



Postpartum fatigue: assessing and improving the psychometric properties of the Fatigue Severity Scale

Nathan Wilson¹ · Karen Wynter^{2,3} · Jane Fisher^{3,4} · Bei Bei¹

Received: 22 October 2017 / Accepted: 6 February 2018 / Published online: 17 February 2018
© Springer-Verlag GmbH Austria, part of Springer Nature 2018

Abstract

A brief, reliable, and valid measure of postpartum fatigue in both clinical and research settings is much needed. This study assessed the psychometric properties of the Fatigue Severity Scale (FSS) among women with unsettled infants and fatigue. Rasch analysis identified an overall ceiling effect for the original FSS, as well as disordered response categories and/or poor fit for 4 items. A modified 5-item version of the FSS (FSS-5R) had improved psychometric properties.

Keywords Fatigue · Postpartum · Postnatal · Depression · Validity · Reliability · Rasch · Psychometric properties

Introduction

Given the prevalence of fatigue symptoms among parents in the weeks and months after birth, there is a need for brief, reliable, and valid measures of postpartum fatigue to facilitate accurate diagnosis and guide treatments. Currently, many fatigue scales are in use, but few have been validated in postpartum samples (e.g., Giallo et al. 2014). The Fatigue Severity Scale (FSS; Krupp et al. 1989) is a widely used scale for measuring fatigue-related impairment in clinical populations, with demonstrated reliability, validity, and sensitivity to treatment (Whitehead 2009), yet to be validated in the postpartum period.

The aim was to (1) examine psychometric properties of the FSS among women admitted to a residential early parenting

program for management of unsettled infant behaviors (UIB) and (2) identify potential improvements to the psychometric properties of the FSS. The sample of women seeking help with UIB (e.g., persistent crying, resistance to soothing, and sleep difficulties) used in this study is ideal for assessing the FSS in a help-seeking postpartum population, because many women in this predicament commonly experience fatigue, depression, and anxiety symptoms (Fisher et al. 2002, 2011).

Method

Participants and procedure

Participants were recruited from the Masada Private Hospital Mother Baby Unit (MPHMBU) in Melbourne, Australia, on admission to a non-psychiatric residential early parenting program for management of UIB. There were no exclusion criteria. Ethics approval was obtained from the Avenue Hospital Research Ethics Committee and Monash University Human Research Ethics Committee.

Measures

The FSS is a 9-item scale of fatigue severity and interference with functioning that uses response options ranging from 1 (“Strongly Disagree”) to 7 (“Strongly Agree”). FSS total scores range from 7 to 63, with scores ≥ 36 suggesting clinically elevated fatigue. The Fatigue Visual Analog Scale (FVAS; Insana and Montgomery-Downs 2010) is a single-

✉ Bei Bei
bei.bei@monash.edu

¹ Monash Institute of Cognitive and Clinical Neurosciences, School of Psychological Sciences, Faculty of Medicine, Nursing and Health Sciences, Monash University, 18 Innovation Walk, Clayton Campus, Clayton, Victoria 3800, Australia

² Centre for Quality and Patient Safety Research-Western Health Partnership, School of Nursing and Midwifery, Deakin University, Burwood, Victoria, Australia

³ Jean Hailes Research Unit, School of Public Health and Preventative Medicine, Monash University, Clayton, Victoria, Australia

⁴ Masada Private Hospital, Mother Baby Unit, East St Kilda, Clayton, Victoria, Australia

item measure of fatigue. The question “How fatigued do you feel right now?” is rated between 0 (“Not at all”) and 100 (“Very Fatigued”) using a 10-cm line. The Depression Anxiety Stress Scale–depression subscale (DASS21-D; Lovibond and Lovibond 1995) and Edinburgh Postnatal Depression Scale (EPDS; Cox et al. 1987) are widely used measures of depressive symptoms. Pre-admission EPDS scores were from consenting participants’ medical records ($n = 141$). Alpha for the DASS21-D was 0.88 and for the EPDS was 0.81. Instrumental social support was assessed with the following item: How often do you feel that you need practical support or help but can’t get it from anyone? (0 = “Never”; 3 = “Very Often”).

Analysis

Rasch analysis (RA) was used to assess the unidimensionality, validity, and reliability of the FSS. The RA was conducted in Winsteps 3.92.1 (Linacre 2016) using a partial credit model. Published RA criteria for acceptable psychometric properties for fatigue scales were used (Lerdal et al. 2016; Mills et al.

2009; Ottonello et al. 2016; see Table 1). Convergent validity was assessed using correlations between the Rasch-generated person scores and FVAS, discriminant validity using the Rasch scores and DASS21-D, EPDS, and social support, all measures of constructs distinct to fatigue.

Results

In total, 44% of women ($N = 167$) admitted to the MPHMBU unit between 1 June 2015 and 12 October 2015 participated on the day of admission. On average, participants were 34.26 years ($SD = 4.23$) and infants, 8.51 months ($SD = 4.16$, range = 2 to 23.50). Seventy percent of participants were born in Australia, 87% spoke English at home, 99% were married or living with a partner, 77% were university educated, and 35% reported past treatment for a psychiatric disorder. There was minimal (0.27%) missing data for FSS items, and the mean total FSS score was high at 47.92 ($SD = 8.85$, range = 21 to 63), with 89% of women scoring above the clinical cutoff. Mean scores were also elevated on the FVAS

Table 1 Rasch analysis criteria and results for FSS and FSS-5R

Aspect measured	Criteria ^a	FSS	FSS-5R
Step 1. Rating scale functioning	a) ≥ 10 observations per category for scale responses b) Item step categories advance monotonically. c) Outfit mean square (MnSq) values < 2.0 logits for step category calibrations	a) < 10 observations for at least one category for all scale items b) Item step categories failed to advance monotonically for items 3, 6, 7 and 9. c) Outfit MnSq < 2.0	a) Item 4 has less than 10 ($n = 7$) observations for category 2, all others have > 10 b) Item steps advance monotonically for all items c) Outfit MnSq < 2.0
Step 2. Internal scale validity	For item goodness of fit statistics: d) Infit MnSq values $0.7 > \text{to} < 1.3$ logits $z < 2.0$	d) Item 2: Infit MnSq = 2.08 ($z = 7.8$) and item 7 Infit MnSq = 0.69 ($z = -3.00$)	d) All $0.7 > \text{to} < 1.3$ logits
Step 3. Interval scale validity	e) $\geq 50\%$ of the variance explained by first latent variable in principal component analysis of the residuals f) Eigenvalue of first contrast in PCA of residuals ≤ 1.7	e) 60.2% variance explained. f) Eigenvalue of 1st contrast = 1.95	e) 61.0% variance explained f) Eigenvalue of first contrast = 1.60
Step 4. Person-response validity	For person goodness of fit statistics: g) $\leq 5\%$ of participants with Infit MnSq values > 1.4 logits and $z > 2.0$	g) 8.98% of participants with Infit MnSq > 1.4 and $z > 2.0$	g) 6.59% of participants with Infit MnSq > 1.4 and $z > 2.0$
Step 5. Person-separation reliability	h) Person separation index (PSI) > 1.5 i) Person reliability index (PRI) ≥ 0.80 j) Cronbach alpha (KR20) for person item measure ≥ 0.80	h) PSI = 2.09 i) PRI = 0.81 j) KR20 = 0.87	h) PSI = 1.99 i) PRI = 0.80 j) KR20 = 0.88
Step 6. Differential item functioning	k) No significant differential item functioning (DIF) using Mantel-Haenszel chi-square statistic between younger (< 35 years) and older women ($p < 0.01$) l) No DIF for mothers of older and younger infants (< 6 months)	k) DIF for young and old women, all $p > 0.02$ l) DIF for young and old infants, all $p > 0.07$	k) DIF for young and old women, all $p > 0.45$ l) DIF for young and old infants, all $p > 0.16$
Step 7. Targeting	m) Difference between mean person ability and mean item difficulty < 1 logit	m) Difference = 1.21 logits Person measure $M = 1.21$ ($SD = 1.05$) Item measure $M = 0.00$ ($SD = 0.71$)	m) Difference < 1 logit Person measure $M = 0.83$ ($SD = 2.04$) Item measure $M = 0.00$ ($SD = 0.60$)

DIF, differential item functioning; FSS, Fatigue Severity Scale; FSS-5R, Fatigue Severity Scale–Recoded 5-item version; MnSq, mean square; PCA, Principal Components Analysis; PSI, Person Separation Index; PRI, Person Reliability Index

^a Criteria for Rasch analysis adapted from Lerdal et al. (2016); Mills et al. (2009); Ottonello et al. 2016

($M = 73.78$, $SD = 16.45$, $n = 129$), DASS21-D ($M = 5.12$, $SD = 3.81$, $n = 162$), and EPDS ($M = 10.87$, $SD = 4.66$, $n = 141$); 50% of mothers reported at least mild depressive symptoms (DASS21-D ≥ 5).

The results of the RA are presented in Tables 1 and 2. Overall, the FSS items were not well targeted and scale items did not assess the latent variable of fatigue at either extreme (see Table 2). Mean Rasch scored person fatigue “ability” was substantially higher (> 1 logit) than the mean Rasch scored item “difficulty” of the FSS, showing an overall ceiling effect for the scale (see Table 1: step 7). Item response categories did not advance monotonically for items 3, 6, 7, and 9 (see Table 1), and there was a poor fit for item 2 and 7 (Infit > 1.3 MeanSquare [MnSq]; see Table 2). There were more participants (9%) with poor fit based on person goodness-of-fit statistics than expected by chance ($\leq 5\%$). Reliability was adequate, and there was no differential item functioning for infant or maternal age.

To improve the scale, the following procedures were carried out based on past FSS validations (Mills et al. 2009; Ottonello et al. 2016). The response options of the FSS were first revised from 1234567 to 1112345, and then items with ongoing lack of monotonic advancement (i.e., the increase of fatigue levels did not correspond to the increase in response options; item 9) or inadequate fit (Infit MnSq > 1.3 ; items 1, 2, 3) were removed in a step-wise fashion. The revised scale (FSS-5R) retained items 4 to 8 of the original FSS and performed better against the RA criteria (see Table 1), with adequate validity and reliability ($\alpha = 0.88$), although still assessed only a modest range of fatigue difficulty (see Table 2). The correlation between the FSS-5R scores and the FVAS was 0.43 ($p < 0.001$). Between the FSS-5R and DASS21-D, EPDS

and social support was 0.34, 0.29, and 0.19 respectively (all $p < 0.05$). For comparison, the correlation between the DASS21-D and EPDS was 0.69 ($p < 0.001$).

Discussion

This is the first study to evaluate the psychometric properties of the FSS during the postpartum period. The original FSS had several limitations among women with elevated fatigue seeking help for UIB, including an overall ceiling effect and poor fit for several items. It did demonstrate adequate reliability and a lack of item bias for different maternal and infant age categories. By revising response options and removing poorly performing items, the revised 5-item FSS (FSS-5R) had improved fit, reduced ceiling effect, adequate validity, and reliability.

Similar problems with the FSS have been found in populations with chronic illnesses (e.g., multiple sclerosis), where improvements in psychometric properties occurred following revision of response options and removal of items with poor fit (Mills et al. 2009; Ottonello et al. 2016). The modified FSS-5R retains items 4 to 8 of the original FSS and measures the interference of fatigue on specific domains of functioning. The removed items related to motivation (item 1), exercise (item 2), ease of fatigability (item 3), and interference of fatigue with work, family, and social life (item 9). Women in our study may have had difficulty simultaneously evaluating fatigue interference on these three different domains specified in item 9, which may be affected to different extents by their fatigue.

Table 2 Rasch item fit statistics for FSS and FSS-5R

FSS (9 item)					FSS-5R				
Item	Item Difficulty (logits) ^a	Model SE	Infit MnSq	Infit z	Item	Item Difficulty (logits)	Model SE	Infit MnSq	Infit z
FSS-1	-1.00	0.10	1.24	1.70	-	-	-	-	-
FSS-2	1.65	0.07	2.08	7.80	-	-	-	-	-
FSS-3	0.08	0.08	1.13	1.10	-	-	-	-	-
FSS-4	-0.38	0.08	0.82	-1.60	FSS5R-4	-0.61	0.11	1.06	0.05
FSS-5	0.29	0.07	0.72	-2.80	FSS5R-5	0.59	0.10	0.95	-0.40
FSS-6	0.34	0.07	0.71	-2.90	FSS5R-6	0.68	0.10	0.82	-1.70
FSS-7	0.03	0.08	0.69	-3.00	FSS5R-7	-0.12	0.11	0.97	-0.20
FSS-8	-0.45	0.08	0.96	-0.30	FSS5R-8	-0.78	0.11	1.26	2.00
FSS-9	-0.57	0.09	1.04	0.40	-	-	-	-	-

FSS, Fatigue Severity Scale; FSS-5R Fatigue Severity Scale–Recoded 5-item version; MnSq, mean square. FSS-5R items are the same as FSS with response options recoded 1234567 to 1112345

^a Item difficulty in the Rasch model ranges from -3 to 3 logits; for the FSS, the lack of items from -1 to -3 and from 1.65 to 3 logits indicates only a narrow range of item difficulty is being assessed

The revised FSS-5R demonstrated discriminant validity, by sharing only a moderate and small positive association with measures of depressive symptoms and a small association with social support needs. These correlations with the FSS-5R were weaker than the strong correlation observed between the two depressive symptom scales and the moderate correlation between the FSS-5R and FVAS. However, evidence for convergent validity of the FSS-5R and FVAS was modest. The lack of a stronger association may be due to the scales targeting different focuses (interference vs. symptoms), different assessment periods and item lengths.

The recruitment rate and sample size were modest but sufficient for item and person measures to be adequately stable. The study sample was predominantly married/partnered, well educated, and Australian born, which may limit generalizability to populations with different demographics. Further research is also required to assess FSS-5R for its fit among other postpartum populations (e.g., women not seeking support for UIB or fatigue, partners who are also vulnerable to fatigue), the longitudinal stability of its item hierarchy, and its sensitivity to treatment. Strengths of this study were use of a rigorous Rasch approach and recruitment of a help-seeking group of women, both firsts in a validation study of a postpartum fatigue scale. Findings from our sample are also likely to generalize to many other women in clinical settings seeking help with fatigue or other forms of psychological distress, given that as many as one in four families report difficulties with infant crying, fussing, and sleep difficulties (Fisher et al. 2011).

Overall, findings suggest that the original FSS has numerous psychometric problems in women with elevated postpartum fatigue symptoms. Clinicians and researchers measuring fatigue in postpartum populations should be cautious about using the FSS in its original form and are advised to consider using the briefer FSS-5R version with more robust psychometric properties that can be easily rescored from the original FSS.

Acknowledgements The authors are grateful to the staff at Masada Private Hospital Mother Baby Unit for their collaboration in implementing this study: Olivia Chung, Hilary Brown, and Hannah Gray for assisting with data collection and to the women who most generously contributed data. NW was supported by an Australian Government Research Training Program Scholarship, KW was supported by a Monash University Advancing Women's Research Success grant, and JF was supported by a Monash Professorial Fellowship and the Jean Hailes Professorial Fellowship, which is supported by a grant to the Jean Hailes Foundation from the H and L Hecht Trust managed by Perpetual Trustees.

Compliance with ethical standards

Conflict of interest Mr. Wilson has nothing to disclose. Dr. Wynter has nothing to disclose. Dr. Fisher reports personal fees from Masada Private

Hospital, which is owned by Ramsay Healthcare, other from Jean Hailes for Women's Health, during the conduct of the study; Dr. Bei has nothing to disclose.

Ethical approval All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

Informed consent Informed consent was obtained from all individual participants included in the study.

References

- Cox JL, Holden JM, Sagovsky R (1987) Detection of postnatal depression. Development of the 10-item Edinburgh Postnatal Depression Scale. *Br J Psychiatry* 150:782–786. <https://doi.org/10.1192/bjp.150.6.782>
- Fisher JR, Feekery CJ, Rowe-Murray HJ (2002) Nature, severity and correlates of psychological distress in women admitted to a private mother-baby unit. *J Paediatr Child Health* 38:140–145. <https://doi.org/10.1046/j.1440-1754.2002.00723.x>
- Fisher JR, Rowe H, Hiscock H, et al (2011) Understanding and responding to unsettled infant behaviour. Australian Research Alliance for Children and Youth. https://www.aracy.org.au/publications-resources/command/download_file/id/158/filename/Understanding_and_responding_to_unsettled_infant_behaviour.pdf. Accessed 7 July 2017
- Giallo R, Wade C, Kienhuis M (2014) Fatigue in mothers of infants and young children: factor structure of the Fatigue Assessment Scale. *Fatigue Biomed Health Behav* 2:119–131. <https://doi.org/10.1080/21641846.2014.925326>
- Insana SP, Montgomery-Downs HE (2010) Maternal postpartum sleepiness and fatigue: associations with objectively measured sleep variables. *J Psychosom Res* 69:467–473. <https://doi.org/10.1016/j.jpsychores.2010.07.004>
- Krupp LB, LaRocca NG, Muir-Nash J, Steinberg AD (1989) The Fatigue Severity Scale. Application to patients with multiple sclerosis and systemic lupus erythematosus. *Arch Neurol* 46:1121–1123
- Lerdal A, Kottorp A, Gay C, Aouizerat BE, Lee KA, Miaskowski C (2016) A Rasch analysis of assessments of morning and evening fatigue in oncology patients using the Lee Fatigue Scale. *J Pain Symptom Manag* 51:1002–1012. <https://doi.org/10.1016/j.jpainsymman.2015.12.331>
- Linacre JM (2016) Winsteps® (Version 3.92.1) [Computer Software]. <http://www.winsteps.com/>. Accessed 7 Jul 2017
- Lovibond PF, Lovibond SH (1995) Manual for the depression, anxiety and stress scales (DASS). Psychology Foundation, Sydney
- Mills R, Young C, Nicholas R, Pallant JF, Tennant A (2009) Rasch analysis of the Fatigue Severity Scale in multiple sclerosis. *Mult Scler* 15:81–87. <https://doi.org/10.1177/1352458508096215>
- Ottone M, Pellicciari L, Giordano A, Foti C (2016) Rasch analysis of the Fatigue Severity Scale in Italian subjects with multiple sclerosis. *J Rehabil Med* 48:597–603. <https://doi.org/10.2340/16501977-2116>
- Whitehead L (2009) The measurement of fatigue in chronic illness: a systematic review of unidimensional and multidimensional fatigue measures. *J Pain Symptom Manag* 37:107–128. <https://doi.org/10.1016/j.jpainsymman.2007.08.019>