screening device. Psychol Assess. 2014;26(2):598-608.

48. Hildebrandt A, Sommer W, Herzmann G, Wilhelm O. Structural invariance and age-related performance differences in face cognition. Psychol Aging. 2010;25(4):794-810.

49. Milfont TL, Fischer R. Testing measurement invariance across groups: Applications in cross-cultural research. International Journal of Psychological Research. 2010;3(1):111-21.

50. Bollen KA. A new incremental fit index for general structural equation models. Sociological Methods and Research. 1989;17:303-16.

51. Cheung GW, Rensvold RB. Evaluating Goodness-of-Fit Indexes for Testing Measurement Invariance. Structural Equation Modeling: A Multidisciplinary Journal. 2002;9(2):233-55.

52. Barton B, Peat J. Medical Statistics: A Guide to SPSS, Data Analysis and Critical Appraisal. . 2nd ed. Chichester: Wiley-Blackwell; 2014.

53. Cohen J. Statistical Power Analysis for the Behavioral Sciences. 2nd ed. Hillsdale, NJ: Lawrence Erlbaum; 1988.

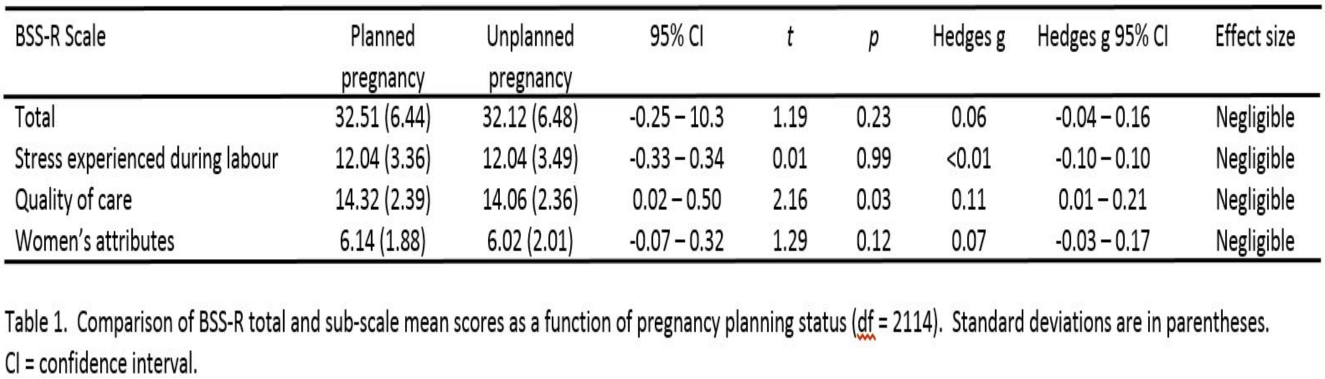
54. Team RC. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing; 2015.

55. Mahalanobis PC. On the generalised distance in statistics. Proceedings of the National Institute of Sciences of India. 1936;2(1):49-55.

56. Faller H. [Significance, effect size, and confidence interval]. Rehabilitation (Stuttg). 2004;43(3):174-8.

57. Marsh HW, Hau K-T. Assessing goodness of fit: Is parsimony always desirable? Journal of Experimental Education. 1996;64(4):363-90.

58. Fielding NG, Lee RM, Blank G. The Sage Handbook of Online Research Methods. London: Sage; 2008.



BSS-R Scale	Planned pregnancy	Unplanned pregnancy	95% CI	*t*	*p*	Hedges g	Hedges g 95% CI	Effect size
Total	32.51 (6.44)	32.12 (6.48)	-0.25 – 10.3	1.19	0.23	0.06	-0.04 – 0.16	Negligible
Stress experienced during labour	12.04 (3.36)	12.04 (3.49)	-0.33 – 0.34	0.01	0.99	<0.01	-0.10 – 0.10	Negligible
Quality of care	14.32 (2.39)	14.06 (2.36)	0.02 – 0.50	2.16	0.03	0.11	0.01 – 0.21	Negligible
Women's attributes	6.14 (1.88)	6.02 (2.01)	-0.07 – 0.32	1.29	0.12	0.07	-0.03 – 0.17	Negligible

Table 1. Comparison of BSS-R total and sub-scale mean scores as a function of pregnancy planning status (df = 2114). Standard deviations are in parentheses. CI = confidence interval.