Tezuka, A., Hiroyama, N, Suzuki M, Matsuoka M, Martin CJH, Martin CR. (2023). Translation and validation of the Japanese version of the Birth Satisfaction Scale-Revised. *Japanese Journal of Nursing Science*. e12569.
doi: 10.1111/jjns.12569.

Translation and Validation of the Japanese Version of the Birth Satisfaction Scale-Revised

Aya Tezuka[a]*, Natsuko Hiroyama[b], Miwa Suzuki[a], Megumi Matsuoka[a], Caroline J. Hollins Martin[c] and Colin R. Martin[d]

[a] Health of Science, Kyorin University,

6-20-2, Shinkawa, Mitaka-shi, Tokyo, 181-8611, Japan

[b] Graduate School of Health Care Sciences, Tokyo Medical and Dental University (TMDU),

1-5-45, Yushima, Bunkyo-ku, Tokyo, 113-8519, Japan

[c] School of Health and Social Care, Edinburgh Napier University (ENU),

Sighthill Campus, Midlothian, Scotland, EH11 4BN, UK

[d] Institute for Health and Wellbeing, University of Suffolk,

Neptune Quay, IPSWICH, Suffolk, England, IP4 1QJ, UK

*** Corresponding author**: Aya Tezuka

Health of Science, Kyorin University

6-20-2, Shinkawa, Mitaka-shi, Tokyo, 181-8611, Japan.

e-mail address: a-taniguchi@ks.kyorin-u.ac.jp

Telephone number: +81-422-47-5514 (Ex 6789)

Abstract

Aim: This study aimed to develop a Japanese version of the Birth Satisfaction Scale-Revised and evaluate its reliability and validity.

Methods: After translating the Birth Satisfaction Scale-Revised into Japanese, we conducted an Internet-based cross-sectional study with 445 Japanese-speaking women within two months of childbirth. Of these, 98 participated in the retest one month later. Data were analyzed using the COSMIN study design checklist for patient-reported outcome measurement instruments. Content validity was evaluated through cognitive debriefing during the translation process into Japanese. Confirmatory factor analysis was conducted to verify structural and cross-cultural validities. For hypothesis testing, we tested correlations with existing measures for convergent and divergent validities, and for known-group discriminant validity, we made comparisons between types of childbirth. Internal consistency was calculated using Cronbach's α, and test-retest reliability was evaluated using the intraclass correlation coefficient.

Results: For the Japanese-Birth Satisfaction Scale-Revised, the established three-factor model fit poorly, whereas the four-factor model fit better. Full metric invariance was observed in both the nulliparous and multiparous groups. Good convergent, divergent, and known-group discriminant validities and test-retest reliability were established. Internal consistency observations were suboptimal; however for vaginal childbirth, the Cronbach's α of the total score was 0.71.

Conclusions: The Japanese-Birth Satisfaction Scale-Revised was a valid and reliable scale, with the exception of internal consistency that requires further investigation. If limited to vaginal childbirth, research, clinical applications, and international comparisons can be drawn.

Keywords

Birth, Factor Analysis, Patient Satisfaction, Quality of Care, Validity and Reliability

Introduction

Pregnancy and childbirth are transformative events in a woman's life, both physically and mentally. Positive birth experiences help women feel their own power and build trusting and supportive relationships with their partners and healthcare providers (Karlström et al., 2015). By contrast, negative birth experiences create feelings of grief, anger, and loss. Furthermore, they can lead to long-term problems, such as postpartum depression, post-traumatic stress disorder, difficulty establishing mother–child relationships, and fear of childbirth (Bell & Andersson, 2016; Simpson & Catling, 2016).

Although birth experience is a subjective event, it can be objectively assessed using tools. They can be used as indicators to improve the quality of care (Sawyer et al., 2013) and examine associations with various factors, such as self-efficacy and postpartum depression (Bell & Andersson, 2016). Intrapartum care has a significant impact on birth satisfaction (World Health Organization, 2018). Thus, many tools have been developed to measure care satisfaction (Britton, 2012; Sawyer et al., 2013), and these tools have been used to improve care. However, various factors influence the experience of childbirth, including the relationship with and support from care providers, pain acceptance, emotions, and sense of control during labor, participation in decision making, obstetric abnormalities, and fulfillment of expectations (Britton, 2012; McKelvin et al., 2021). Therefore, an assessment tool that encompasses these factors would be ideal (Britton, 2012; Rudman et al., 2007; Waldenström & Rudman, 2008).

The Birth Satisfaction Scale-Revised (BSS-R) (Hollins Martin & Martin, 2014) is a multidimensional, psychometrically robust tool developed in the UK to measure women's experience of childbirth. The original BSS was developed by collecting "expressions indicating satisfaction or dissatisfaction with childbirth" through a literature review (Martin & Fleming, 2011). After a concurrent analysis and model re-specification, it was compiled into the 10-item BSS-R (Hollins Martin et al., 2012; Hollins Martin & Martin, 2014), consisting of three subscales: quality of care provision, women's personal attributes, and stress experienced during labor. It was confirmed to have a good fit to the three-factor model by confirmatory factor analysis as well as criterion-

related validity and internal consistency (Hollins Martin & Martin, 2014). It was adopted by the International Consortium for Health Outcome Measurement (ICHOM) as the standard set for pregnancy and childbirth (Nijagal et al., 2018). The BSS-R has been translated into several languages, including Italian, Dutch, and Czech. Furthermore, its validity and reliability have been confirmed (Emmens et al., 2021; Nespoli et al., 2018, 2021; Ratislavová et al., 2022). However, a Japanese version does not yet exist.

In Japan, among several other scales, the Childbirth Experience Self-Rating Scale is widely used for evaluating childbirth experiences (Tokiwa & Imazeki, 2000). This scale encompasses care and other factors, and its reliability and validity have been verified. However, 35 items are burdensome for clinical use, and international comparisons are difficult. Hence, the development of a Japanese version of the BSS-R (Japanese-BSS-R) would make it possible to measure satisfaction with childbirth in Japan, while allowing international comparisons.

This study aimed to develop and validate the Japanese-BSS-R and examine its key psychometric parameters of reliability and validity.

Methods

Design

This cross-sectional study used an Internet-based questionnaire survey. The survey was conducted between January and March 2022. The inclusion criteria were native Japanese speakers, aged between 20 and 49 years, within two months postpartum, and those who had given birth at 37–41 weeks gestation. Since it was recommended that the BSS-R be measured approximately six to eight weeks postpartum, women within two months of childbirth were included. The survey used convenience sampling. Among the registered panels of Internet research firms (Cross Marketing Inc.) in Japan, women who were scheduled to give birth two months prior to the survey were presented with the study and asked to participate. Then, a questionnaire survey was administered to women who met the inclusion criteria in the screening survey and gave their consent to participate. The women who participated in the study received monetary incentives. To validate the test-retest

reliability, a secondary survey was conducted one month later, that is, within three months postpartum.

With respect to sample size, at least 150 or five times the number of questions (29 x 5 = 145) is required to examine cross-cultural validity (Mokkink et al, 2019). In this study, 150 participants were assigned to caesarean section (CS) to examine measurement invariance across parities and types of childbirth. The overall sample size was set at 450 participants because the number of participants in a prior study that examined the reliability and validity of the BSS-R in an Internet survey was 459 (Martin et al., 2020). The sample size for the retest was set at 100 participants because more than 50 participants are needed to examine test-retest reliability (Mokkink et al, 2019). Both the main and secondary surveys were closed when the set number of participants was reached. There were no missing data, as any omitted responses would not allow the women to proceed to the next question on the Internet-based survey. However, five multivariate outliers were identified by Mahalanobis distances, which exceeded the cut-off value of $\chi^2 > 29.59$ for a 10-item measure (Mahalanobis, 1936).

Ethical Approval

Ethical approval was obtained from the Institutional Review Board of Kyorin University (Approval number: 2021-64 for cognitive debriefing, 2021-84 for validation survey). All participants were provided with a written explanation of the study purpose and background, methods, their voluntary participation, that they may discontinue at any time, personal information protection, and the publication of the study results. If they consented, they were asked to check a box to confirm their consent before answering.

Instruments or Measures

The Birth Satisfaction Scale-Revised

The BSS-R is a 10-item self-report measure of birth experience that comprises three subscales of stress experienced during labor (SE sub-scale, four items), women's personal attributes

(WA sub-scales, two items), and the quality of care (QC sub-scales, four items). The BSS-R consists of a 5-point Likert scale that ranges from 0 (strongly disagree) to 4 (strongly agree), and the scores range from 0 to 40. Higher scores indicate greater satisfaction with childbirth.

The original version (Hollins Martin & Martin, 2014) has a good fit to the three factors, confirmed via a confirmatory factor analysis (CFA). It also has criterion-related validity and internal consistency (Cronbach's α=0.79).

Visual Analog Scale (VAS) for Birth Satisfaction

The participants were asked to move the cursor to the appropriate part of the line and rate their satisfaction with their overall childbirth experience using the VAS, with the left and right ends of the 100-mm horizontal line (gauge) as "not satisfied at all" and "very satisfied," respectively. The distance from the left end to the cursor was scored (0–100 points), and higher scores were interpreted as higher satisfaction.

Client Satisfaction Questionnaire (CSQ-8J)

The CSQ-8 was developed by Larsen et al. (1979) to measure and assess consumer satisfaction with health and human services. The CSQ-8 has no subscales and reports a single score measuring a single dimension of overall satisfaction. The scale consists of eight items rated on a 4-point Likert scale that ranges from 1 (not good) to 4 (very good). Items include questions about respondents' opinions and conclusions about services they have received or are currently receiving. The total score ranges from 8 to 32, and higher scores indicate greater satisfaction (Larsen et al., 1979). We used the Japanese 8-item version of the CSQ-8J, which was shown to have sufficient internal consistency and criterion-related validity (Tatemori & Ito, 1999). Permission was obtained to use the CSQ-8J in the Internet survey according to the method specified by the scale's developer.

Edinburgh Postnatal Depression Scale (EPDS)

The EPDS is a scale developed by Cox et al. (1987) to screen for postpartum depression.

The scale consists of 10 items and rates mental status over the past week on a 4-point Likert scale that ranges from 0 (never) to 3 (always). The total score ranges from 0 to 30. The Japanese version of the EPDS (Okano et al, 1996) has high validity and reliability. The EPDS is a face-to-face survey, and the practicality of online surveys has been under consideration (Cox, 2019). We used the EPDS to verify the discriminant validity of the BSS-R and not to screen for postpartum depression. Approval for its use was obtained from an ethics review committee.

Translation Processes of the BSS-R into Japanese

The Japanese-BSS-R was developed based on the 10 steps of a report by the International Society for Pharmacoeconomics and Outcomes Research (ISPOR) Task Force (Wild et al., 2005), which provided guidelines for the translation of patient-reported outcome measures.

First, permission for the Japanese translation was obtained from Caroline J. Hollins Martin, the developer of the scale. Subsequently, four Japanese midwifery researchers were enlisted: three translated the original version into Japanese, and the translations were discussed with the researchers. The discussions were recorded each time. During the discussion, the meanings of the words “unscathed” in Item 1 (“I came through childbirth virtually unscathed”) and “distress” in Items 7 (“I found giving birth a distressing experience”) and 9 (“I was not distressed at all during labor”) were unclear. Hence, we asked two other researchers who were native English speakers that were familiar with Japanese to confirm the meanings. The forward translation was then backtranslated into English by a native English-speaking translator not involved in the midwifery field. The back-translation was checked by the scale’s developer. We were asked to revise the wording of one question. Therefore, we reviewed the forward-translated version, and the revised version was approved. Finally, cognitive debriefing was conducted with Japanese women who had given birth, and based on the opinions obtained, the Japanese expressions were revised after a re-check with a scale’s developer to finalize the Japanese-BSS-R.

Data Analysis

We analyzed the data using the COSMIN study design checklist for patient-reported outcome measurement instruments (PROM) (Mokkink et al., 2019). SPSS Amos version 28 and SPSS version 28 were used for statistical analyses (IBM, 2021).

Descriptive statistics of participant characteristics and the Japanese BSS-R were calculated. The correlations between the total and subscale scores of the Japanese BSS-R were calculated and compared with those of the original version. (Diedenhofen & Musch, 2015).

Content Validity

Cognitive debriefing was conducted with 10 Japanese women within five years of childbirth to assess whether there were any expressions or content that were difficult to understand in the explanatory text, question items, and answer choices. Furthermore, we also assessed whether they could answer the questions based on their childbirth experiences. To obtain various opinions, participants were selected to allow for variation in parity and type of childbirth and were interviewed face-to-face using an interview guide. In the translation processes, the researchers discussed whether the questions were appropriate for the various types of childbirth in Japan and whether the translation was consistent with the Japanese language while maintaining the meaning of the original text.

Construct Validity

Structural Validity. CFA was used to evaluate the established three-dimensional measurement model of the BSS-R. Model fit was determined using the comparative fit index (CFI), root mean squared error of approximation (RMSEA), and square root mean residual (SRMR). Threshold values of >0.90 (CFI), <0.08 (RMSEA), and <0.08 (SRMR) were set to determine the model fit adequacy (Vandenberg & Lance, 2000).

Cross-cultural Validity / Measurement Invariance. Measurement invariance was assessed across parities (nulliparous vs. multiparous) and types of childbirth (vaginal childbirth vs.

CS). There were three models in the multigroup CFA: Models 1 (configural invariance model: same structure across the groups), 2 (metric model: factor loadings constrained equally across the groups), and 3 (scalar model: factor loading and intercepts constrained similarly across the groups). $\Delta CFI<0.01$ implied that the invariance assumption still held (Cheung & Rensvold, 2002).

Hypotheses Testing. To evaluate convergent validity, the correlations between the Japanese-BSS-R subscale scores and total score and the VAS for birth satisfaction and CSQ-8J were examined. Pearson's correlation coefficient (r) was calculated. Since the VAS for birth satisfaction and CSQ-8J measured the same concept of "satisfaction," it was predicted that the correlations between them would be statistically significant and moderately positive.

To evaluate divergent validity, the correlations between the Japanese-BSS-R subscale scores and total score and EPDS and participant age were examined. Pearson's correlation coefficient (r) was calculated. Since postpartum depression was a different concept, although related to birth satisfaction, we predicted that the correlations between the Japanese-BSS-R subscale scores, total score, and EPDS would be statistically significant and moderately negative, with a strength of association (absolute value of the correlation coefficient) smaller than that of the VAS or CSQ-8J. No statistically significant correlation was predicted between the Japanese-BSS-R subscale scores and total score and age at childbirth.

To evaluate known-group discriminant validity, the Japanese-BSS-R subscale scores and total score as a function of childbirth type were compared. Type of childbirth was categorized as normal vaginal childbirth, forceps or ventouse childbirth, painless childbirth, elective CS, and emergency CS. We defined painless childbirth as vaginal childbirth with regional anesthesia to relieve pain. A one-way analysis of variance (ANOVA) was used to compare the differences among these groups on the Japanese-BSS-R subscale and total scale scores. If a statistically significant effect was observed, post-hoc testing was performed using the Bonferroni correlation to control for Type 1 error. Aligned with the preceding literature review (Coates et al., 2020), it was anticipated that the emergency CS group would have significantly lower scores compared with those in the

normal vaginal childbirth and elective CS groups.

Reliability

Internal Consistency. Cronbach's α was used to assess the internal consistency of the Japanese-BSS-R subscale and total score, and a value of ≥ 0.70 was considered acceptable. Comparisons with the original validation study (Hollins Martin & Martin, 2014) were made using Diedenhofen and Musch's (2016) method, based on the Cronbach's α sampling error theory of Feldt et al. (1987), which detailed how Cronbach's α estimations could be compared and evaluated for statistically significant differences based on chi-square distribution.

Test-retest Reliability / Measurement Error. To examine test-retest reliability based on a single measurement, intraclass correlation coefficient (ICC) and its 95% confidence interval of the Japanese-BSS-R subscale and total score were calculated under a single measurement, absolute-agreement, two-way mixed-effects model (Koo & Li, 2016). ICCs between 0.50 and 0.75 and 0.75 and 0.90 were considered to have moderate and good test-retest reliability, respectively (Koo & Li, 2016). Additionally, the standard error of measurement (SEM) and the smallest detectable change (SDC) were calculated as a parameter of measurement error. Specifically, the SEM was calculated using the formula [$SEM = SD_{difference} / \sqrt{2}$], where $SD_{difference}$ is the standard deviation of the difference between the test and retest scores (De Vet et al., 2011). The $SDC_{indivisual}$ was calculated using the formula [$SDC_{indivisual} = 1.96 \times \sqrt{2} \times SEM$], and the SDC_{group} was calculated by dividing the $SDC_{indivisual}$ by the square root of the number of participants (De Vet et al., 2011).

Results

Participants

The dataset for the psychometric evaluation comprised 445 participants. The participants' characteristics are shown in Table 1.

The dataset for test-retest reliability comprised 98 participants as two of the 100

participants in the secondary survey were excluded as outliers.

Japanese-BSS-R

The descriptive and distributional characteristics of each Japanese BSS-R item are summarized in Table 2. No evidence of excessive skewness or kurtosis was observed.

The Japanese-BSS-R SE, WA, and QC subscale scores were highly correlated with the Japanese-BSS-R total score, r=0.82, p<0.001, r=0,62, p<0.001 and r=0.61, p<0.001, respectively. The SE subscale was significantly correlated with the WA (r=0.43, p<0.001) and QC (r=0.19, p<0.001) subscales. No statistically significant correlation was observed between the WA and QC subscale scores (r=0.02, p=0.63). When compared with the original BSS-R, statistically significant differences between the studies were observed on three scale combinations: (i) SE and WA subscales, (ii.) WA and QC subscales; and (iii) BSS-R total score and WA subscale. No other statistically significant differences were observed between the groups. (Table 3).

Content Validity

Since “labor,” “birth,” and “delivery” were translated into the same word in Japanese, some women in the cognitive debriefing commented that it was difficult to know which period of experience this referred to. The same comment was made regarding a CS or a painless childbirth, as the situation differed from that of a normal vaginal childbirth. Other women also commented that they could not distinguish between Items 7 and 9, where the same word “distress” was used. We discussed this with the researchers involved in the Japanese translation and the original scale’s developer; consequently, we added that respondents should answer regarding their experience from the beginning of labor until birth. If their experience did not start at labor, we asked them to use the timing of when they felt the birth began. For the latter comment, we checked the question’s intent with the scale’s developer again and modified the Japanese wording so that Items 7 and 9 referred to the second and first stages of childbirth, respectively. In addition, expressions that were not appropriate in Japanese owing to their direct translation from English were revised after discussion.

Structural Validity

Table 4 and Figure 1 show the results of the CFA. The model fit for the established three-factor measurement model did not meet these criteria (CFI=0.863, RMSEA=0.108, SRMR=0.083). Therefore, we evaluated a four-factor model with SE divided into SE1 (BSS-R1 and BSS-R9) and SE2 (BSS-R2 and BSS-R7) and found that the four-factor model offered a generally excellent fit to the data (CFI=0.941, RMSEA=0.074, SRMR=0.046).

Cross-cultural Validity / Measurement Invariance

The results (Table 5) suggested full metric invariance between nulliparas and multiparas. The configural model had a good model fit. There were negligible changes in the CFI, RMSEA, and SRMR values between the configural and metric models (0.001, 0.002, and < 0.001, respectively). However, notable changes were observed in the CFI and RMSEA values between the metric and scalar models (0.058 and -0.015, respectively). Item 4 was freed in the scalar model and showed little change in the CFI and RMSEA values. However, their changes and partial scalar invariance did not meet the criteria. The measurement invariance across types of childbirth could not be analyzed owing to the influence of elective CS data.

Hypotheses Testing

The convergent and divergent validity results are summarized in Table 6. Statistically significant (p<0.05 for SE1, otherwise p<0.01) and positive correlations were observed between the Japanese-BSS-R subscale scores and total score and the VAS for birth satisfaction (SE r=0.24, SE1 r=0.12, SE2 r=0.24, WA r=0.17, QC r=0.50, total r=0.45). Similarly, statistically significant (p<0.01) and positive correlations were observed between the Japanese-BSS-R sub-scale scores other than SE1 and the total score and CSQ-8J (SE r=0.19, SE2 r=0.22, WA r=0.15, QC r=0.56, total r=0.44).

Statistically significant (p<0.01) and negative correlations were observed between the

Japanese-BSS-R subscale scores other than SE1 and the total score and EPDS (SE r=-0.22, SE2 r=-0.33, WA r=-0.37, QC r=-0.27, total r=-0.39). No statistically significant correlations were observed between the Japanese-BSS-R SE and QC subscale and total scores and participant age. However, a statistically significant (p<0.05), small positive correlation was observed between the WA subscale score and participant age (r=0.11). WA scores were higher for older age.

The one-way ANOVA showed that the differences in the scores between the types of childbirth were statistically significant, except for the QC subscale score (SE $F(4, 440)=5.27$, p<0.001; SE1 $F(4, 440)=8.72$, p<0.001; SE2 $F(4, 440)=4.62$, p=0.001; WA $F(4, 440)=3.49$, p=0.008; total score $F(4, 440)=3.54$, p=0.007). Post-hoc testing using the Bonferroni method (Games–Howell method as SE was not equally distributed) showed that the emergency CS scored significantly lower than the normal vaginal childbirth group in the SE and SE1 subscales and the total scale. Furthermore, they also scored significantly lower than the elective CS group in the SE, SE1, and WA subscales. Women who had forceps or ventouse childbirth had significantly lower scores on the SE2 subscale than those who had normal vaginal childbirth and elective and emergency CS (Table 7).

Internal Consistency

The Cronbach's alphas for the Japanese BSS-R total scale and subscales are summarized in Table 8. Only the QC subscale was > 0.70. Compared with the original validation study (Hollins Martin & Martin, 2014), Cronbach's alphas for the total and SE subscales were significantly low. However, the Cronbach's α for the QC subscale was significantly high compared with that of the original study. No significant difference was observed between the Japanese-BSS-R WA subscale and that reported by Hollins Martin and Martin (2014).

For all vaginal childbirths, including normal, forceps or ventouse, and painless, Cronbach's α for the total score was 0.71.

Test-retest Reliability / Measurement Error

The test-retest reliability and measurement error results are summarized in Table 9. The

test-retest reliability of the Japanese-BSS-R resulted in an ICC of 0.76 (95% CI = 0.66–0.83) for the total score. The ICCs for the subscales were all between 0.5 and 0.75. The SEM for the total score was 2.65, and those of the subscales varied between 1.03 and 1.83. The $SDC_{indivisual}$ for the total score was 7.34, and for the subscales, it varied between 2.86 and 5.08. The SDC_{group} for the total score was 0.74.

Discussion

This study aimed to develop the Japanese-BSS-R and evaluate it according to the COSMIN's PROM, content, structural and cross-cultural validities, hypothesis testing (convergent, discriminant, and known-group discriminant validities), internal consistency, and test-retest reliability (Mokkink et al., 2019).

First, regarding content validity, COSMIN required that the relevance, comprehensiveness, and comprehensibility of the scale be verified (Terwee et al., 2018). In this study, the relevance and comprehensibility were subjected to cognitive debriefing, and the Japanese translation was revised based on the opinions obtained and discussed among the researchers. Thus, a certain degree of content validity was ensured. However, since we were unable to interview the participants regarding comprehensibility, this issue should be considered in the future.

Regarding structural validity, the CFA results showed that the Japanese-BSS-R had a poor fit to the three-factor model of the original version. However, the four-factor model of SE1 (Items 1 and 9), SE2 (Items 2 and 7), WA (Items 4 and 8), and QC (Items 3, 5, 6, and 10) satisfied the model fit index and showed structural validity. Most of the current translated validation studies showed a good fit to the three-factor model (Emmens et al., 2021; Nespoli et al., 2018, 2021; Ratislavová et al., 2022), but only an Indian validation study had problems with the WA sub-scale items, resulting in poor fit to the three-factor model (Tiwari et al., 2023). The Indian study states that in childbirth in India, unlike in Western countries, practicality and safety are prioritized over women's perceptions and experiences; moreover, it discusses the need to review the WA items and re-examine their fit to the three-factor model (Tiwari et al., 2023). For the Japanese-BSS-R, there are two possible reasons

for the poor fit to the three-factor model. First, culturally, the Japanese perception of childbirth may have had an influence. In Japan, it is believed that pain during childbirth is necessary to become a mother (Takegata et al., 2018). The Childbirth Experience Self-Rating Scale (Tokiwa & Imazeki, 2000), a commonly used scale in Japan, includes items such as "I endured the pain during childbirth because it was natural," "I did not make weak complaints, such as 'It hurts' or 'Help me,'" and "I did my best for the baby even though it was painful." Therefore, it was possible that Items 2 and 7 were viewed as expected stress and had different factors from Items 1 and 9. Hence, SE1 could be considered "stress experienced during labor" and SE2 "stress tolerance." BSS-R translated validation studies have been conducted in English-speaking countries, as well as in Europe and the Middle East, but studies from Asia are still limited. Further research is needed to determine whether cultural factors affect the factor structure. Another reason is that Items 2 and 7 were reversed items, which may have resulted in different expressions when translated into Japanese, forming factors different from those of Items 1 and 9. Accordingly, revising the Japanese translation and verifying the factor structure again are necessary in future studies.

Measurement invariance was not well tested in previously translated papers and was a novel finding of this study. The results suggested full metric invariance between nulliparas and multiparas. The Japanese BSS-R assessed the same factor loadings for both nulliparous and multiparous women.

In addition, the VAS showed significant positive correlations with all sub-scales and total scores, and the CSQ-8J showed significant positive correlations (especially moderate positive correlations with the QC subscale), except for SE1. Hence, the hypothesis was generally supported and confirmed to have convergent validity. The correlation coefficients with the EPDS were significantly negative, except for SE1, and the absolute values of the correlation coefficients for the total score were smaller than those of the VAS and CSQ-8J. Participant age showed a significantly weak positive correlation with the WA subscale. However, no other significant correlations were found. These results generally supported the divergent validity. Regarding known-group discriminant validity, the emergency CS group had significantly lower SE, SE1 subscale, and total

scale scores than the normal vaginal childbirth group. By contrast, for emergency CS and elective CS, the SE, SE1, and WA subscales were significantly lower in the emergency CS group. However, there was no significant difference in the total scale scores. The small sample size of 48 for emergency CS and the lower power of the Bonferroni method may have contributed to this difference. Since the hypothesis itself was supported, it was confirmed to have known-group discriminant validity.

Regarding internal consistency, Cronbach's α was considerably lower than that in the original study and other translated studies, except for QC, which did not exceed the criterion of 0.70. Similar results were observed in the Italian and Indian validation studies (Nespoli et al., 2021; Tiwari et al., 2023). Since Cronbach's α is lower for factors consisting of two items (Eisinga et al., 2012), Cronbach's α may have been lower for factors consisting of two items in the Japanese-BSS-R. However, the alpha coefficient of the total score was also low, and sufficient internal consistency could not be confirmed. Furthermore, when limited to vaginal childbirths, Cronbach's α of the total score exceeded the 0.70 criterion. It was possible that the questions were not worded appropriately or that items not related to satisfaction with CS childbirth were included. Therefore, these groups should be interviewed again for further discussion.

Finally, regarding test-retest reliability, the ICC for the total score confirmed good test-retest reliability. The measurement error results suggested that the Japanese-BSS-R total score of an individual would have to change with at least 7.34 points (on a scale of 0–40) before the observed change can be considered a true change in a participant's birth satisfaction and not potentially a result of measurement error. However, the ICCs of the subscales were moderate at 0.75 or less. This may have been influenced by the fact that the survey was conducted when the perception of childbirth was likely to change owing to postnatal hormonal balance and environmental changes. In future studies, we believe that more accurate ICC, SEM, and SDC can be calculated by identifying the unchanged population in the retest using an anchor scale that evaluates the change in birth satisfaction during the test and retest (Kamber et al., 2009; Storheim et al., 2012).

In conclusion, the Japanese-BSS-R was a valid and reliable scale, with the exception of

internal consistency, which requires further investigation. However, internal consistency was confirmed in vaginal childbirth, which could be applied to research, clinical practice, and international comparisons.

There are several limitations to this study. First, the Japanese-BSS-R fits the four-factor model well, but SE1, SE2, and WA become two-item factors. Research suggests that at least 3-4 items should be included in the factor because a higher number of items will result in higher validity and reliability (Emons et al., 2007; Marsh et al., 1998). However, the increased need for short scales to conduct large-scale social surveys has resulted in more scales with fewer items than before, and there are many psychological scales with two-item factors, such as the original BSS-R (Rammstedt, 2014; Schweizer, 2011). These short scales also demonstrate sufficient reliability, leaving room for improvement when reliability is low, as with the Japanese-BSS-R in this study. Therefore, in the future, to use the Japanese-BSS-R widely, we should review the Japanese expressions concerning comprehensiveness and the interpretation of reversal item questions and evaluate the reliability and validity again. Another issue is the representativeness of the sample associated with the Internet survey (Ball, 2019; Bethlehem, 2010). This study was conducted on a registered panel of an Internet research firm, and respondents were offered monetary incentives in exchange; thus, the sample was likely biased. The distribution of age and parity among the participants in this study was close to that of all births in Japan in 2020 (Minister of Health, Labour and Welfare of Japan, 2021). However, this study assigned 150 participants to the CS to examine cross-cultural validity, which resulted in a higher proportion of CS than all births in Japan (18.6% in 2014) (Maeda et al., 2021). These sample biases may have contributed to the poor model fit to the three factors, as well as low internal consistency. Finally, the credibility of Internet survey data is also an issue (Ball, 2019). An Internet survey firm and monitors were contracted, and survey companies also took measures to eliminate short-time and straight-line responses. However, to further increase the credibility of the data, the consistency with other questions should be examined and trap questions should be established to detect satisficers. Based on these limitations, new methods of data collection and sampling should be devised in future studies to generalize the study results.

Conclusions

The Japanese-BSS-R was found to be a valid and reliable translation of the original BSS-R with the exception of internal consistency, which requires further investigation. If limited to vaginal childbirth, research, clinical applications, and international comparisons can be drawn. However, to use the Japanese-BSS-R widely, participants should be interviewed regarding the comprehensiveness of the questions and translation of the reversed items, and the scale's reliability and validity should be re-evaluated.

Author Contributions

All authors meet the criteria for authorship and are listed as authors.

Aya Tezuka, Natsuko Hiroyama,Megumi Matsuoka and Colin R. Martin contributed to the conception and design of this study. Aya Tezuka collected the data. All authors contributed to data analysis and interpretation. Aya Tezuka and Natsuko Hiroyama drafted the manuscript. All authors reviewed the manuscript draft and revised it critically on intellectual content. All authors have given final approval to the manuscript and also agreed to be responsible for all aspects of this research.

Acknowledgements

We want to express our deepest gratitude to all the women who participated in our study, Professor Robin Sakamoto of Kyorin University, and Assistant Professor Janelle Moross of Tokyo Medical and Dental University, for consultation on the Japanese translation. In addition, we would like to thank Editage (www.editage.com) for English language editing. This work was supported by the JSPS KAKENHI (grant number JP17K17485) and by the operating grant from Department of Reproductive Health Nursing, Graduate School of Health Care Sciences, Tokyo Medical and Dental University (TMDU).

Disclosure and Conflict of Interests

The authors have no potential conflicts of interest to declare.

References

Ball, H. L. (2019). Conducting online surveys. *Journal of Human Lactation*, *35*(3), 413–417. https://doi.org/10.1177/0890334419848734.

Bell, A. F., & Andersson, E. (2016). The birth experience and women's postnatal depression: A systematic review. *Midwifery*, *39*, 112–123. https://doi.org/10.1016/j.midw.2016.04.014.

Bethlehem, J. (2010). Selection bias in web surveys. *International Statistical Review*, *78*(2), 161–188. https://doi:10.1111/j.1751-5823.2010.00112.x.

Britton, J. R. (2012). The assessment of satisfaction with care in the perinatal period. *Journal of Psychosomatic Obstetrics and Gynaecology*, *33*(2), 37–44. https://doi.org/10.3109/0167482X.2012.658464.

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

Coates, D., Thirukumar, P., & Henry, A. (2020). Women's experiences and satisfaction with having a cesarean birth: An integrative review. *Birth*, *47*(2), 169–182. https://doi.org/10.1111/birt.12478.

Cox, J. (2019). Thirty years with the Edinburgh Postnatal Depression Scale: Voices from the past and recommendations for the future. *British Journal of Psychiatry*, *214*(3), 127–129. https://doi.org/10.1192/bjp.2018.245.

Cox, J. L., Holden, J. M., & Sagovsky, R. (1987). Detection of postnatal depression. Development of the 10-item Edinburgh Postnatal Depression Scale. *British Journal of Psychiatry*, *150*, 782–786. https://doi.org/10.1192/bjp.150.6.782.

De Vet, H. C. W., Terwee, C., Mokkink, L. B., & Knol, D. L. (2011). *Measurement in medicine: A practical guide*. Cambridge: Cambridge University Press.

Diedenhofen, B., & Musch, J. (2015). cocor: A comprehensive solution for the statistical comparison of correlations. *PLOS ONE*, *10*(3), e0121945. https://doi.org/10.1371/journal.pone.0121945.

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

Eisinga, R., Grotenhuis, Mt., & Pelzer, B. (2013). The reliability of a two-item scale: Pearson, Cronbach, or Spearman–Brown? *International Journal of Public Health*, *58*(4), 637–642. https://doi.org/10.1007/s00038-012-0416-3.

Emmens, B., Hollins Martin, C. J., & Martin, C. R. (2021). Translation and validation of

the Dutch version of the Birth Satisfaction Scale-Revised (BSS-R). *Journal of Reproductive and Infant Psychology*, 1–15. https://doi.org/10.1080/02646838.2021.1979200.

Emons, W. H., Sijtsma, K., & Meijer, R. R. (2007). On the consistency of individual classification using short scales. *Psychological Methods*, *12*(1), 105–120. https://doi.org/10.1037/1082-989X.12.1.105.

Feldt, L. S., Woodruff, D. J., & Salih, F. A. (1987). Statistical inference for coefficient alpha. *Applied Psychological Measurement*, *11*(1), 93–103. https://doi.org/10.1177/014662168701100107.

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

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. https://doi.org/10.1080/02646838.2012.710833.

Kamper, S. J., Maher, C. G., & Mackay, G. (2009). Global rating of change scales: A review of strengths and weaknesses and considerations for design. *Journal of Manual and Manipulative Therapy*, *17*(3), 163–170. https://doi.org/10.1179/jmt.2009.17.3.163.

Karlström, A., Nystedt, A., & Hildingsson, I. (2015). The meaning of a very positive birth experience: Focus groups discussions with women. *BMC Pregnancy and Childbirth*, *15*, 251. https://doi.org/10.1186/s12884-015-0683-0.

Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine*, *15*(2), 155–163. https://doi.org/10.1016/j.jcm.2016.02.012.

Larsen, D. L., Attkisson, C. C., Hargreaves, W. A., & Nguyen, T. D. (1979). Assessment of client/patient satisfaction: Development of a general scale. *Evaluation and Program Planning*, *2*(3), 197–207. https://doi.org/10.1016/0149-7189(79)90094-6.

Maeda, E., Ishihara, O., Tomio, J., Miura, H., Kobayashi, Y., Terada, Y., Murata, K., & Nomura, K. (2021). Cesarean delivery rates for overall and multiple pregnancies in Japan: A descriptive study using nationwide health insurance claims data. *Journal of Obstetrics and Gynaecology Research*, *47*(6), 2099–2109. https://doi.org/10.1111/jog.14772.

Mahalanobis, P. C. (1936). On the generalised distance in statistics. *Proceedings of the National Institute of Sciences in India*, *2*(1), 49–55.

Marsh, H. W., Hau, K. T., Balla, J. R., & Grayson, D. (1998). Is more ever too much? The number of indicators per factor in confirmatory factor analysis. *Multivariate Behavioral Research*, *33*(2), 181–220. https://doi.org/10.1207/s15327906mbr3302_1.

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

Martin, C. R., Jefford, E., & Hollins Martin, C. J. (2020). Crisis, what crisis? Replicability of the key measurement characteristics of the Australian version of the Birth Satisfaction Scale-Revised. *International Journal of Childbirth*, *10*(3), 140–150. https://doi.org/10.1891/IJCBIRTH-D-20-00006.

McKelvin, G., Thomson, G., & Downe, S. (2021). The childbirth experience: A systematic review of predictors and outcomes. *Women and Birth*, *34*(5), 407–416. https://doi.org/10.1016/j.wombi.2020.09.021.

Minister of Health, Labour and Welfare of Japan. (2021). *Handbook of health and welfare statistics*. https://www.mhlw.go.jp/english/database/db-hh/1-2.html.

Mokkink, L. B., Prinsen, C. A. C., Patrick, D. L., Alonso, J., Bouter, L. M., De Vet, H. C. W., & Terwee, C. B. (2019). *COSMIN Study Design checklist for Patient-reported outcome measurement instruments.* COnsensus-based Standards for the selection of health Measurement Instruments. https://www.cosmin.nl/wp-content/uploads/COSMIN-study-designing-checklist_final.pdf.

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

Nespoli, A., Colciago, E., Pedroni, S., Perego, S., & Fumagalli, S. (2018). The Birth Satisfaction Scale-Revised (BSS-R): Process of translation and adaptation in an Italian context. *Annali dell'Istituto Superiore di Sanita*, *54*(4), 340–347. https://doi.org/10.4415/ANN_18_04_11.

Nijagal, M. A., Wissig, S., Stowell, C., Olson, E., Amer-Wahlin, I., Bonsel, G., Brooks, A., Coleman, M., Devi Karalasingam, S., Duffy, J. M. N., Flanagan, T., Gebhardt, S., Greene, M. E., Groenendaal, F., R Jeganathan, J. R., Kowaliw, T., Lamain-de-Ruiter, M., Main, E., Owens, M., Petersen, R., . . . Franx, A. (2018). Standardized

outcome measures for pregnancy and childbirth, an ICHOM proposal. *BMC health services research*, *18*(1), 953. https://doi-org.ezproxy.kyorin-u.ac.jp/10.1186/s12913-018-3732-3.

Okano, T., Murata, M., Masuji, F., Tamaki, R., Nomura, J., Miyaoka, H., & Kitamura, T. (1996). Validation and reliability of a Japanese version of the EPDS. *Archives of Psychiatric Diagnosis and Clinical Evaluation*, *7*, 525–533. (in Japanese).

Rammstedt, B., & Beierlein, C. (2014). Can't we make it any shorter? The limits of personality assessment and ways to overcome them. *Journal of Individual Differences*, *35*(4), 212–220. https://doi-org.ezproxy.kyorin-u.ac.jp/10.1027/1614-0001/a000141.

Ratislavová, K., Hendrych Lorenzová, E., Hollins Martin, C. J., & Martin, C. R. (2022). Translation and validation of the Czech Republic version of the Birth Satisfaction Scale-Revised (BSS-R). *Journal of Reproductive and Infant Psychology*, 1–17. https://doi.org/10.1080/02646838.2022.2067837.

Rudman, A., El-Khouri, B., & Waldenström, U. (2007). Women's satisfaction with intrapartum care – A pattern approach. *Journal of Advanced Nursing*, *59*(5), 474–487. https://doi.org/10.1111/j.1365-2648.2007.04323.x.

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. https://doi.org/10.1186/1471-2393-13-108.

Schweizer, K. (2011). Some thoughts concerning the recent shift from measures with many items to measures with few items. *European Journal of Psychological Assessment*, *27*(2), 71–72. https://doi-org.ezproxy.kyorin-u.ac.jp/10.1027/1015-5759/a000056.

Simpson, M., & Catling, C. (2016). Understanding psychological traumatic birth experiences: A literature review. *Women and Birth*, *29*(3), 203–207. https://doi.org/10.1016/j.wombi.2015.10.009.

Storheim, K., Brox, J. I., Løchting, I., Werner, E. L., & Grotle, M. (2012). Cross-cultural adaptation and validation of the Norwegian version of the Core Outcome Measures Index for low back pain. *European Spine Journal*, *21*(12), 2539–2549. https://doi.org/10.1007/s00586-012-2393-x.

Takegata, M., Haruna, M., Morikawa, M., Yonezawa, K., Komada, M., & Severinsson, E. (2018). Qualitative exploration of fear of childbirth and preferences for mode of birth among Japanese primiparas. *Nursing and Health Sciences*, *20*(3), 338–345. https://doi.org/10.1111/nhs.12571.

Tatemori, H., & Ito, H. (1999). Reliability and validity of the Japanese version on client satisfaction questionnaire. *Clinical Psychiatry*, *41*, 711–717. https://doi.org/10.11477/mf.1405905056. (in Japanese).

Terwee, C. B., Prinsen, C. A. C., Chiarotto, A., De Vet, H. C. W., Bouter, L. M., Alonso, J., … Mokkink, L. B. (2018). *COSMIN methodology for assessing the content validity of PROMs user manual version 1.0.* COnsensus-based Standards for the selection of health Measurement Instruments. https://www.cosmin.nl/wp-content/uploads/COSMIN-methodology-for-content-validity-user-manual-v1.pdf.

Tiwari, S. K., Murry, L., Joshi, P., Tallanao, T., Zined, R., Hollins Martin, C. J., & Martin, C. R. (2023). Translation and validation of the Hindi-Indian version of the Birth Satisfaction Scale-Revised. *Journal of Obstetrics and Gynaecology Research*, *49*(3), 938–945. https://doi: 10.1111/jog.15520.

Tokiwa, Y., & Imazeki, S. (2000). Design of self-evaluation scale for experience of delivery and study of its reliability and validity. *Journal of Japan Academy of Nursing Science*, *20*(1), 1–9. https://doi.org/10.5630/jans1981.20.1_1. (in Japanese).

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*(1), 4–70. https://doi.org/10.1177/109442810031002.

Waldenström, U., & Rudman, A. (2008). Satisfaction with maternity care: How to measure and what to do. *Women's Health*, *4*(3), 211–214. https://doi.org/10.2217/17455057.4.3.211.

Wild, D., Grove, A., Martin, M., Eremenco, S., McElroy, S., Verjee-Lorenz, A., Erikson, P., & ISPOR Task Force for Translation and Cultural Adaptation. (2005). Principles of Good Practice for the Translation and Cultural Adaptation Process for Patient-Reported Outcomes (PRO) Measures: Report of the ISPOR Task Force for Translation and Cultural Adaptation. *Value in Health*, *8*(2), 94–104. https://doi.org/10.1111/j.1524-4733.2005.04054.x.

World Health Organization. (2018). *WHO recommendations Intrapartum care for a positive childbirth experience*. https://apps.who.int/iris/bitstream/handle/10665/260178/9789241550215-eng.pdf.

Figure Legends

Figure 1

Standardized factor loadings of the four-dimensional measurement of the Japanese-BSS-R.

Note: The arrows leading from the factor (*circle*) to the BSS-R item (*box*) represent the standardized regression coefficient weight of the factor on the indicator. The value of each box (*BSS-R item*) indicates the proportion of variance of the item explained by the factor. The double-headed arrow between the factors represents the covariance between factors.

Abbreviations: BSS-R: Birth Satisfaction Scale-Revised, SE1: stress experienced during labor, SE 2: stress tolerance, WA: women's personal attributes, QC: quality of care, CFI: comparative fit index, RMSEA: root mean square error of approximation, SRMR: standardized root mean square residual

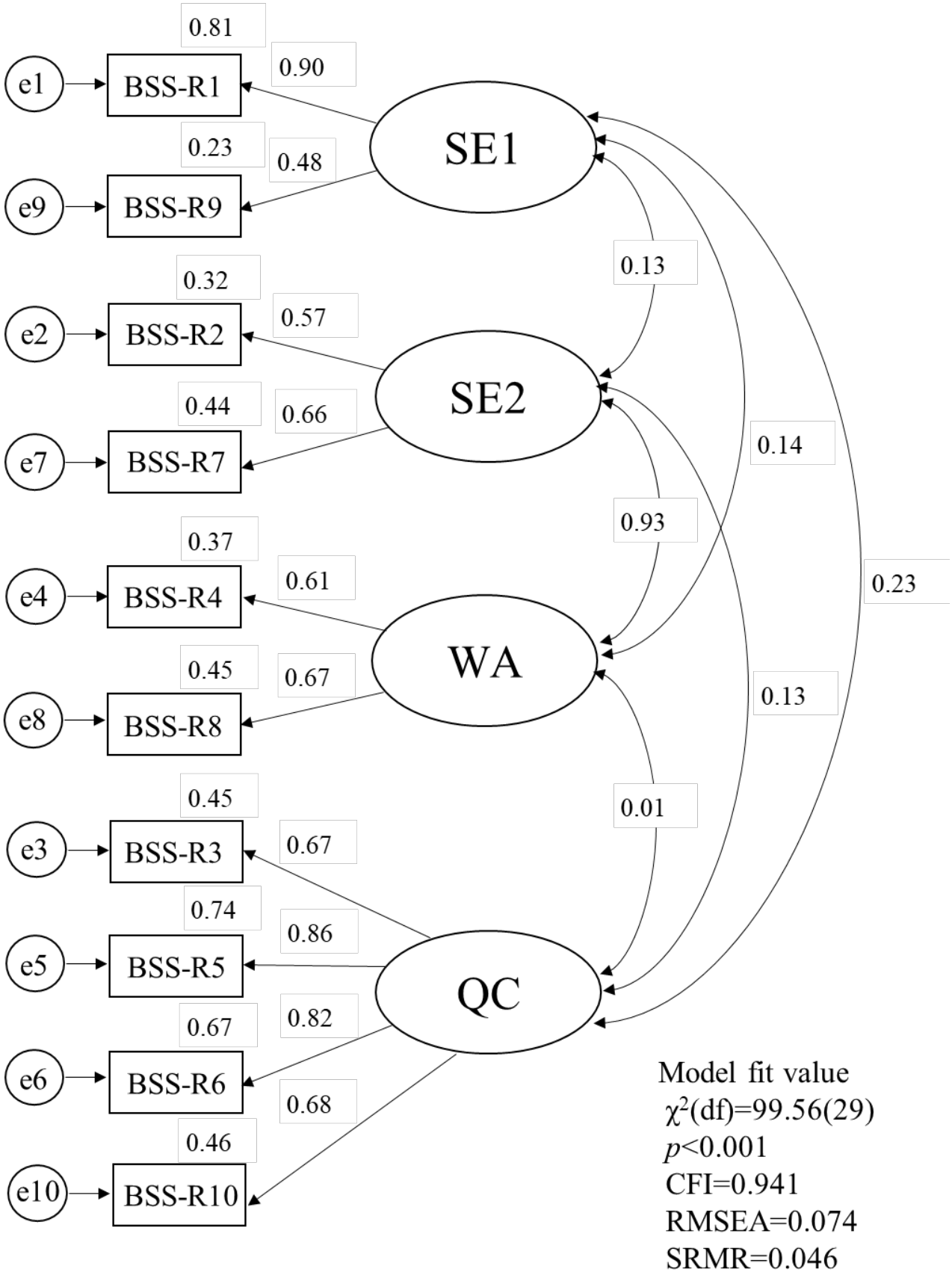
0.81
0.90
e1
BSS-R1
0.23
0.48
e9
BSS-R9
SE1
0.13
0.32
0.57
e2
BSS-R2
SE2
0.44
0.66
e7
BSS-R7
0.14
0.93
0.37
0.61
e4
BSS-R4
0.23
WA
0.45
0.67
e8
BSS-R8
0.13
0.01
0.45
0.67
e3
BSS-R3
0.74
0.86
e5
BSS-R5
QC
0.67
0.82
e6
BSS-R6
0.68
0.46
e10
BSS-R10
Model fit value
χ²(df)=99.56(29)
p<0.001
CFI=0.941
RMSEA=0.074
SRMR=0.046

Table1. Participants' characteristics (n=445).

Age (years)	n (%)
20-24	28 (6.3%)
25-29	114 (25.6%)
30-34	163 (36.6%)
34-39	108 (24.3%)
40-44	29 (6.5%)
45-49	3 (0.7%)
Parity	**n (%)**
First child	231 (51.9%)
Second child	148 (33.3%)
Third child or more	66 (14.8%)
Type of childbirth	**n (%)**
Normal Vaginal	238 (53.5%)
Forceps or Ventouse	35 (7.9%)
Painless[a]	23 (5.2%)
Elective CS	101 (22.7%)
Emergency CS	48 (10.8%)
Birthing unit	**n (%)**
Hospital	305 (68.5%)
Clinic	118 (26.5%)
Maternity home	22 (4.9%)
Home	0 (0%)

† Painless childbirth is vaginal childbirth with regional anesthesia to relieve pain.

‡ CS: caesarean section

Table 2. Descriptive and distributional characteristics of the Japanese-BSS-R (n=445).

Item	Item content	Domain	Mean	SD	Median (IQR)	Min	Max	Skew	Kurtosis	SE
BSS-R 1	I came through childbirth virtually unscathed.	Stress	2.0	±1.3	2 (1-3)	0	4	0.05	-1.16	0.06
BSS-R 2	I thought my labor was excessively long.	Stress	2.6	±1.2	3 (2-4)	0	4	-0.59	-0.60	0.06
BSS-R 3	The delivery room staff encouraged me to make decisions about how I wanted my birth to progress.	Quality	2.9	±0.9	3 (2-4)	0	4	-0.44	-0.28	0.04
BSS-R 4	I felt very anxious during my labor and birth.	Attributes	1.7	±1.2	2 (1-3)	0	4	0.30	-0.78	0.05
BSS-R 5	I felt well supported by staff during my labor and birth.	Quality	3.3	±0.88	3 (3-4)	0	4	-1.18	1.41	0.04
BSS-R 6	The staff communicated well with me during labor.	Quality	3.2	±0.8	3 (3-4)	0	4	-0.73	0.14	0.04
BSS-R 7	I found giving birth a distressing experience.	Stress	1.9	±1.3	2 (1-3)	0	4	0.13	-0.99	0.06
BSS-R 8	I felt out of control during my birth experience.	Attributes	2.2	±1.2	2 (1-3)	0	4	-0.25	-0.89	0.06
BSS-R 9	I was not distressed at all during labor.	Stress	1.5	±1.2	1 (1-2)	0	4	0.41	-0.62	0.06
BSS-R 10	The delivery room was clean and hygienic.	Quality	3.3	±0.8	4 (3-4)	1	4	-0.84	-0.15	0.04
Stress			8.1	±3.1		0	16	-0.04	0.12	0.15
Attributes			3.9	±2.0		0	8	-0.03	-0.61	0.09
Quality			12.7	±2.7		3	16	-0.70	0.29	0.13
Total			24.7	±5.4		8	40	0.26	0.06	0.26

† BSS-R: Birth Satisfaction Scale-Revised, SD: standard deviation, SE: standard error, Stress: stress experienced during labor, Attributes: women's personal attributes, Quality: the quality of care

Table 3. Correlations of the Japanese-BSS-R sub-scales and total score (n=445) and comparison with the original UK BSS-R validation study (n=228).

Scale combination	Japanese r	UK[a] r	Z	95% CI	*p*	
SE-WA	0.43	0.57	2.29	(-0.25, -0.02)	0.02	*
SE-QC	0.19	0.26	0.90	(-0.22, 0.08)	0.37	
WA-QC	0.02	0.35	4.22	(-0.47, -0.18)	<0.001	**
Total score-SE	0.82	0.86	1.67	(-0.08, 0.01)	0.10	
Total score-WA	0.62	0.80	4.56	(-0.25, -0.10)	<0.001	**
Total score-QC	0.61	0.63	0.40	(-0.12, 0.08)	0.69	

† [a] Hollins Martin, C. J., Martin, C. R., 2014. Development and psychometric properties of the Birth Satisfaction Scale-Revised (BSS-R). Midwifery, 30(6), 610-619.

‡ BSS-R: Birth Satisfaction Scale-Revised, CI: confidence interval, SE: stress experienced during labor, WA: women's personal attributes, QC: the quality of care

§ *p<0.05, **p<0.01

Table 4. Confirmatory factor analysis and model fit of the Japanese-BSS-R.

Model	χ^2(df)	*p*	CFI	RMSEA	SRMR
Single factor	485.00(35)	<0.001	0.626	0.170	0.144
Two-factor	204.75(34)	<0.001	0.858	0.106	0.086
Three-factor	197.42(32)	<0.001	0.863	0.108	0.083
Four-factor	99.56(29)	<0.001	0.941	0.074	0.046

† BSS-R: Birth Satisfaction Scale-Revised, df: degrees of freedom, CFI: comparative fit index, RMSEA: root mean square error of approximation, SRMR: standardized root mean square residual

Table 5. Measurement invariances of the Japanese-BSS-R across parity groups (nulliparous vs multiparous).

Model	χ^2(df)	*p*	CFI	RMSEA	SRMR
Nulliparous	68.169(29)	<0.001	0.934	0.077	0.050
Multiparous	60.310(29)	0.001	0.945	0.071	0.054
Configural	128.48(58)	<0.001	0.939	0.052	0.050
Metric	135.19(64)	<0.001	0.938	0.050	0.050
Scalar	212.85(74)	<0.001	0.880	0.065	0.052
Partial Scalar (except item 4)	193.23(73)	<0.001	0.896	0.061	0.051

† BSS-R: Birth Satisfaction Scale-Revised, df: degrees of freedom, CFI: comparative fit index, RMSEA: root mean square error of approximation, SRMR: standardized root mean square residual

Table 6. Correlations between the Japanese-BSS-R total score and sub-scale scores and VAS of birth satisfaction, CSQ-8J, EPDS, and participant age.

BSS-R scale	VAS		CSQ-8J		EPDS		participant age	
SE	0.24	**	0.19	**	-0.22	**	0.09	
SE1	0.12	*	0.06		0.001		0.07	
SE2	0.24	**	0.22	**	-0.33	**	0.06	
WA	0.17	**	0.15	**	-0.37	**	0.11	*
QC	0.50	**	0.56	**	-0.27	**	-0.01	
Total	0.45	**	0.44	**	-0.39	**	0.09	

† BSS-R: Birth Satisfaction Scale-Revised, VAS: Visual Analog Scale, CSQ-8J: Client Satisfaction Questionnaire, EPDS: Edinburgh Postnatal Depression Scale, SE: stress experienced during labor, SE 1: stress experienced during labor, SE 2: stress tolerance, WA: women's personal attributes, QC: quality of care,

‡ *p<0.05, **p<0.01

Table 7. Comparison of Japanese-BSS-R total and sub-scale scores by childbirth type.

	Normal vaginal (n=238)	Forceps/Ventouse (n=35)	Painless (n=23)	Elective CS (n=101)	Emergency CS (n=48)				
BSS-R Scale	Mean(SD)	Mean(SD)	Mean(SD)	Mean(SD)	Mean(SD)	F	*p*	ω2	Effect size
SE	8.5(±3.2)$_{a}$	6.3(±3.8)$_{c,d}$	7.6(±2.6)$_{a,b,c}$	8.3(±2.5)$_{a,b}$	6.9(±2.7)$_{c,d}$	5.27	<0.001	0.04	Small
SE1	4.0(±2.0)$_{a}$	3.0(±2.5)$_{a,b,c,d}$	3.2(±2.0)$_{a,b,c}$	3.4(±2.0)$_{a,b}$	2.3(±1.9)$_{c,d}$	8.72	<0.001	0.07	Small
SE2	4.5(±2.0)$_{a,b,c}$	3.3(±2.1)$_{d}$	4.4(±1.7)$_{a,b,c,d}$	4.9(±2.0)$_{a}$	4.7(±1.9)$_{a,b}$	4.62	0.001	0.03	Small
WA	4.0(±2.0)$_{a,b,c}$	3.5(±1.6)$_{a,b,c,d}$	4.5(±1.7)$_{a}$	4.3(±2.0)$_{a,b}$	3.2(±2.0)$_{a,c,d}$	3.49	0.008	0.02	Small
QC	12.7(±2.7)	13.1(±2.2)	12.7(±2.9)	12.7(±2.9)	12.5(±2.5)	0.21	0.931	-0.007	Negligible
Total score	25.2(±5.7)$_{a,b}$	22.9(±5.4)$_{a,b,c,d}$	24.8(±4.2)$_{a,b,c}$	25.3(±4.9)$_{a}$	22.7(±4.8)$_{a,c,d}$	3.54	0.007	0.02	Small

† The one-way analysis of variance was used. Post-hoc testing was performed using the Bonferroni correlation. Different alphabets indicate statistically significant (p<0.05).

‡ Painless childbirth is vaginal childbirth with regional anesthesia to relieve pain.

§ BSS-R: Birth Satisfaction Scale-Revised, CS: caesarean section, SD: standard deviation, SE: stress experienced during labor, SE1: stress experienced during labor, SE2: stress tolerance, WA: women's personal attributes, QC: quality of care

Table 8. Cronbach's alpha of the Japanese-BSS-R sub-scales and total score and comparison with the original UK BSS-R study.

BSS-R scale	Japanese	UK[a]	χ^2	*p*
SE	0.47	0.71	15.10	<0.001 **
SE1	0.61			
SE2	0.55			
WA	0.58	0.64	0.58	0.45
QC	0.84	0.74	11.06	<0.001 **
Total score	0.67	0.79	11.85	<0.001 **

† [a] Hollins Martin, C. J., Martin, C. R., 2014. Development and psychometric properties of the Birth Satisfaction Scale-Revised (BSS-R). Midwifery, 30(6), 610-619.

‡ BSS-R: Birth Satisfaction Scale-Revised, SE: stress experienced during labor, SE 1: stress experienced during labor, SE 2: stress tolerance, WA: women's personal attributes, QC: quality of care

§ **$p<0.01$

Table 9. ICC, SEM, and SDC of the Japanese-BSS-R sub-scales and total score.

BSS-R scale	ICC	95%CI	SEM	$SDC_{individual}$	SDC_{group}
SE	0.67	(0.54, 0.76)	1.83	5.08	0.51
SE1	0.56	(0.40, 0.68)	1.41	3.92	0.40
SE2	0.63	(0.49, 0.73)	1.21	3.35	0.34
WA	0.72	(0.61, 0.80)	1.03	2.86	0.29
QC	0.56	(0.41, 0.69)	1.54	4.27	0.43
Total score	0.76	(0.66, 0.83)	2.65	7.34	0.74

†ICC: intraclass correlation coefficient, SEM: standard error of measurement, SDC: smallest detectable change, BSS-R: Birth Satisfaction Scale-Revised, CI: confidence interval, SE: stress experienced during labor, SE1: stress experienced during labor, SE2: stress tolerance, WA: women's personal attributes, QC: quality of care