

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. DOI: 10.1080/02646838.2017.1310374

Cross-Cultural Invariance of the Birth Satisfaction Scale-Revised (BSS-R): Comparing UK and US Samples

Ekaterina Burduli ^{a*}, Celestina Barbosa-Leiker ^b, Susan Fleming ^c, Caroline J. Hollins Martin ^d, and Colin R. Martin ^e.

^aInitiative for Research and Education to Advance Community Health (IREACH), Elson S. Floyd College of Medicine, Washington State University, Spokane, WA, USA; ^bCollege of Nursing, Washington State University, Spokane, WA, USA; ^cCollege of Nursing, Seattle University, Seattle, WA; ^dSchool of Nursing, Midwifery and Social Care, Edinburgh Napier University (Sighthill Campus), Midlothian, Scotland, UK; ^eFaculty of Society and Health, Buckinghamshire New University, Uxbridge, England, UK

*Corresponding author: Ekaterina Burduli, Washington State University, Initiative for Research and Education to Advance Community Health (IREACH), Elson S. Floyd College of Medicine, PO Box 1495 Spokane, WA, USA, 99210-1495; Phone: +1-509-358-7794; Fax: +1-509-324-7341; Email: eburduli@wsu.edu

We thank G. Leonard Burns, PhD, and Craig D. Parks, PhD, for assistance with invariance analyses and for observations that greatly improved the manuscript.

Cross-Cultural Invariance of the Birth Satisfaction Scale-Revised (BSS-R): Comparing UK and US Samples

Abstract

Objective: This research sought to test the measurement invariance of the Birth Satisfaction Scale-Revised (BSS-R) across United States (US) and United Kingdom (UK) samples. Multiple-group measurement was tested and latent means analysis compared levels of birth satisfaction across the samples. **Method:** Using Confirmatory Factor Analysis (CFA), data previously collected from 409 mothers (181 US mothers; 228 UK mothers) were used to examine the multiple-group measurement invariance of the BSS-R across US and UK samples. **Results:** A correlated factors BSS-R model demonstrated partial measurement invariance. US mothers had significantly lower birth satisfaction levels on the three BSS-R subscales. **Conclusions:** This research demonstrates that the BSS-R is a robust tool that can be used to reliably measure women's birth satisfaction within and across the US and UK.

Keywords: Childbirth, Birth Satisfaction, Measurement Invariance

Cross-Cultural Invariance of the Birth Satisfaction Scale-Revised (BSS-R): Comparing UK and US Samples

Recent healthcare advances around the world emphasize providing exceptional health care service that in turn leads to greater patient satisfaction. The development of the Affordable Care Act in the US highlights the need for affordable patient care of the highest quality (Patient Protection and Affordable Care Act, 2010; Sakala, 2010). Additionally, the Institute of Healthcare Improvement (IHI) recently formed “The Triple Aim”: a framework designed to enhance the patient care experience, improve overall health, and reduce costs (IHI, 2014). In the UK, the Department of Health (DH) continuously emphasized enriching childbearing women’s care by placing women at the center of care provision, and pledging to focus on the needs of childbearing women by enhancing choice and control for those women (DH, 2004). These goals were further underlined in later DH publications ‘our health, our care, our say’ (2006) and ‘Maternity matters: choice, access and the continuity of care in a safe service’ (DH, 2007).

Assessing birth satisfaction is especially relevant in the US. Compared to other industrialized countries, maternity care in the US is characterized by poor outcomes and expensive practices (Bezruchka, 2012; Clark, Belfort, Byrum, Meyers, & Perlin, 2008; Stones & Arulkumaran, 2014). For example, pregnancy and childbirth care are currently considered a more expensive hospital stay than major conditions such as coronary artery disease (Weir & Andrews, 2011). Additionally, the US underutilizes midwifery-led maternal care, which is known to be associated with low-severity, cost-saving procedures that often result in more positive health outcomes (Martin, Hamilton, Curtin, & Matthews, 2013; Renfrew, Homer, Downe, McFadden, Muir, & Prentice, 2014). Alternatively, the majority of maternity units in UK hospitals are run by midwives (Barbosa-Leiker, Fleming, Hollins Martin, & Martin, 2015;

Renfrew et al., 2014). In addition, more births are carried out in birth centers rather than hospitals in the UK, leading to greater satisfaction with birth, better maternal health outcomes, and lower overall costs (Renfrew et al., 2014). In contrast, over 98% of births in the US currently occur in a hospital setting; however, only 7.6% of those births are attended by certified nurse midwives (Martin et al., 2013). Because the vast majority of US mothers give birth in a hospital setting, it is more likely that their birth will be physician dominated. As a result, mothers are less likely to feel in control of their labor, and in turn, less satisfied with their birth. (Goodman, Mackey, & Tavakoli, 2004; Sawyer, Ayers, Abbott, Gyte, Rabe, & Duley, 2013; Knapp, 1996). Another contributing factor to the maternal health decline in the US may be its present rate of caesarean deliveries. At 32.8%, the US has one of the highest rates of caesarean deliveries in the world (Betran et al., 2007; Bezruchka, 2012; Martin et al., 2013). Meanwhile, UK's 24.8% rate of caesarean delivery is significantly lower (NHS, 2011). Lower rates of caesarean delivery are associated with greater maternal satisfaction, lower costs, and better health outcomes (Harvey et al., 2002; Renfrew et al., 2014).

In order to provide care of the highest quality and improve maternal health outcomes, it is essential to assess birth satisfaction. Because satisfaction is one of the most widely reported outcome measures of quality of care, an evaluation of maternal birth satisfaction can be considered a direct measure of maternal care quality (Goodman et al., 2004; Sawyer et al., 2013). However, an accurate assessment of maternal care largely depends on the use of reliable and valid measures of birth satisfaction.

The Birth Satisfaction Scale – Revised (BSS-R), a quantitative measure examining women's satisfaction with labour experiences and outcomes was developed and psychometrically validated in the UK (Hollins Martin & Martin, 2014) as a short-form version of

the thematically-developed Birth Satisfaction Scale (BSS; Hollins Martin & Fleming, 2011). Recently the BSS-R has been validated in the US (Barbosa-Leiker et al., 2015) as well as in Greece (Vardavaki, Hollins Martin & Martin, 2015). Moreover, the tool's measurement invariance was also demonstrated across UK and Greek samples, providing further evidence of the tool's validity and cross-cultural utility (Vardavaki, Martin, & Hollins Martin, 2016). Prior to utilizing the BSS-R in the US, and comparing birth satisfaction levels between the US and UK, it is necessary to examine the measurement invariance of the BSS-R across these two cultures.

Measurement invariance tests the equivalence of the measured construct (i.e., birth satisfaction) between groups of interest (i.e., US vs. UK) to determine that identical constructs are being measured between the groups (Brown, 2015; Chen, Sousa, & West, 2005). Without a prior demonstration of measurement invariance for scales measuring birth satisfaction, it is impossible to determine whether any observed differences in satisfaction levels are due to true differences or due to differences in how the particular scale assesses birth satisfaction in the US and the UK. Such an assessment will shed light on the usefulness and accuracy of the BSS-R within and across cultures. Having a cross-culturally validated measure of birth satisfaction will serve as an essential tool in improving women's health outcomes, reducing costs, and meeting national and international health goals.

The aims of this study were twofold:

1. Test the measurement invariance of the BSS-R across UK and US mothers.
2. If adequate measurement invariance is demonstrated (i.e., scalar invariance), compare latent factor means across US and UK mothers.

Method

Study Population

UK Sample. The UK sample consists of 228 postnatal women (48% first time mothers) receiving care from maternity service providers in the West of Scotland (UK), within 10 days of giving birth (Hollins Martin & Martin, 2014). Mothers with medical conditions, problematic obstetric history, pre-term (< 37 weeks), post-term (> 42 weeks), and mothers under 16 years of age or over 50 years of age were excluded from participation in the study. Data for this sample was collected between October 2010 and January 2011 (Hollins Martin & Martin, 2014).

Detailed sample information is available in Hollins Martin & Martin, 2014.

US Sample. Data collection for this sample took place in 2013. One hundred and eighty-one first-time mothers were recruited electronically via personal and professional contacts to participate in the original study. Participants in this study had to be 18 years old, and be English-speaking. Participants also had to have given birth in a US hospital or birthing center in the past 5 years.

While the UK sample was assessed within 10 days of delivery, the US sample rated their satisfaction of birth within five years of delivery. Nevertheless, birth is considered a seminal event and research has established that mothers can recall their birthing experience even 20 years later with astonishing precision (Simkin, 1991).

This study received certification of exemption from the participating universities' Institutional Review Boards. Written, informed consent was given by all participating women.

Measure

BSS-R. Developed in the UK, the BSS-R is a 10-item, self-report scale that was reduced from the original 30-item BSS, and assesses women's perceptions of birth to determine women's

satisfaction with their birth experience (Hollins Martin & Fleming, 2011; Hollins Martin et al., 2012; Hollins Martin & Martin, 2014). The BSS-R (Appendix A) is a Likert-type scale asking participants to rate their level of agreement with each item (4 = Strongly agree, 3 = Agree, 2 = Neither Agree or Disagree; 1 = Disagree; 0 = Strongly Disagree). Four of the items are reverse-coded (e.g. “I found giving birth a distressing experience”). The scale consists of one, higher-order factor, *experience of childbearing*, containing three lower-order factors: *quality of care provision* (four items reflecting home assessment, birth environment, support, and relationships with health care professionals); *women’s personal attributes* (two items reflecting ability to cope during labour, feeling in control, childbirth preparation, and relationship with baby); and *stress experienced during labour* (four items reflecting distress, obstetric injuries, receiving sufficient medical care, obstetric intervention, pain, long labour, and baby’s health) (Hollins Martin & Fleming, 2011) (figure 1). The BSS-R can be scored to provide three separate subscale scores, as well as a total score reflecting overall satisfaction with birth. To make the BSS-R culturally relevant to US mothers, two primary changes were previously made to the scale. First, both spellings of “labour” in the scale items were included (e.g. “I was not distressed at all during labour/labor”). Next, an item was added at the end of the scale that used a different term for “unscathed” (“I came through childbirth virtually unscathed”). Researchers thought that this term was not often used in the US so the item “I came through childbirth virtually unharmed” was included. Results indeed showed that US mothers responded significantly differently when asked if they came through childbirth unscathed vs. unharmed; thus, researchers recommend using “I came through childbirth virtually unharmed” in US samples in order to get more precise BSS-R scores (Barbosa-Leiker et al., 2015).

Analytic strategy

The one-factor, higher-order BSS-R model containing three lower-order factors, as well as a three, correlated factor BSS-R model (figure 2) were compared across the samples using Confirmatory Factor Analysis (CFA). Descriptive statistics were examined using IBM SPSS version 22 (IBM Corporation, 2013). Factor analyses were conducted using Mplus version 6.1 (Muthen & Muthen, 1998-2012). The Full Information Maximum likelihood Robust estimation (FIMLR) was used to estimate data and account for any abnormal and missing data patterns.

Model Evaluation. Several goodness of fit indices were used to evaluate the fit of the models. The Chi-Square Statistic (χ^2) evaluates the absolute fit of the hypothesized model to the data; however, this statistic is sensitive to sample sizes and abnormally distributed data (Bollen & Long 1993; Hu & Bentler, 1999). Therefore additional, more robust fit indices were employed: The Comparative Fit Index (CFI), the Standardized Root Mean Square Residuals (SRMR), and the Root Mean Square Error of Approximation (RMSEA). Substantial research recommends the following cut-off criteria: A good model fit is illustrated when the CFI is close to or above .95, the SRMR is close to or below .06 and the RMSEA is close to or below .08 (Hu & Bentler, 1999; Marsh, 2004).

Similarly to the χ^2 statistic, the χ^2 difference test may also be influenced by sample size and multivariate abnormalities (Chen et al., 2005); therefore decrement in fit across the invariance models was examined by evaluating additional fit indices in accordance with current guidelines. A decrement in CFI between comparison and nested models of > 0.010 , in addition to a change in RMSEA of $\geq .015$ and a change in SRMR of ≥ 0.030 (for loading invariance) and ≥ 0.010 (for intercept invariance) indicates non-invariance (Chen, 2007; Cheung & Rensvold, 2002; Wang & Wang, 2012).

Measurement Invariance. Original BSS-R validation research supported a multi-dimensional classification of the scale with three correlated factors (e.g., Hollins Martin & Fleming, 2011; Hollins Martin et al., 2012). Later work allowed for a hierarchical classification of the BSS-R with the three factors being subsumed under one, higher-order factor (e.g., Barbosa-Leiker et al., 2015; Hollins Martin & Martin, 2014). In this study, the higher-order model as well as the three, correlated-factors model provided equivalently good model fit, as well as a better model fit than a one-factor, or two-factor models in these samples. Therefore, this study compared both the higher-order model, as well as the three correlated factors BSS-R model across the samples.

Evaluating measurement invariance of a higher-order model consists of testing metric invariance of both the lower-order and higher-order factors, and by testing scalar invariance of both the measured variables and the lower-order factors (Chen et al., 2005). As suggested by Chen and colleagues (2005), the following testing order was executed: Configural invariance examining equivalent factor structures across the groups was evaluated first. This is the most basic level of invariance and it tests to see if similar, though not necessarily identical, factors exist in the two groups (Brown, 2015; Vandenberg & Lance, 2000). Next, first-order metric invariance testing the equivalence of the lower-order item-factor loadings across groups was examined. Invariant item-factor loadings indicate that the underlying factors' units of measurement are identical across the groups. Third, higher-order metric invariance testing the equivalence of the higher-order factor loadings across groups was evaluated. Invariant higher-order factor loadings indicate that the units of measurement of the underlying higher-order factor are identical across the groups. If previous invariance steps hold, lower-order scalar invariance would be examined next. This step constrains lower-order indicator loadings as well as indicator

intercepts equal across groups to test if, for any given lower-order factor value, the observed values for indicators are statistically equivalent across the two groups (Brown, 2015; Chen et al., 2005). Subsequently, higher-order scalar invariance would be tested. This level of invariance implies that scores from different groups have the same units of measurement (factor loading) as well as the same origin (intercept) for the higher-order factor. Scalar invariance indicates a common origin for the scale across the groups. Therefore succeeding this step, the higher-order latent factor mean can be compared across groups (Chen et al., 2005).

Similarly to the higher-order BSS-R invariance steps, invariance analyses for the three correlated factors BSS-R model included configural invariance first, followed by metric invariance, and concluding with scalar invariance. However, because this model is tri-dimensional with the three factors correlated instead of being subsumed under one, higher-order factor, the above tests are performed only once. Following the verification of scalar invariance, latent means of the three factors were compared across the samples.

More advanced testing such as strict invariance, factor variance invariance, and factor covariance invariance was not pursued in this study because these criteria are often exceptionally challenging to achieve. Therefore, configural invariance, metric invariance, and scalar invariance are the most frequently tested forms of measurement invariance. In addition, comparing latent factor means across US and UK mothers were of utmost interest, and in order to do so, it is sufficient to demonstrate scalar invariance.

Partial Measurement Invariance. Partial invariance refers to circumstances in which not all model parameters are constrained statistically equal across groups (Brown, 2015; Byrne, Shavelson, & Muthen, 1989). If full invariance was not demonstrated at any given step of the

measurement invariance process, partial invariance was considered based on strong theoretical justification and information obtained from the examination of model fit and modification indices. Partial metric invariance is deemed acceptable as long as at least one loading other than the marker indicator loading demonstrates invariance across the samples (Brown, 2015; Byrne et al., 1989). Similarly, partial scalar invariance is deemed acceptable as long as at least one item intercept other than the marker indicator intercept demonstrates invariance (Brown, 2015; Byrne et al., 1989). Latent means comparisons require at least two invariant indicators per factor (Brown, 2015; Steenkamp & Baumgartner, 1998).

Results

Descriptive Statistics

Descriptive statistics including means, standard deviations, skewness, and kurtosis scores of the BSS-R items are presented in Table 1. None of the items showed high skewness and kurtosis scores (>3 for skewness and >10 for kurtosis), indicating a fairly normal distribution. Cronbach's alpha values were excellent with $\alpha = .88$ for the total scale; Stress subscale $\alpha = .72$; Women's Attributes subscale $\alpha = .70$; and Quality of Care subscale $\alpha = .93$.

Higher-Order BSS-R

Invariance Models. Table 2 shows standardized loadings and intercepts at the configural invariance step. Lower-order and higher-order loadings were statistically significant and largely substantial ranging from .41 to .91. Table 3 shows the fit of the baseline models (i.e., three lower-order factors subordinate to a single higher-order factor) for each country separately. The χ^2 statistics for the baseline models were not statistically significant, and all of the subjective fit indices (CFI values $\geq .991$, SRMR values $\leq .048$, and RMSEA values $\leq .032$) met recommended

criteria, suggesting that the second-order BSS-R model represented a good fit within each country. Table 3 also shows the fit of the higher-order BSS-R invariance models across the samples. Tests of configural invariance, and first-order metric invariance showed good model fit (CFI values $\geq .987$, SRMR values $\leq .064$, and RMSEA values $\leq .031$). While the test of second-order metric invariance showed acceptable model fit (CFI = .974, SRMR = .090, and RMSEA = .043), when compared to the previous model (i.e., first-order metric invariance) it revealed a significant decrement in model fit (Δ CFI = -.013, SRMR = +.026, Δ RMSEA = +.012). This indicates that the units of measurement of the higher-order birth satisfaction factor are not equivalent across the samples. Inspection of the model modification indices does not present a clear problem area in the model. Results suggest that the higher-order portion of the BSS-R model is demonstrating a lack of measurement invariance across the UK and US, leading us to examine the correlated, three-factor BSS-R model across the samples.

Three-Factor BSS-R

Invariance Models. Table 4 shows the fit of the baseline model (i.e., three, correlated factors) for each country separately. The χ^2 statistics for the baseline models were not statistically significant, and all of the subjective fit indices met recommended criteria (CFI values $\geq .99$, SRMR values $\leq .048$, and RMSEA values $\leq .031$), suggesting that the correlated, three-factor BSS-R model represented a good fit within each country. Table 4 also shows the fit of the three-factor BSS-R invariance models across the samples. Tests of configural, and metric invariance also showed good model fit (CFI values $\geq .98$, SRMR values $\leq .064$, RMSEA values $\leq .031$). However, the test of scalar invariance showed poor model fit (CFI = .54, SRMR = .406, RMSEA = .174), and when compared to the previous model (i.e., metric invariance) it resulted in a significant decrement in model fit (Δ CFI = -.443, SRMR = +.342, Δ RMSEA = +.143).

Inspection of the modification indices revealed that four of the 10 BSS-R items need to be released from equality constraints. In the case of partial scalar invariance, releasing the non-invariant item intercepts from equality constraints for all further invariance analyses is recommended (Brown, 2015). Following the release of the four item intercepts, all changes across the sequential models met the criteria for measurement invariance. Following the confirmation of partial scalar invariance, the three latent factor means were compared across the samples. Results showed that the US sample had significantly lower birth satisfaction levels than the UK sample, on the stress scale (US latent mean = 2.01; UK latent mean = 3.25; $z = -0.94$, $p = .000$), the quality of care scale (US latent mean = 1.10; UK latent mean = 6.22; $z = -2.07$, $p = .000$), and the women's attributes scale (US latent mean = 2.24; UK latent mean = 3.67; $z = -0.63$, $p = .000$). Correlations among the three factors were statistically significant and largely substantial for both samples, ranging from .30 to .93.

Discussion

This study sought to examine the measurement invariance of the BSS-R across US and UK samples, and to compare the levels of UK and US mothers' birth satisfaction on the three BSS-R scales. Results indicate that when comparing birth satisfaction across the US and UK, it is best to examine the three-factor BSS-R model and look at mean levels of each BSS-R sub-scale, in place of the higher-order *experience of childbearing* factor. While the BSS-R has been conceptualized as a hierarchical structure proposing to provide a total BSS-R score alongside the three subscale scores (Barbosa-Leiker et al, 2015; Hollins Martin & Martin, 2014), this study shows that the uni-dimensional structure of the BSS-R is only plausible within the US and UK samples separately. When comparing birth satisfaction across these two cultures, it may only be

feasible to compare mean birth satisfaction levels on the three subscales and not the mean higher-order satisfaction score of the BSS-R.

Initial development and validation research of the Birth Satisfaction Scale indeed support the existence of three correlated, yet distinct birth satisfaction domains (Hollins Martin & Fleming, 2011; Hollins Martin, Snowden, Martin, 2012; Hollins Martin & Martin, 2014). Hollins Martin and colleagues originally developed the 30-item long form BSS by transcribing evidence-based reports of women's birth satisfaction into written statements that identified three related, yet theoretically distinct domains that affect 'birth satisfaction' (Hollins Martin & Fleming, 2011). The existence of the three birth satisfaction domains was further supported through a Concurrent Validity Analysis (Hollins Martin et al., 2012). Most recently, a Greek version of the BSS-R was also validated, offering further support for a correlated factor BSS-R model and not a uni-dimensional structure.

Assessing measurement invariance of the correlated BSS-R model across the US and UK resulted in partial measurement invariance. The correlated BSS-R model demonstrated equal factor structure and equal units of measurement across the US and UK samples through the tests of configural and metric invariance. However, the test of scalar invariance showed a significant decrement in model fit when compared to the previous model (i.e., metric invariance). This means that at a constant level of the factors, the item means differ across the samples. Inspection of the modification indices revealed that four out of the 10 BSS-R items need to be released from equality constraints. The four non-invariant items included two Quality of Care subscale items: *The delivery room was clean and hygienic; The delivery room staff encouraged me to make decisions about how I wanted my birth to progress*, and two Stress Experienced during Labour

subscale items: *I was not distressed at all during labour/labor; I thought my labour/labor was excessively long.*

In case of partial invariance, ample research recommends proceeding to subsequent invariance steps by releasing non-invariant items from further analyses (Brown, 2015). Partial scalar invariance as is evident in this research is deemed acceptable as long as at least one item intercept other than the marker indicator intercept demonstrates invariance (Brown, 2015; Byrne, Shavelson, & Muthen, 1989). In this research, while four of the items do not have statistically equivalent scores (intercepts) for the same level on the latent factors, six of the item intercepts do demonstrate scalar invariance. Thus, invariance tests were continued with the non-invariant item intercepts released from constrains in further invariance analyses.

The lack of scalar invariance of some of the items may reflect sample differences. Specifically, all of the mothers in the US sample were first-time mothers, whereas the UK sample consisted of a mixture of newer and older mothers. Being a first-time mother may influence one's expectation about how the birth should proceed, and this in turn may affect a mother's perception of her birth experience. Differences on the items may also reflect the differing health economies and outcomes between the US and UK. Maternal health in the US is characterized by expensive medical practices and poorer outcomes possibly due to high caesarean section rates, under-utilization of midwives, and higher rates of hospital births (Bezruchka, 2012; Clark et al., 2008; Harvey et al., 2002; MacDorman & Matthews, 2009; Martin et al., 2013; Renfrew et al., 2014; Stones & Arulkumaran, 2014). Additionally, because births in the US are physician dominated (versus midwife dominated in the UK), it is possible that US mothers may feel they have less say in their birth (Harvey et al., 2002; Martin et al., 2013; Renfrew et al., 2014). On the other hand, US mothers may respond differently to the non-

invariant item *I was not distressed at all during labour/labor* because of the higher rates of epidural use (67%) in the US versus the UK (~30%) (Declercq, Sakala, Corry, Applebaum, & Herrlich, 2013; NHS, 2015; McGrady & Litchfield, 2004).

After establishing partial scalar invariance, the three factor means were compared across the samples with US mothers reporting significantly lower birth satisfaction levels than UK mothers on all three BSS-R scales. These cultural differences strongly suggest that higher caesarean rates, costly practices, higher rates of hospital births, and under-utilization of midwives are negatively influencing women's birth satisfaction in the US (Bezruchka, 2012; Clark et al., 2008; Martin et al., 2013; Renfrew et al., 2014). In order to improve maternal health outcomes in the US and meet national and international health goals, it is imperative to reflect on the reasons behind lower birth satisfaction rates in the US and apply these findings to changes in maternal health policy.

This study presents several limitations that are important to note. The divergence of the samples in regards to parity and timeline of the most recent birth may impact the differences found between the samples. Specifically, while the US sample consists of all first-time mothers, only 48% of the UK sample includes first-time mothers. In addition, the UK sample completed the BSS-R within 10 days of giving birth, whereas the US mothers gave birth within the past five years. Finally, recruitment for the two samples occurred several years apart, and during this gap several maternal health care policies were updated. For example, the importance of vaginal deliveries over caesarean deliveries was re-emphasized by the American College of Obstetricians and Gynecologists in the US and by the National Institute for Health and Care Excellence in the UK (The American College of Obstetricians and Gynecologists, Committee on Obstetric Practice, Number 559, 2013; The National Institute for Health and Care Excellence, 2011). In the

future, it is also vital to cross-validate this study in a much larger sample of mothers in the US and the UK. The modest sample sizes in this study may hinder the generalizability of the findings. In the future, in order to establish that the BSS-R assesses birth satisfaction equivalently for different subgroups of the population, it is also of interest to examine the measurement invariance of the BSS-R across age (i.e., younger vs. older mothers), marital status (single vs. married), birth type (vaginal vs. caesarean), and birth setting (Home birth vs. Hospital birth).

Conclusion

This research aimed to test the measurement invariance of the BSS-R and to compare birth satisfaction levels across US and UK mothers. Results demonstrate that the BSS-R is a robust tool that can be used by midwives, obstetricians and maternity care experts to reliably measure women's birth satisfaction within and across the US and UK.

References

- 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. doi: 10.1080/02646838.2015.1024211
- Betran, A.P., Merialdi, M., Lauer, J.A., Bing-Shun, W., Thomas, J., Van Look, P., & Wagner, M. (2007). Rates of caesarean section: Analysis of global, regional and national estimates. *Paediatric and Perinatal Epidemiology, 21*, 98–113.
- Bezruchka, S. (2012). The hurrier I go the behinder I get: The deteriorating international ranking of the U.S. health status. *Annual Review of Public Health, 33*, 157–173. doi:10.1146/annurev-publhealth-031811-124649
- Bollen, K. A., & Long, S. J. (1993) (Eds.). *Testing structural equation models*. Newbury Park: Sage.
- Brown, T. (2015). *Confirmatory factor analysis for applied research*. (2nd ed.). New York: The Guilford Press.
- Byrne, B. M., Shavelson, R. J., & Muthén, B. (1989). Testing for the equivalence of factor covariance and mean structures. *Psychological Bulletin, 105* (3), 456-466. doi:10.1037/00332909.105.3.456
- Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling, 14*, 464-504. doi: 10.1080/10705510701301834.
- Chen, F. F., Sousa, K. H., & West, S. G. (2005). Testing measurement invariance of second-order factor models. *Structural Equation Modeling, 12* (3), 471-492.

- Cheung, G.W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling, 9*, 233–255.
- Clark, S. L., Belfort, M. A., Byrum, S. L., Meyers, J. A., & Perlin, J. B. (2008). Improved outcomes, fewer cesarean deliveries, and reduced litigation: Results of a new paradigm in patient safety. *American Journal of Obstetrics & Gynecology, 199* (2):105, e1-7. doi: 10.1016/j.ajog.2008.02.031.
- Declercq, E. R., Sakala, C., Corry, M. P., Applebaum, S., & Herrlich, A. (2013). Listening to mothers III: Report of the third national U.S. survey of women's childbearing experiences. New York: Childbirth Connection 2013. Available from <http://transform.childbirthconnection.org/reports/listeningtomothers/>
- U.K. Department of Health (2004). *National Services Framework for Children, Young People and Maternity Services* (Publication ref. 3779). Retrieved from https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/199952/National_Service_Framework_for_Children_Young_People_and_Maternity_Services_-_Core_Standards.pdf
- U.K. Department of Health (2007). *Maternity Matters: Choice, access and continuity of care in a safe service* (Publication ref. 7586). Retrieved from http://webarchive.nationalarchives.gov.uk/20130107105354/http://www.dh.gov.uk/prod_consum_dh/groups/dh_digitalassets/@dh/@en/documents/digitalasset/dh_074199.pdf
- Goodman P., Mackey M. C., Tavakoli A. S. (2004). Factors related to childbirth satisfaction. *Journal of Advanced Nursing, 46* (2), 212-219.
- Harvey, S., Rach, D., Stainton, M. C., Jarrell, J., & Brant, R. (2002). Evaluation of satisfaction with midwifery care. *Midwifery, 18* (4), 260-7.

- Hollins Martin, C. J., & Fleming, V. (2011). The birth satisfaction scale. *International Journal of Health Care Quality Assurance, 24*, 124–135.
- Hollins Martin, C. J., & Martin, C. R. (2014). Development and psychometric properties of the Birth Satisfaction Scale-Revised (BSS-R). *Midwifery, 30*, 610–619.
- Hollins Martin, C. J., Snowden, A., & Martin, C. R. (2012). Concurrent analysis: Validation of the domains within the Birth Satisfaction Scale. *Journal of Reproductive and Infant Psychology, 30*, 247–260.
- IBM Corp. Released 2013. IBM SPSS Statistics for Windows, Version 22.0. Armonk, NY: IBM Corp.
- Institute for Healthcare Improvement (IHI). (2014). The IHI Triple AIM. Retrieved July 1, 2014, from <http://www.ihl.org/offerings/Initiatives/TripleAim/Pages/default.aspx>
- Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6*, 1-55.
- Knapp, L. (1996). Childbirth satisfaction: The effects of internality and perceived control. *Journal of Perinatal Education, 5*, 7-17.
- McGrady, E., & Litchfield, K. (2004). Epidural analgesia in labour. *Continuing Education in Anaesthesia, Critical Care, & Pain, 4*(4), 114–117.
- Marsh, H. (2004). In search of golden rules: Comment on hypothesis-testing approaches to setting cutoff values for fit indexes and dangers in overgeneralizing Hu and Bentler's (1999) findings. *Structural Equation Modeling, 11* (3), 320-341.
- Martin, J. A., Hamilton, B. E., Osterman, M. J. K., Curtin, S. C., & Mathews, T. J. (2013). Births: Final data for 2012. *National Vital Statistics Report, 62*, pp. 1–102. Hyattsville, MD: National Center for Health Statistics.

Muthen, L., & Muthen, B. (1998-2012). *Mplus user's guide*. (6 ed.). Los Angeles: Muthen & Muthen.

National Institute for Health and Clinical Excellence. (2011). *Clinical guideline 132: Caesarean section*, 2011. Retrieved from <https://www.nice.org.uk/guidance/cg132>

NHS. (2011). Hospital Episode Statistics. *NHS Maternity Statistics, 2010-2011*. Retrieved from <http://www.hscic.gov.uk/catalogue/PUB03071/nhs-mate-eng-2010-2011-rep.pdf>

NHS. (2015). Hospital Episode Statistics. *NHS Maternity Statistics-England, 2013-2014*. Retrieved from <http://www.hscic.gov.uk/catalogue/PUB16725/nhs-mate-eng-2013-14-summ-repo-rep.pdf>

Patient Protection and Affordable Care Act, 42 U.S.C. § 18001 et seq. (2010).

Renfrew, M., Homer, C., Downe, S., McFadden, A., Muir, N., & Prentice, T. (2014). Midwifery: An executive summary for The Lancet's series. *The Lancet*. Published online at http://download.thelancet.com/flatcontentassets/series/midwifery/midwifery_exec_sum.pdf

Sakala, C. (2010). U.S. health care reform legislation offers major new gains to childbearing women and newborns. *Birth, 37*, 337–340.

Sawyer, A., Ayers, S., Abbott, J., Gyte, G., Rabe, H., & Duley, L. (2013). Measures of satisfaction with care during labour and birth: A comparative review. *BMC Pregnancy and Childbirth, 13*, 108. doi:10.1186/1471-2393-13-108.

Simkin, P. (1991). Just another day in a woman's life? Women's long-term perceptions of their first birth experience, Part I. *Birth, 18*, 203–210.

Steenkamp, J. B. E. M. & Baumgartner, H. (1998). Assessing measurement invariance in cross-national research. *Journal of Consumer Research, 25* (1), 78-90.

- Stones, W., & Arulkumaran, A. (2014). Health-care professionals in midwifery care. *The Lancet*, *384*, 1169–1170.
- Vandenberg R. J., Lance C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods*, *3*, 4–70.
- 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), 448-503.
- Wang, J., & Wang, X. (2012). *Structural equation modeling: Applications using SEM*. (1st ed.). Chichester, UK: John Wiley & Sons.
- Weir, L., & Andrews, R. (2011, March). The national hospital bill: The most expensive conditions by payer, 2008. HCUP Statistical Brief, 107. Retrieved from <http://www.hcup-us.ahrq.gov/reports/statbriefs/sb107.jsp>

Appendix A. Items of the Birth Satisfaction Scale-Revised (BSS-R) 1-factor higher-order model containing 3 lower-order factors.

	Lower-order factors		
	Stress	Quality of care	Women's attributes
The delivery room staff encouraged me to make decisions about how I wanted my birth to progress		X	
I found giving birth a distressing experience (R)	X		
I came through childbirth virtually unharmed	X		
I felt very anxious during my labour/labor and birth (R)			X
I felt out of control during my birth experience (R)			X
I was not distressed at all during labour/labor	X		
I thought my labour/labor was excessively long (R)	X		
I felt well supported by staff during my labour/labor and birth		X	
The staff communicated well with me during labour/labor		X	
The delivery room was clean and hygienic		X	
	Higher-order factor loadings		
	Experience of childbearing		
Stress	X		
Quality of care	X		
Women's attributes	X		

Note. (R) = reverse-coded items

Table 1. *Descriptive statistics of the BSS-R in the US and UK samples*

Items	US			UK		
	<i>M</i> (SD)	Skewness	Kurtosis	<i>M</i> (SD)	Skewness	Kurtosis
Stress						
Unscathed/Unharmed	1.57 (1.28)	.45	-.95	2.40 (1.13)	-.41	-.82
Long	1.43 (1.32)	.65	-.73	2.86 (1.11)	-.94	.27
Distressed	1.71 (1.27)	.41	-.94	2.48 (1.12)	-.45	-.64
Not Distressed	2.43 (1.11)	-.55	-.69	1.95 (1.14)	.15	-.85
Quality of Care						
Encourage	1.60 (1.37)	.45	-.98	3.07 (.84)	-.95	1.38
Support	1.19 (1.20)	.86	-.27	3.47 (.76)	-1.95	5.27
Communicate	1.32 (1.22)	.84	-.32	3.52 (.67)	-1.70	4.25
Clean	.49 (.71)	1.86	4.91	3.69 (.52)	-1.41	1.04
Women's Attributes						
Anxious	1.98 (1.21)	.17	-1.10	2.30 (1.22)	-.37	-.95
Out of Control	1.81 (1.32)	.16	-1.16	2.61 (1.10)	-.67	-.18

Note. *M* = Mean. SD = Standard Deviation.

Table 2. *Configural Invariance US/UK Item-Factor Loadings and Intercepts*

Items	Lower-order Factors		
	Stress	Quality of care	Women's attributes
The delivery room staff encouraged me to make decisions about how I wanted my birth to progress		.83/.59 (1.17/3.68)	
I found giving birth a distressing experience	.72/.73 (1.35/2.22)		
I came through childbirth virtually unscathed/unharmmed	.64/.59 (1.23/2.13)		
I felt very anxious during my labour/labor and birth			.71/.64 (1.65/1.89)
I felt out of control during my birth experience			.85/.75 (1.37/2.38)
I was not distressed at all during labour/labor	.75/.78 (2.19/1.72)		
I thought my labour/labor was excessively long	.53/.37 (1.08/2.59)		
I felt well supported by staff during my labour/labor and birth		.87/.76 (0.99/4.58)	
The staff communicated well with me during labour/labor		.91/.81 (1.08/5.24)	
The delivery room was clean and hygienic		.46/.50 (0.70/7.17)	
Lower-order Factors	Higher-order Factor		
	Experience of childbearing		
Stress	.90/.84		
Quality of care	.65/.41		
Women's attributes	1.00/1.00		

Note. All loadings and Intercepts are standardized. Intercept values are in parentheses.

Table 3. *Invariance of the Higher-Order BSS-R Model*

	χ^2	<i>df</i>	CFI	RMSEA	SRMR	$\Delta\chi^2$	Δdf	ΔCFI	$\Delta RMSEA$	$\Delta SRMR$
US Sample	38.97	33	.991	.032	.039					
UK Sample	36.22	33	.992	.021	.048					
Configural Invariance	75.09	66	.991	.026	.044					
1 st Order Metric Invariance	87.21	73	.987	.031	.064	12.12	7	-.004	+.005	+.020
2 nd Order Metric Invariance	103.08*	75	.974	.043	.090	15.87*	2	-.013	+.012	+.026

Note. χ^2 = Chi-square. *df* = Degrees of Freedom. *CFI* = Comparative Fit Index. *SRMR* = Standardized Root Mean Square Residuals. *RMSEA* = Root Mean Square Error Approximation. * $p \leq .05$

Table 4. *Invariance of the Correlated Three-Factor BSS-R Model*

	χ^2	<i>df</i>	CFI	RMSEA	SRMR	$\Delta\chi^2$	Δdf	ΔCFI	$\Delta RMSEA$	$\Delta SRMR$
US Sample	37.72	32	.991	.031	.038					
UK Sample	35.38	32	.992	.022	.048					
Configural Invariance	73.01	64	.992	.026	.044					
Metric Invariance	85.26	71	.987	.031	.064	12.25	7	-.005	+.005	+.002
Scalar Invariance	563.54*	78	.544	.174	.406	478.28	7	-.443	+.143	+.342
Partial scalar Bssr-10 (intercepts)	298.40*	77	.792	.119	.128	264.14	1	+.248	-.055	-.278
Partial scalar Bssr-10 & 9 (intercepts)	152.03*	76	.929	.070	.093	146.37	1	+.137	-.049	-.035
Partial scalar Bssr-10 & 9 & 2 (intercepts)	118.26*	75	.959	.053	.075	33.77	1	+.030	-.017	-.018
Partial scalar Bssr-10 & 9 & 2 & 3 (intercepts)	94.87	74	.980	.037	.067	23.39	1	+.021	-.016	-.008
Final Scalar vs. Metric						9.61	4	-.007	+.006	+.003

Note. χ^2 = Chi-square. *df* = Degrees of Freedom. *CFI* = Comparative Fit Index. *SRMR* = Standardized Root Mean Square Residuals. *RMSEA* = Root Mean Square Error Approximation. * $p \leq .05$

Figure List:

Figure 1. Higher-order BSS-R Model

Figure 2. Three-Factor BSS-R Model