

**Construct and content validity of the Greek version of the
Birth Satisfaction Scale (G-BSS)**

Zoi Vardavaki¹

Caroline J. Hollins Martin²

&

Colin R.Martin³

¹Midwife in University Hospital Southampton NHS, SO16 6YD, Southampton, UK. E-mail: zoevardavaki@gmail.com

² Professor in Midwifery, School of Nursing, Midwifery and Social Work, University of Salford, UK. M6 6PU. E-mail: C.J.Hollins-Martin@salford.ac.uk

³Professor of Mental Health, Faculty of Society and Health, Buckinghamshire New University, Uxbridge, UK, UB8 1NA. E-mail: colin.martin@bucks.ac.uk

Address for correspondence

Address for correspondence: Professor Colin R Martin, Room 2.11, Faculty of Health and Society, Buckinghamshire New University, Uxbridge Campus
106 Oxford Road, Uxbridge, Middlesex, UB8 1NA, UK. Tel: 01494 522141 Extension 2349; Fax: 01494 603179; Email: colin.martin@bucks.ac.uk

Abstract

Background: Birth Satisfaction' is a term that encompasses a woman's evaluation of her birth experience. The term includes factors such as her appraisal of the quality of care she received, a personal assessment of how she coped, and her reconstructions of what happened on that particular day. Her accounts may be accurate or skewed, yet correspond with her reality of how events unfolded.

Objective: To evaluate properties of an instrument designed to measure birth satisfaction in a Greek population of postnatal women.

Study design: We assessed factor structure, internal consistency, divergent validity and known-groups discriminant validity of the *30-item-Greek-Birth-Satisfaction-Scale-Long-Form (30-item-G-BSS-LF)* and its revised version the *10-item-Greek-BSS-Revised (10-item-G-BSS-R)*, using survey data collected in Athens.

Participants: A convenience sample of healthy Greek postnatal women (n=162) aged 22-46 years who had delivered between 34-42 weeks gestation.

Results: The *30-item-G-BSS-LF* performed poorly in terms of factor structure. The short-form *10-item-G-BSS-R* performed well in terms of measurement replication of the English equivalent version as a multidimensional instrument. The short-form *10-item-G-BSS-R* comprises 3 sub-scales which measure distinct but correlated domains of: (1) *quality of care provision* (4-items), (2) *women's personal attributes* (2-items), and (3) *stress experienced during labour* (4-items).

Key conclusions:

The *10-item-G-BSS-R* is a valid and reliable multi-dimensional psychometric instrument for measuring birth satisfaction in Greek postnatal women.

Key Words: Birth Satisfaction Scale (BSS), childbearing women, Greek, intranatal, labour, psychometric properties

Introduction

Birth satisfaction represents a woman's subjective and uniquely personal evaluation of her birth experience. This complex multi-faceted construct includes elements of perceived quality of care, coping efficacy and reflections of the birth experience as a whole and in context. Birth satisfaction is thus a retrospective reconstruction related directly to the salient events surrounding the experience of birth (Hollins Martin and Fleming, 2011). The woman's individual evaluation of her own birth experience is important, since this may potentially be a potent indicator of perinatal mental health outcome, for example, birth trauma, which would be anticipated to be experienced as a negative event, may be associated with the experience and manifestation of postpartum post-traumatic stress disorder (Fones, 1996) and a consequential enduring impact on mother-child interactions and family relationships (Iles et al., 2011; McDonald et al., 2011). It is also implicitly important to consider the experience of birth from the woman's perspective, a central tenet of care delivery philosophy enshrined in policy as early as the influential Changing Childbirth (Department of Health, 1993) document.

The effective and valid assessment of birth satisfaction as a self-report measure has remained elusive until relatively recently with the advent of the Birth Satisfaction Scale (Hollins Martin and Fleming, 2011). Hollins Martin and Fleming (2011) developed the original English *30-item-Birth-Satisfaction-Scale-Long-Form (30-item-BSS-LF)*, through transcribing evidence-based reports of women's birth satisfaction into statements written in English. Three main areas (sub-scales) that affect 'birth satisfaction' were identified in the literature:

(1) *Quality of Care (QC)* (includes sub-themes (1a) *home assessment*¹, (1b) *birth environment*, (1c) *sufficient support*, and (1d) *relationships with health professionals*).

(2) *Women's Attributes (WA)* (includes sub-themes (2a) *ability to cope in labour*, (2b) *feeling in control*, (2c) *preparation for childbirth*, and (2d) *relationship with baby*).

(3) *Stress Experienced (SE)* (includes sub-scales (3a) *distress during labour*, (3b) *obstetric injuries*, (3c) *perception of sufficient medical care*, (3d) *obstetric intervention*, (3e) *pain experienced*, (3f) *long labour*, and (3g) *health of baby*).

It is of interest that a number of thematic elements that comprise the original BSS subscales from Hollins Martin and Fleming's (2011) review include absolute categorical variables and perceived variables. Obstetric intervention for example, could include unplanned Caesarian Section or a ventouse delivery. In contrast, a long labour could be a perceived longer than anticipated length of labour since there is considerable variability in the 'average' labour length.

A second qualitative assessment of scale content took place using a research method called *Concurrent Analysis* (Hollins Martin et al., 2012). It is worthy of note, that though the BSS was developed by a sophisticated review and thematic appraisal of the literature (Hollins Martin and Fleming, 2011), the concurrent analysis approach to scale content veracity utilised by Hollins Martin et al., (2012) is a relatively novel and recent approach within the validity literature. Consequently, only relatively recently was a psychometric appraisal of the measurement properties of the BSS

¹ Home assessment was described by Hollins Martin and Fleming (2011) in extensive detail in their original BSS paper as critically important in helping women to feel in control of their labour.

conducted (Hollins Martin and Martin, 2014). Two-hundred and twenty-eight women who were less than 10 days postpartum and resident in the West of Scotland (UK) participated in the study to evaluate the measurement characteristics of the BSS. This study resulted in not only the detailing of the fundamental measurement attributes of the original measure, but also the development of a shortened 10-item version of the BSS, the BSS-Revised (BSS-R) which revealed generally excellent psychometric properties and a good statistical fit to empirical data. This version of the tool assessed the three same dimensions as the original BSS, these being but with fewer items and was deemed to be ideal for use in survey studies where birth satisfaction represented an important construct to assess. The *10-item-BSS-R* is thus comprised of 3 sub-scales that measure distinct but correlated domains of: (1) *Quality of Care (QC)* (4-items), (2) *Women's Attributes (WA)* (2-items), and (3) *Stress Experienced (SE)* (4-items).

The use and applicability of the BSS beyond English-speaking countries is both theoretically and clinically appropriate since birth satisfaction represents a construct of relevance to a broad International community (Fleming & Vandermause, 2011). This would also provide the opportunity to examine the relationship of birth satisfaction to a myriad of other meaningful psychological constructs of relevance to perinatal mental health (Alderdice et al., 2013; Jomeen and Martin, 2005; Redshaw and Martin, 2009; Redshaw et al., 2009; Wylie et al., 2011) within the Greek population. To date however, the tool is currently available only in the English language. Once such country where there is contemporary interest in birth satisfaction is Greece. However, the nurturing and evolution of this research area in

Greece is limited by the lack of availability of a suitable assessment measure of the construct.

The purpose of the current investigation was consequently to develop a Greek language version of the BSS and evaluate its psychometric properties in terms of key indices of validity and reliability. Both versions of the original English-language version of the BSS would be translated into Greek for psychometric evaluation, these being the *30-item-Greek-Birth-Satisfaction-Scale-Long-Form (30-item-G-BSS-LF)* and the *10-item-Greek-BSS-Revised (10-item-G-BSS-R)*.

The following research questions were asked:

- (1) Does the *30-item-G-BSS-LF* conform to a uni-dimensional or multidimensional measure consistent with the model of BSS proposed by Hollins Martin and Fleming (2011)?
- (2) Are the thematically embedded and postulated sub-scales of the *30-item-G-BSS-LF* robust and reliable?
- (3) Does the *30-item-G-BSS-LF* and the postulated sub-scales demonstrate acceptable internal consistency and divergent validity?
- (4) Is the known-groups discriminant validity of the *30-item-G-BSS-LF* satisfactory?
- (5) Is the *30-item-G-BSS-LF* the most appropriate and psychometrically valid formulation of the tool?
- (6) Does the shortened version of the *30-item-G-BSS-LF* (the *10-item-G-BSS-Revised*), perform similarly in terms of measurement characteristics as the English-language version of the BSS-R consistent with the model of Hollins Martin and Martin (2014)?

Method

A quantitative survey was carried out using the 30-item-G-BSS-LF. Participants also completed the short-form 10-item-G-BSS-R since this measure is intrinsic to the longer version G-BSS-LF. The study used a cross-sectional design for initial data capture and for most reliability and validity evaluations. Embedded within this cross-sectional design, was a between-subjects design to evaluate the known-groups discriminant validity of both version of the tool. In summary, the study follows a sequential process of instrument evaluation using classical and contemporary psychometric approaches (Byrne, 2010; Kline, 2000) applied to a single cohort that is differentiated by clinical attributes to rigorously evaluate instrument measurement properties.

Ethical approval

Ethics approval was gained from the appropriate organisational structures within the maternity unit and from the Technological Education Institute of Athens (Greece).

Participants

Participants were a convenience sample of healthy Greek postnatal women (n=162) aged 22-46 years (Mean=32; SD=4.75), who had delivered between 34-42 weeks gestation at a maternity unit in Athens. Those who had experienced a stillbirth, perinatal or neonatal death were excluded from taking part due to the impact grief has upon constructions of childbirth (Hollins-Martin & Forrester, 2013).

Translation of the 30-item-BSS-LF into Greek

The scale was translated into Greek by two native Greek midwives, both of whom were fluent in oral and written English. The first author has lived in England for

several years. Post translation, a pilot study was carried out with 5 Greek service users to test comprehension and interpretation of items. In response to feedback, minor adjustments were made and translation procedures repeated where appropriate to further refine the 30-item-G-BSS-LF. Consistent with both the original English-language version of the tool and contemporary questionnaire scoring practice, a number of BSS items are reverse-scored. The reverse scoring protocol and scoring algorithm is identical to the English-language version of the BSS and BSS-R.

The 30-item-G-BSS-LF

Participants responded to statements on the 30-item-G-BSS-LF by circling a 5-point Likert scale. The range of scores is 30-150, with a score of 30 representing least satisfied and 150 most satisfied. An example question follows:

(Q27) The staff communicated well with me during labour.					
Είχα καλή επικοινωνία με το προσωπικό καθ' όλη τη διάρκεια του τοκετού.					
Strongly Agree	Agree	Neither Agree or Disagree	Disagree	Strongly Disagree	
Συμφωνώ Απόλυτα	Συμφωνώ	Ούτε συμφωνώ Ούτε διαφωνώ	Διαφωνώ	Διαφωνώ Απόλυτα	
Scores	5	4	3	2	1
Comments(Σχόλια): _____					

To view full content of the 30-item-BSS-LF see Appendix 1.

Data collection

Data was collected in 2012. Initially the *30-item-G-BSS-LF* was sent by email to interested participants, with (n=250) responding. Prior to completing the *30-item-G-BSS-LF*, an information sheet and consent form were sent by email, with opportunity provided to ask and have questions answered via email or telephone. The completed survey instrument (n=185; 75% return rate) was anonymously tagged and stored on a password protected computer.

Statistical analysis

The objectives of the study in terms of the evaluation of the psychometric properties of the *30-item-G-BSS-LF* required the use of Confirmatory Factor Analysis (CFA) (Kline, 1993, 2000); CFA being a special case of the Structural Equation Modelling (SEM) approach to data analysis (Byrne, 2010; Kline, 1998). The robustness of parametric tests (including CFA and SEM) against violations of the fundamental parametric assumptions (Martin & Thompson, 2000) have resulted in the contemporary use of ordinal or ordered categorical data, which represents the common reality of questionnaire data, with these statistical techniques (Friedrich et al., 2011; Kind & Barmby, 2011; Shulruf et al., 2004). However, data exhibiting significant deviation from the normal distribution assumption can lead to an erroneous outcome of a statistical analysis based on assumed parametric acceptable data distributional characteristic and consequently, an incorrect and potentially misleading interpretation of statistical findings (Flora & Curran, 2004; Lubke & Muthen, 2004; Martin & Thompson, 2000; Muthen & Kaplan, 1992). Skew and kurtosis characteristics of each item were examined and those exhibiting any significant deviation from normality were rejected from the *30-item-G-BSS-LF* item pool prior to further statistical analysis based on normality assumptions. The criteria

for item rejection was based on univariate skew and kurtosis characteristics and absolute skew values equal to, or greater than 3 and absolute kurtosis values of equal to, or greater than 10, based on the non-normality cut-off recommendations of Kline (2005). Statistical analysis for all quantitative studies in this thesis were conducted using PASW version 18 (SPSS, 2009a,b), Analysis of Moment Structures (AMOS) version 18 (Arbuckle, 1995-2009) and Mplus version 3 (Muthen & Muthen, 1998-2004).

Confirmatory Factor Analysis (CFA)

Four CFA models were evaluated, two multi-dimensional models and two uni-dimensional models. The multi-dimensional models represent three-factor models comprising three related thematically determined sub-scales of *Quality of Care (QC)*, *Women's Attributes (WA)*, and *Stress Experienced (SE)*, consistent with the themes identified in the original English language *30-item-BSS-LF* study by Hollins Martin and Fleming (2011). The two multi-dimensional models evaluated are the full 30-item version of the scale and the 10-item short-form version. The two uni-dimensional models evaluated represent single factor models of the full 30-item version of the BSS and the 10-item short-form version. Consistent with the assumption of multivariate normality, a maximum-likelihood (ML) approach to model estimation was adopted (Byrne, 2010; Kline, 1993; 2000). Multiple goodness of fit tests (Bentler & Bonett, 1980) were used to evaluate the models, these being the Comparative Fit Index (CFI) (Bentler, 1990), and the root mean squared error of approximation (RMSEA). A CFI greater than 0.90 indicates an acceptable fit to the data (Bentler & Bonett, 1980; Bollen, 1989; Hu & Bentler, 1995; Kline, 1998; Marsh et al, 1988), while a CFI equal to or greater than 0.95 indicates a good fit to the data (Hu & Bentler, 1999). A RMSEA with values of less than 0.08 indicate an acceptable fit to the data

(Browne & Cudeck, 1993), while values of less than 0.05 indicate a good fit to the data (Schumaker & Lomax, 2010). A statistically significant χ^2 indicates a significant proportion of variance within the data is unexplained by the model (Bentler & Bonett, 1980), however trivial and inconsequential variations in the data can promote a significant χ^2 statistic (Hu & Bentler, 1995), hence model evaluation is almost universally determined by model fits statistics such as CFI and RMSEA (Byrne, 2010; Hooper et al., 2008).

Divergent validity

Divergent validity was determined by correlating scale scores (30-item and 10-item versions) with the number of weeks pregnant when the baby was born. It was predicted that there would be no significant relationship between scores and the duration of pregnancy.

Known-groups discriminant validity

Known-groups discriminant validity was evaluated by testing for differences in scores in relation to birth type, comparing **unassisted vaginal delivery** to non-normal (forceps, ventouse, prearranged section, emergency section). It was predicted that total scores would be significantly higher for normal compared to non-normal delivery. In relation to theoretically circumscribed sub-scales embedded within the instrument, it is anticipated that while there would be no significant difference in *quality of care* provision sub-scale scores, and *women's personal attributes* sub-scales scores as a function of delivery type. However, it is predicted that scores on the *stress experienced during labour* sub-scale would be significantly higher, since higher scores relate to greater satisfaction within this domain, in the **unassisted vaginal delivery** type group.

Internal consistency

An internal consistency analysis was conducted to ensure that the measures satisfied the criteria for clinical and research purposes using the Cronbach coefficient alpha statistical procedure. A Cronbach's alpha reliability statistic of 0.70 is considered as the minimum acceptable criterion of instrument internal reliability (Kline, 1993, 2000).

Results

Descriptive results

A total of 162 women completed the *30-item-G-BSS-LF*, with 116 (72%) of these participants primigravidas. The average duration of pregnancy was 39 (SD 1.31) weeks, and the average duration of labour was 6.81 (SD 5.43) hours. The total *30-item-G-BSS-LF* mean score was 112.57 (SD 16.62, Range 66-146), and the thematically determined sub-scale mean scores of the *quality of care provision* (8-items), *women's personal attributes* (8-items), and *stress experienced during labour* (14-items) sub-scales were 30.67 (SD 5.25, Range 10-40), 29.60 (SD 5.15, Range 12-39) and 52.31 (SD 8.52, Range 29-69) respectively.

Multivariate normality

The distribution of the *30-item-G-BSS-LF* items revealed no significant evidence of skew or kurtosis with the exception of item 6. '*I gave birth to a normal healthy baby*' (skew = 3.54; kurtosis = 12.72).

Measurement evaluation of predicted models

The structure of the thematically derived 30-item three-factor model was found to be poor, $\chi^2_{(df = 403)} = 1188.51$, $p < 0.001$, CFI = 0.55 and RMSEA = 0.11. The alternative 30-item single-factor model was also revealed to have a poor fit to the data, $\chi^2_{(df = 406)} = 1209.93$, $p < 0.001$, CFI = 0.54 and RMSEA = 0.11. The 10-item single-factor model based on the *10-item-BSS-R* revealed a poor fit to the data, $\chi^2_{(df = 35)} = 178.08$, $p < 0.001$, CFI = 0.66 and RMSEA = 0.16. The structure of the 10-item three-factor model based on Hollins Martin and Martin (2014) and the *10-item-G-BSS-R* was also found to be improved, with model fit approaching acceptability, $\chi^2_{(df = 32)} = 76.20$, $p < 0.001$, CFI = 0.90 and RMSEA = 0.09. Application of the Satorra-Bentler scaled χ^2 , which controls for data distributional non-normality², improved fit of this model to acceptability $\chi^2_{(df = 32)} = 64.77$, $p < 0.001$, CFI = 0.91 and RMSEA = 0.08. To view the 10-BSS items comprising three-factor best-fit measurement model and associated domains reported by Hollins Martin and Martin (2014) (see [Appendix 1](#)).

Rescaling of the 10-item short-form version (10-item-G-BSS-R)

Using the three-subscale model of the *10-item-BSS-R* and the suggested rescaling to a zero point as suggested by Hollins Martin and Martin (2014), the 10-item revised version of the *10-item-G-BSS-R* was rescaled to produce a zero score point across the total scale and the three subscales, thus the revised 10-item instrument would be scored along a 0-4 Likert scale instead of a 1-5 Likert scale. Using this approach the *10-item-G-BSS-R* total score was 27.45 (SD 6.22, range 6-40), and the thematically

² It should be noted that the 10-item G-BSS-R demonstrates no evidence of non-normality based on the criteria highlighted earlier in the paper. However, as this version of the scale is embedded in the long-form which does demonstrate non-normality and that responses to the short-form may be influenced by responses to the longer version, a consistent approach to data analysis between both versions of the measure was chosen.

determined sub-scale mean scores of the *quality of care provision* (BSS-QC), *women's personal attributes* (BSS-WA), and *stress experienced during labour* (BSS-SL) sub-scales were 11.75 (SD 3.10, range 2-16), 5.21 (SD 1.94, range 0-8) and 10.49 (SD 2.99, range 1-16) respectively. All three sub-scales were observed to be moderately to highly correlated and indeed, correlated with the full *30-item-G-BSS-LF* version and the associated thematically-derived sub-scales. The relationship between short-form and full-length total and sub-scales versions are shown in Table 1.

TABLE 1. ABOUT HERE

Divergent validity

No significant correlation was observed between the *10-item-G-BSS-R* total score and the number of weeks of the pregnancy, $r = 0.02$, $p=0.84$. Neither was there evidence of any significant relationships with the *10-item-G-BSS-R* short form sub-scale scores; between G-BSS-SL-SF scores, G-BSS-WA-SF scores, G-BSS-QC-SF and pregnancy duration, $r = 0.02$, $p = 0.82$; $r = 0.08$, $p = 0.31$; $r = 0.01$, $p = 0.99$; respectively. Similarly, no significant correlation was observed between the *30-item G-BSS-LF* total score and the number of weeks of the pregnancy, $r = 0.03$, $p=0.73$. Neither was there evidence of any significant relationships with the *30-item-G-BSS-LF* sub-scale scores: between G-BSS-SL scores, G-BSS-WA scores, G-BSS-QC and pregnancy duration, $r = 0.01$, $p = 0.96$; $r = 0.11$, $p = 0.17$; $r = 0.03$, $p = 0.73$; respectively.

Known-groups discriminant validity

The mean *10-item-G-BSS-R* total score and *G-BSS-SL-R*, *G-BSS-WA-R* and *G-BSS-QC-R* sub-scale scores as a function of delivery type are shown in Table 2.

TABLE 2. ABOUT HERE

A significant difference between groups differentiated by delivery type was observed on the *10-item-G-BSS-R* total score, $t_{(160)} = 3.96$, $p < 0.001$, and *G-BSS-SE-R* sub-scale score, $t_{(160)} = 2.23$, $p = 0.03$ in the direction predicted. Against prediction, statistically significant differences were observed in *BSS-WA-R* sub-scale scores, $t_{(160)} = 1.98$, $p = 0.05$, and *BSS-QC-R* sub-scale scores, $t_{(160)} = 4.51$, $p < 0.001$, as a function of delivery type. Similarly, examination of the impact of delivery type on the *30-item-G-BSS-LF* derived scores revealed significant differences between groups on total score, $t_{(160)} = 5.32$, $p < 0.001$, and sub-scale score, $t_{(160)} = 4.27$, $p < 0.001$ in the direction predicted. Consistent with short-form sub-scales scores and against prediction, statistically significant differences were observed in *G-BSS-WA-R* sub-scale scores, $t_{(160)} = 4.91$, $p < 0.001$, and *G-BSS-QC-R* sub-scale scores, $t_{(160)} = 4.78$, $p < 0.001$, as a function of delivery type. The mean scores as a function of delivery type are shown in Table 2.

Internal consistency

Calculated Cronbach's alpha's of the *10-item-G-BSS-R* total scale (0.78) and subscales *G-BSS-SE-R* (0.76), *G-BSS-WA-R* (0.51), and *G-BSS-QC-R* (0.56).

Cronbach's alpha of full scale derived scores for the *30-item-G-BSS-LF* total scale (0.75) and subscales *G-BSS-SE-LF* (0.89), *G-BSS-WA-LF* (0.80), and *G-BSS-QC-LF* (0.67). Thus, the *10-item G-BSS-WA-R* and *G-BSS-QC-R* subscales and the *G-BSS-QC-LF* subscale did not reach the conventional threshold of internal consistency reliability.

Discussion

The findings from this study are generally consistent with that of the Hollins Martin and Martin (2014) paper, which demonstrates support for the robustness of the *10-item-BSS-R* in terms of factor structure in the English version of the instrument. With similarity, our 30-item Greek version performed badly in terms of model fit, again, a finding consistent with Hollins Martin and Martin (2014). Our findings were also consistent with Hollins Martin and Martin (2014) in relation to known-groups discriminant validity findings, where those differentiated by having a normal birth scored significantly higher on all items of the *30-item-G-BSS-LF* and *10-item-G-BSS-R* sub-scales. It was originally predicted that there would be no difference in scores as a function of birth type on *Quality of Care (QC)* and *Women's Attributes (WA)* sub-scales, however statistically significant higher scores were found in the normal birth group.

It may be that our original predictions were actually in error, since these sub-scales represent domains of birth satisfaction and the self-perception and self-evaluation of these attributes may well be influenced by a non-normal birth experience. Indeed, it is clear from the Hollins Martin and Martin (2014) paper that the profile of scores on the *10-item-BSS-R* sub-scales and total score was identical in interpretation to what was found in the current study. Consequently, it is plausible that the *G-BSS-R-QC* and *G-BSS-R-WA* sub-scales are sensitive to germane contextual aspects of the birth experience, **for example, the reassurance that may be perceived from experiencing a normal vaginal delivery**, rather than being representative of absolute quality of care or trait-specific aspects of women's individual attributes.

The lack of support for a uni-dimensional model of either the *30-item-G-BSS-LF* or the *10-item-G-BSS-R* is again consistent with Hollins Martin and Martin's (2014) critique of earlier work on the original development of the English (UK) version of the scale, which suggested scoring the tool as a uni-dimensional measure with a single sum score (Hollins Martin & Fleming, 2011). However, it should be acknowledged that within the early development of the scale, Hollins Martin and Fleming (2011) had identified a plausible thematic structure that they recommended for further evaluation in terms of possible sub-scale differentiation. Our findings regarding poor model fit of the *30-item-G-BSS-LF* do however contradict the assertion of Hollins Martin et al. (2012) that the *30-item-BSS-LF* could be promoted as a fundamentally robust measure of birth satisfaction under the rubric that the instrument accounted for all the analysed data. This contradiction between findings of the *Concurrent Analysis* paper (Hollins Martin et al., 2012) and the psychometric paper (Hollins Martin & Martin, 2014) was also highlighted by Hollins Martin and Martin (2014). Consequently and consistent with Hollins Martin and Martin (2014), our findings would suggest that there may be tautological issues inherent within the *Concurrent Analysis* methodology used by Hollins Martin et al. (2012).

The short-form *10-item-G-BSS-R* (Greek version) performed well in terms of measurement replication of the *10-item-BSS-R* (English version) in our Greek sample. The one concern raised with respect to this short version of the tool, is the internal consistency estimations. The total score and *Quality of Care* (QC) subscale scores reached satisfactory consistency. However, the *Stress Experienced* (SE) and *Women's Attributes* (WA) subscales did not reach threshold levels of acceptable consistency. It is established that alpha levels are influenced by the number of items

in the scale, therefore, with just two items it is unsurprising that the *10-item-G-BSS-R-WA* sub-scale performed sub-optimally. Hence, it is surprising that the *4-item-G-BSS-R-SE* did not perform well, particularly in comparison with the findings of Hollins Martin and Martin (2014), where this sub-scale revealed satisfactory internal consistency properties. It may be that less than exemplary internal consistency characteristics of this particular sub-scale may have been influenced by the translation process and deleteriously impacted on this measurement index. It is suggested that further work be conducted to look at any impact of the translation process on this sub-scale in future use of the *10-item-G-BSS-R*, and also to consider whether additional items may be required for the *10-item-G-BSS-R-WA* sub-scale to improve its internal consistency. A Cronbach's alpha in the range of 0.5 to 0.6 has been described as poor but not unacceptable (George and Mallery, 2003), thus consideration of the implications of alpha levels at this level for the sub-scales described is important to address and if possible remedy with both a parsimonious and theoretically as well as statically valid approach. A further limitation is the use of the *10-item-G-BSS-R* embedded within the full 30-item instrument. It is possible that responses on the 10-item version of the scale could be influenced by the content of the twenty additional items of the full scale. In this regard, further evaluation of the measurement characteristics of the *10-item-G-BSS-R* used in isolation from the full 30-item-G-BSS is suggested in order to confirm the findings of the current investigation. Finally, the convenience nature of the participants who participated in the study meant that there were a very small number (N=4) who had a premature (<37) weeks delivery. It is plausible that for these individuals may have been exposed to additional stress as a consequence of early delivery and if so, this may also have a potential influence on both G-BSS scale scores and associated sub-

scales. It is desirable in future evaluative work with the G-BSS measures to determine the influence of premature birth on birth satisfaction, accepting that comparisons will require larger numbers for statistical comparisons to be meaningful and the utilisation of additional measures of stress to determine the relationship with birth satisfaction within the context of premature birth.

The availability of these short-form and long-form versions of the Greek-language versions of the original BSS, notwithstanding the limitations highlighted above, does facilitate the opportunity to explore the relationship of birth satisfaction in relation to salient psychological domains of relevance to childbirth such as postnatal depression (Wylie et al., 2011), anxiety (Jomeen and Martin, 2005), perceptions of care (Redshaw and Martin, 2009), worry (Redshaw et al., 2009), prenatal distress (Alderdice et al., 2013) within the Greek context. Future studies with the G-BSS might also usefully examine alternative measures to assess divergent validity, beyond gestational age which was used in the current study. Not only would this furnish additional confirmatory data regarding divergent validity, it would also allow the opportunity to determine if any of these selected variables did indeed have a significant relationship with G-BSS scores, against hypothesis, and thus present evidence for variables which may need to be controlled in statistical analysis (e.g. entering the variable as a covariate) within a clinical context. A final, pleasing observation from the current study was the high return rate. The high return rate was considered to have been achieved through the Foot-In-The-Door (FITD) technique (Burger, 1999), which involves conscripting people through an initial first request followed by a second request to complete the survey.

Conclusion

The *30-item-G-BSS-LF* and *10-item-G-BSS-R* and all their associated sub-scales were shown to have excellent divergent validity characteristics, suggesting that in the domain evaluated, that the instrument can be used reliably and is not influenced by key perinatal characteristics, such as length of pregnancy. Further evaluation of the divergent validity characteristics in other aspects of postnatal application of the tool would appear worthy of evaluation to define further the parameters of instrument use.

It must be emphasised however in view of the limitations of this study highlighted previously, that additional research endeavour is highly desirable to corroborate our initial findings and provide additional validation evidence for this Greek-language version of the BSS. The data from the current study has shown considerable promise in the development and initial validation of the *10-item-G-BSS-R* as a robust instrument for health researchers and health professionals, including midwives, obstetricians, maternity care managers to use to measure Greek speaking women's levels of birth satisfaction (see Appendix 2). It is anticipated that the future development and use of the *G-BSS* will help contribute to the evidence-base in relation to birth satisfaction and associated psychological parameters within Greek populations of women.

References

Alderdice, F., Savage-McGlynn, E., Martin, C.R., McAuliffe, F.M., Hunter, A., Unterscheider, J., Daly, S., Geary, M.P., Kennelly, M.M., O'Donoghue, K., Morrison, J.J., Burke, G., Dicker, P., Tully, E.C. & Malone, F.D. (2013). The Prenatal Distress Questionnaire: An investigation of factor structure in a high risk population. *Journal of Reproductive and Infant Psychology*, 31(5), 456-464.

Arbuckle, J.L. (1995-2009). *AMOS 18 User's Guide*. Chicago, IL: AMOS Development Corporation.

Bentler, P.M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin*, 107, 238-246.

Bentler, P.M., & Bonett, D.G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin*, 88, 588-606.

Bollen, K.A. (1989). A new incremental fit index for general structural equation models. *Sociological Methods and Research*, 17, 303-316.

Browne, M.W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K.A. Bollen & J.S. Long (Eds.), *Testing Structural Equation Models* (pp. 136-162). Newbury Park, CA: Sage.

Burger, J. M. (1999). The foot-in-the-door compliance procedure: a multiple process and analysis review. *Personality and Social Psychology Review*. 3: 303-325.

Byrne, B.M. (2010). *Structural equation modelling with AMOS: Basic concepts, applications, and programming (2nd ed.)*. New York, NY US: Routledge/Taylor & Francis Group.

Department of Health: Great Britain, (1993). *Changing Childbirth: Report of the Expert Maternity Group (Cumberlege Report)*. HMSO, London.

Fleming S., & Vandermause R. (2011). Grand Multiapare's evolving experiences of birthing and technology in U.S. hospitals. *JOGNN: Journal of Obstetric, Gynecologic, and Neonatal Nursing*. 40(6), 742-752.

Flora, D.B., & Curran, P.J. (2004). An empirical evaluation of alternative methods of estimation for confirmatory factor analysis with ordinal data. *Psychological Methods*, 9(4), 466-491.

Fones, C. (1996). Posttraumatic stress disorder occurring after painful child birth. *Journal of Nervous and Mental Disease*, 18, 195-196.

Friedrich, F., Alexandrowicz, R., Benda, N., Cerny, G., & Wancata, J. (2011). The criterion validity of different versions of the General Health Questionnaire among non-psychiatric inpatients. *Social Psychiatry & Psychiatric Epidemiology*, 46(7), 635-641.

George, D., & Mallery, P. (2003). *SPSS for Windows Step by Step: A Simple Guide and Reference*. 11.0 update (4th ed.). Boston: Allyn and Bacon.

Hollins Martin, C.J., & Fleming, V. (2011). The Birth Satisfaction Scale (BSS). *International Journal of Health Care Quality Assurance*. 24(2), 124-135.

Hollins Martin, C.J., Forrest, E. (2013). *Bereavement Care for Childbearing Women and Their Families: An Interactive Workbook*. Routledge, Abingdon, Oxon (UK). ISBN 978-0-415-82724-9

Hollins-Martin, C.J., Martin, C. (2014). Development and psychometric properties of the Birth Satisfaction Scale-Revised (BSS-R). *Midwifery*.
<http://dx.doi.org/10.1016/j.midw.2013.10.006>

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(3), 247-260.

Hooper, D., Coughlan, J., & Mullen, M.R. (2008). Structural equation modelling: Guidelines for determining model fit. *Electronic Journal of Business Research Methods*, 6(1), 53-60.

Hu, L.T., & Bentler, P.M. (1995). Evaluating model fit. In R.H. Hoyle (Ed.), *Structural equation modelling: concepts, issues and applications*. Thousand Oaks, CA: Sage.

Iles, J., Slade, P. and Spiby, H. (2011). Posttraumatic stress symptoms and postpartum depression in couples after childbirth: the role of partner support and attachment. *Journal of Anxiety Disorders*, 25(4):520-30.

Jomeen, J. & Martin, C.R. (2005). Confirmation of an occluded anxiety component within the Edinburgh Postnatal Depression Scale (EPDS) during early pregnancy. *Journal of Reproductive and Infant Psychology*, 23, 143-154.

Kind, P., & Barmby, P. (2011). Defending attitude scales. In M. Russell & M. Kavanaugh (Eds.), *Assessing students in the margins: Challenges, strategies, and techniques*. (pp. 117-135). Charlotte, NC US: IAP Information Age Publishing.

Kline, P. (1993). *The handbook of psychological testing*. London: Routledge.

Kline, R.B. (1998). *Principles and practice of structural equation modeling*. New York: Guilford.

Kline, P. (2000). *A psychometrics primer*. London: Free Association Books.

Kline, R.B. (2005). *Principles and Practice of Structural Equation Modeling* (2nd ed.). New York: Guilford Press.

Lubke, G.H., & Muthén, B.O. (2004). Applying Multigroup Confirmatory Factor Models for Continuous Outcomes to Likert Scale Data Complicates Meaningful Group Comparisons. *Structural Equation Modeling*, 11(4), 514-534.

Marsh, H.W., Balla, J.R., & McDonald, R.P. (1988). Goodness-of-fit indices in confirmatory factor analysis: the effect of sample size. *Psychological Bulletin*, 103, 391-410.

Martin, C.R., & Thompson, D.R. (2000). *Design and Analysis of Clinical Nursing Research Studies*. London: Routledge.

McDonald, S., Slade, P., Spiby, H. and Iles, J. (2011). Post-traumatic stress symptoms, parenting stress and mother-child relationships following childbirth and at 2 years postpartum. *Journal of Psychosomatic Obstetrics and Gynaecology*, 32(3), 141-146.

Muthen, B., & Kaplan, D. (1992). A comparison of some methodologies for the factor analysis of non-normal Likert variables: A note on the size of the model. *British Journal of Mathematical and Statistical Psychology*, 45(1), 19-30.

Muthen, L.K., & Muthen, B.O. (1998-2004). *Mplus User's Guide* (3rd ed.). Los Angeles, CA: Muthen and Muthen.

Redshaw, M., Martin, C.R., Rowe, R. & Hockley, C. (2009). The Oxford Worries about Labour Scale: Women's experience and measurement characteristics of a measure of maternal concern about labour and birth. *Psychology, Health and Medicine*, 14, 354-366.

Redshaw, M. & Martin, C.R. (2009). Validation of a perceptions of care adjective check list. *Journal of Evaluation in Clinical Practice*, 15, 281-288.

Schumacker, R.E., & Lomax, R.G. (2010). *A beginner's guide to structural equation modelling* (3rd ed.). New York, NY US: Routledge/Taylor & Francis Group.

Shulruf, B., Hattie, J., & Dixon, R. (2008). Factors affecting responses to Likert type questionnaires: introduction of the ImpExp, a new comprehensive model. . *Social Psychology of Education*, 11(1), 59-78.

SPSS (2009a). *PASW Statistics 18 Core System User's Guide*. Chicago, IL: SPSS, Inc.

SPSS (2009b). *PASW Advanced Statistics 18*. Chicago, IL: SPSS, Inc.

Wylie, L., Hollins Martin, C.J., Marland, G., Martin, C.R. & Rankin, J. (2011). The enigma of postnatal depression: an update. *Journal of Psychiatric and Mental Health Nursing*, 8, 48-58.

Table 1. Correlations between the *10-item-G-BSS-R* total and sub-scale scores and the full *30-item-G-BSS-LF* total* and sub-scale* scores (all correlations statistically significant at $p < 0.01$)

Scale	G-BBC-R(total)	G-BSS-SE	G-BBS-WA	G-BSS-QC	G-BSS*	G-BSS-SE*	G-BBS-WA*	G-BSS-QC*
G-BSS-R(total)		0.85	0.69	0.75	0.92	0.85	0.78	0.76
G-BSS-SE			0.57	0.39	0.79	0.86	0.57	0.56
G-BSS-WA				0.21	0.52	0.48	0.61	0.28
G-BSS-QC					0.75	0.58	0.63	0.80
G-BSS*						0.93	0.82	0.86
G-BSS-SE*							0.63	0.70
G-BSS-WA*								0.59
G-BSS-QC*								

G-BSS-R=10-item-Greek-Birth-Satisfaction-Scale-Revised

G-BSS-SE=10-item-Greek-Birth-Satisfaction-Scale-Revised (SE=Stress Experienced)

G-BSS-WA=10-item-Greek-Birth-Satisfaction-Scale-Revised (WA=Women's Attributes)

G-BSS-QC=10-item-Greek-Birth-Satisfaction-Scale-Revised (QC=Quality of Care)

G-BSS*=30-item-Greek-Birth-Satisfaction-Long-Form

G-BSS-SE*=30-item-Greek-Birth-Satisfaction-Long-Form (SE=Stress Experienced)

G-BSS-WA*=30-item-Greek-Birth-Satisfaction-Long-Form (WA=Women's Attributes)

G-BSS-QC*=30-item-Greek-Birth-Satisfaction-Long-Form (QC=Quality of Care)

Table 2. Mean *10-item-G-BSS-R* and *30-item-G-BSS-LF* sub-scale scores as a function of delivery type (standard deviations in parentheses).

Variable	Unassisted vaginal delivery (n=140)	Non-normal delivery (n=22)
10-item-G-BSS-R total	28.19 (5.90)	22.77 (6.34)
10-item-G-BSS-R-Stress Experienced	10.69 (2.94)	9.18 (3.05)
10-item- G-BSS-R-Woman's Attributes	5.33 (1.97)	4.45 (1.56)
10-item-G-BSS-R-Quality of Care	12.16 (2.84)	9.14 (3.47)
30-item-G-BSS total	115.12 (14.71)	96.36 (19.16)
30-item-G-BSS-Stress Experience	53.39 (7.79)	45.45 (9.91)
30-item-G-BSS-Women's Attributes	30.34 (4.88)	24.91 (4.41)
30-item-G-BSS-Quality of Care	31.40 (4.58)	26.00 (6.82)

Subscale (1) Quality of Care = home assessment, birth environment, sufficient support, relationships with health professionals;
 Subscale (2) Women's Attributes = ability to cope in labour, feeling in control, preparation for childbirth, relationship with baby;
 Subscale (3) Stress Experienced = distress during labour, obstetric injuries, perception of sufficient medical care, obstetric intervention, pain experienced, long labour, health of baby

Appendix 1: The statements that comprise the Birth Satisfaction Scale	
BSS item	Domain
(Q1) I coped well during my birth.	WA
<i>*(Q2) The delivery room staff encouraged me to make decisions about how I wanted my birth to progress.</i>	QC
(Q3) I was well prepared for my labour, i.e., read a lot of literature and/or attended parenthood education classes.	WA
(Q4) <i>I found giving birth a distressing experience.</i>	SE
(Q5) <i>I came through childbirth virtually unscathed.</i>	SE
(Q6) I gave birth to a healthy normal baby.	SE
(Q7) During labour I received outstanding medical care.	SE
(Q8) I received a lot of medical intervention, i.e., induction, forceps, section etc.	SE
(Q9) I had a swift and speedy labour.	SE
(Q10) I felt well supported by my partner during labour and birth.	QC
(Q11) I was encouraged to hold my baby for a substantial amount of time after birth.	WA
(Q12) My birth experience was considerably different to what I intended.	QC
(Q13) I had the same midwife throughout the entire process of labour and delivery.	QC
(Q14) I felt that the delivery room was unthreatening and comfortable.	QC
(Q15) <i>I felt very anxious during my labour and birth.</i>	WA
(Q16) <i>I felt out of control during my birth experience.</i>	WA
(Q17) I felt it was better not to know in advance about the processes of giving birth.	WA
(Q18) <i>I was not distressed at all during labour.</i>	SE
(Q19) I felt mutilated by my birth experience.	SE
(Q20) My baby was avoidably hurt during birth.	SE
(Q21) The staff provided me with insufficient medical care during my birth.	SE
(Q22) I had a natural labour, i.e., minimal medical intervention.	SE
(Q23) <i>I thought my labour was excessively long.</i>	SE
(Q24) <i>I felt well supported by staff during my labour and birth.</i>	QC
(Q25) I was separated from my baby for a considerable period of time after my birth.	WA
(Q26) My birth proceeded as I planned it.	QC
(Q27) <i>The staff communicated well with me during labour.</i>	QC
(Q28) <i>The delivery room was clean and hygienic.</i>	QC
(Q29) Giving birth was incredibly painful.	SE
(Q30) Labour was not as painful as I imagined.	SE

Note: Stress experienced = SE; Quality of Care = QC; Women's Attributes = WA. Short-form items of the BSS-R are in italics.

*Question 2 was originally circumscribed by Fleming and Martin (2011) as a 'Women's Attributes' domain item, but was content reviewed and statistical evaluated by Hollins Martin and Martin (2014) and determined to be a 'Quality of Care' domain item and thus realigned to that domain in their revision of the scale and development of the short-form version.

Appendix 2: Valid and reliable 10-item-Greek Birth Satisfaction-Scale-Revised (10-item-G-BSS-R)

Strongly Agree (Συμφωνώ Απόλυτα) = (SA); Agree (Συμφωνώ) = (A); Neither Agree or Disagree (Ούτε συμφωνώ Ούτε διαφωνώ) = (NA/D), Disagree (Διαφωνώ) = (D), Strongly Disagree (Διαφωνώ Απόλυτα) = (SD)

(1) I came through childbirth virtually unscathed (Μέσα από τον τοκετό βγήκα σχεδόν αλώβητη).

	SA	A	NA/D	D	SD
Scores	4	3	2	1	0

Comments (Σχόλια):

(2) I thought my labour was excessively long (Νομίζω ότι ο τοκετός μου κράτησε υπερβολικά πολύ ώρα).

	SA	A	NA/D	D	SD
Scores	0	1	2	3	4

Comments (Σχόλια):

(3) The delivery room staff encouraged me to make decisions about how I wanted my birth to progress (Το προσωπικό στην αίθουσα τοκετού με ενθάρρυνε να παίρνω μόνη μου τις αποφάσεις για το πώς επιθυμώ να εξελιχθεί το τοκετός).

	SA	A	NA/D	D	SD
Scores	4	3	2	1	0

Comments (Σχόλια):

(4) I felt very anxious during my labour and birth (Ένιωθα πολύ αγχωμένη κατά την διάρκεια του τοκετού).

	SA	A	NA/D	D	SD
Scores	0	1	2	3	4

Comments (Σχόλια):

(5) I felt well supported by staff during my labour and birth (Αισθανόμουν ότι το προσωπικό με στήριζε κατά τη διάρκεια του τοκετού και της γέννας).

	SA	A	NA/D	D	SD
Scores	4	3	2	1	0

Comments (Σχόλια):

(6) The staff communicated well with me during labour (Είχα καλή επικοινωνία με το προσωπικό καθ' όλη τη διάρκεια του τοκετού).

	SA	A	NA/D	D	SD
Scores	4	3	2	1	0

Comments (Σχόλια):

(7) I found giving birth a distressing experience (Βρήκα την διαδικασία του τοκετού ως μια οδυνηρή εμπειρία).

	SA	A	NA/D	D	SD
Scores	0	1	2	3	4

Comments (Σχόλια):

(8) I felt out of control during my birth experience (Αισθανόμουν ότι ήμουν εκτός ελέγχου κατά τον τοκετό).

	SA	A	NA/D	D	SD
Scores	0	1	2	3	4

Comments (Σχόλια):

(9) I was not distressed at all during labour (Δεν ήμουν καθόλου αγχωμένη κατά την διάρκεια του τοκετού).

	SA	A	NA/D	D	SD
Scores	4	3	2	1	0

Comments (Σχόλια):

(10) The delivery room was clean and hygienic (Το δωμάτιο όπου έγινε ο τοκετός ήταν καθαρό και υγιεινό).

	SA	A	NA/D	D	SD
Scores	4	3	2	1	0

Comments (Σχόλια):

