**ASSESSMENT REVISION**

**EDITOR’S FEEDBACK**

1. You need to provide a clear rational on the importance of the construct and not only why a new measure is needed above the existing ones.
   * I have begun to do this, but the section needs a bit more work. I have asked Kathrin if she can finish it up.
2. Please explain in more detail the implications on the representativeness of your sample (e.g. is 50% prior pregnancy loss representative?).
   * We apologize. Our rate of miscarriage was indeed high. This was an error on our part in that women who were experiencing their first pregnancy were not included in the denominator of the equation. This has now been corrected in Table 1, as has the rate of stillbirth in our sample (30.3% and 2.4% respectively).
3. SPSS does not provide fit statistics for exploratory factor analysis. In order to convince me of the structure of your measure, you have to provide fit statistics. For example, M plus provides fit statistics for such analyses (also see reviewer 3; and M plus will provide geomin rotation), besides chi-squared, RMSEA (and confidence interval), also report at least CFI and TLI.

* The table has been produced. See attached report.

1. This reviewer is also absolutely correct that using the standard scree plot is an outdated approach to select the number of factors. So please follow his advice and apply a more advanced method such as parallel analysis or the MAP test or if you use a scree plot, do this on basis of fit statistics like for example the Hull method described by Lorenzo-Seva, Timmerman and Kiers (2011).
   * We have now applied both Parallel Analysis and the MAP test to obtain a suggested number of factors. Parallel analysis resulted in…

See attached report. It appears that the data can support up to 9 factors

1. Also demonstrate the unidimensionality of the Interference scale.
   * We have run an unweighted least squares factor analysis of the 7-item interference scale. The results indicate a single Eigen-value greater than one, and a very clear, one-factor solution based on the Scree plot.
2. An important question is the stability of your factor structure. You can perfectly demonstrate factorial invariance in your sample by calculating congruency coefficients for the factors in important subgroups: 1) country: congruency factors Canada, US, GB 2) delivery method: congruency factors vaginal vs. cesarean. Demonstrating invariance over subgroups is a prerequisite for acceptation of the manuscript.
   * We have now calculated congruency coefficients for both Canada (n = 409) and the United States (n = 198). The sample size of each other the other nation groups (UK, Australia/New Zealand, and other) was less than 20, and therefore too small to produce a meaningful factor structure.
   * The congruency coefficients are as follows:
     1. Full sample and Canada: 0.84.
     2. Full sample and the United States: 0.99.
     3. Canada and the United States: 0.87.
   * As we did not have mode of delivery data for this sample (this is a prenatal sample only), we instead calculated congruency coefficients for nulliparous and multiparous women.
   * Specifically, the congruency coefficients are as follows:
     1. Full sample and nulliparous women: 1.00 (0.998).
     2. Full sample and multiparous women: 0.86.
     3. Nulliparous and multiparous women: 0.83.
3. Table 3: add means and SD over countries. Explain the \* in the table (suppose it are p levels) and add the effect sizes (Cohen d) for important subgroups (to compare over countries and over delivery method).
   * I think what is being asked for here is that I also add in major country groups to T3, and provide the M & SD for each of the CFQ subscales across country groups. I will do this once we have finalized the measure.
4. Subtitles are not needed in the abstract.
   * Thank you. This has been corrected.
5. Provide a concrete example of excluded content domains in your abstract.
   * This has now been done.
6. End the abstract with a concrete example of why the measure can be useful (instead of hoping it will be useful).
   * This has now been done.

**REVIEWER 1**

***Introduction***

1. Given that the journal Assessment has a wide audience including those who may not have specific interest/expertise in childbirth, authors may want to consider providing a brief summary of the construct of childbirth fear and why a sound assessment tool in this area would be important in advancing women’s perinatal health.
   * Working on this, but not quite done. Kathrin is helping.

***Methods and Results***

1. Given the increased interest in perinatal mood and anxiety disorders, please provide information regarding the EPDS and MQ scores in Table 1 or Participants section of the Results.
   * The means and standard deviations for the MQ have now been added to Table 1.
2. How many women scored positively on the EPDS? This information is important in further understanding the generalizability of the study’s findings.
   * 138 (21.3%) participants scored a 12 or greater on the EPDS. This has now been reported in the participant section of the manuscript.
3. Authors may also want to consider providing information on whether and how the number of weeks in pregnancy, pregnancy complications, and prior pregnancy losses affected the CFQ scores. These factors can affect perinatal woman’s perception and fear of childbirth.
   * Thank you very much for the relevant feedback. We currently have a subsequent journal submission under review in which we assess the relationship between the CFQ subscale and full scale scores and demographic, obstetric and reproductive variable. As such, we do not think it would be helpful to include this information here in the psychometric development paper.
4. How did the CFQ correlate with the fear domain/factor of the W-DEQ-A?
   * I’ve asked Kathrin to help me sort out which items to use.
5. Did women who show positive vs. negative EPDS scores show differences in the total CFQ score and/or the pattern of subscale scores?
   * Run this when the measure is finalized.
6. Was EPDS more strongly correlated with certain subscales of the CFQ than others?
7. Did the interference scale scores distinguish those who score high vs. low on EPDS?
8. Similar questions apply to the MQ.
9. I understand that the CFQ assesses a construct that is different from perinatal depression or blood injury fears. However, given the authors’ emphasis on the role of the CFQ (and particularly its interference scale) on predicting mental health outcomes, these associations would be important to consider and report.

**REVIEWER: 2**

1. In the Results section (p. 10), the authors report deleting items from the scale based on a “combination of the number of items per factor, the conceptual coherence of each factor, and the factor loadings.” It would be helpful if the methods for removing items were made more explicit with regard to these methods, especially factor loadings.

* Thank you. That is helpful feedback. Items, in this iteration, were removed as follows:
  + Any item loading below 0.40 across all five factors was removed (items 5, 16, 18, 45, 46, and 49).
  + Any item loading 0.30 or greater on one or more factor was also removed (items 31 and 32).
  + There were two exceptions to the above:
    1. Two items (#1 and #28) were retained for conceptual reasons. Specifically, although item #28 loaded below 0.40 (i.e., 0.387), it was retained because we judged that having two items pertaining to fear of harm to the mother was important to the Fear of Harm to Mother or Infant subscale.

1. It would be helpful to know if there were any differences in patterns of response between women from each of the different countries. If so, that additional variable should be included in the model for analyses.
2. The authors reported prorating the W-DEQ-A from a 5-point scale to a 6-point scale. I presume that simply means that the total scores were prorated, although it is not clear. It would be helpful to provide more explicit details about how the items were prorated.

* Thank you. Yes, this is correct. Specifically, we took original W-DEQ-A scores, divided them by four and then multiplied them by 5 to shift scores from a 0-4 scale to a 0-5 scale (i.e., W-DEQ-A-Corrected = W-DEQ-A-Uncorrected / 4 \* 5). This has now been explained in the manuscript (para 1, page 10).

1. The methods used to compare correlations between the CFQ and other measures were calculated using a Lee & Preacher; however, the citation in the reference list is incomplete. I presume the authors used the calculator provided at the following website:

Lee, I. A., & Preacher, K. J. (2013, September). Calculation for the test of the difference between two dependent correlations with one variable in common [Computer software]. Available from <http://quantpsy.org>

1. I was able to locate that calculator online and used it to try to replicate the results presented by the authors. My results were slightly different than those reported in the results, which may be that the n for the r-to-z transformation was different than 643. Please double check the values obtained when comparing these correlations, and also report the number of participants for whom the correlations were calculated.

* We conducted PA with various software, the results of all is supplied in the appendix. N = 643

1. Was there an association between number of weeks pregnant and CFQ results? If so, that variable should be used or controlled in the models.
2. On page 12, line 13, there is a t-test with df listed as 33.98. That appears to be a typo.
3. In the sample, of those with prior pregnancy (296), 50% had prior pregnancy loss before 20 weeks. That figure seems very high. Please comment as to whether that rate of prior pregnancy loss is atypically high, and whether it may make the sample less representative.

* This was an error on our part in that women who were experiencing their first pregnancy were not included in the denominator of the equation. This has now been corrected in Table 1, as has the rate of stillbirth in our sample (30.3% and 2.4% respectively).

1. Please clarify if the 7-item Interference scale is in addition to 38 items of the CFQ, thus resulting in a measure that is 45 items in length.

* Yes, this is correct.

**REVIEWER: 3**

***Major points:***

1. The introduction provides a nice review of existing fear of childbirth measures and their limitations. However, I was left wondering why it is important to assess fear of childbirth. What are the negative consequences of fearing childbirth? A rational should be provided as to why this is an important construct to assess.

* This has now been added to the introduction. We agree that this was an omission from our original submission.

1. Relatedly, although existing measures do not assess childbirth-related fears as broadly as the CFQ, I was left wondering why the additional scales assessed by the CFQ might be important. What important outcomes might these previously unassessed content domains be related to, or predict?

* While we agree that dimensions assessed by the W-DEQ may be relevant to women’s experience of childbirth, they do not assess fear. What concerned us, when developing the CFQ, was that existing measures of fear of childbirth fail to give us a clear picture of fear of childbirth and its correlates. Clinicians who identify women reporting high levels of childbirth fear may be identifying a different group of women from those detected via existing measures such as the W-DEQ-Q. Consequently, relationship found to exist between the W-DEQ-A and outcomes may not consistently apply to women who report high levels of childbirth fear. For example, there is some evidence suggesting that fear of childbirth may increase risk of caesarean birth. It could, however, be the case the other constructs (e.g., joy, hopelessness) assessed by existing measures (e.g., the W-DEQ-A) are driving this relationship. Consequently, advice and education given to women regarding their childbirth related fears may be misleading.

1. Using scree plots to determine the number of factors to retain in an EFA is an outdated approach. Parallel analysis is now considered the gold standard for determining how many factors to retain in EFA models (Fabrigar et al. 1999; Velicer et al. 2000; Hayton et al. 2004). Parallel analysis should be used to confirm the six-factor structure. Relatedly, the authors should report the eigenvalues for at least the first 6 factors

Eigen values have been reported. See html report.

1. The authors need to present goodness of fit statistics for their final solution – at the very least chi-squared, RMSEA, and the RMSEA test of close-fit.

* Table has been produced with these statistics. See html report.

1. Parallel analysis should also be used to confirm the unidimensional structure of the interference scale. Again, eigenvalues should also be presented

* I don’t know how to do it, sorry.

***Other points:***

1. The authors identify seven domains of interest to assess using the CFQ? Were any additional domains considered, but ultimately omitted?
   * The only fear domain which we considered but chose not to include is that of fears related to maternity care providers. Specifically, women may be fearful that they will feel unsupported by or not listened to by their care provider during labour/delivery, that their own care provider will be unavailable, or that their care provider may not make choices that the woman would want, or finally that the care provider may be incompetent in some way. We decided not to include items pertaining to this domain for two reasons: (1) this arena of fear is less central to the birth experience that those we included (i.e., is more similar to perceptions of partner support during labour/birth) and (2) there exist a number of self-report tools which assess this directly.
2. Do the authors have citations to support the argument that at least three items are needed to ensure a reasonable degree of internal consistency? –

This is a rule of thumb, a recommendation, not a prescription.

1. The authors should provide the within-sample alphas for all the questionnaires administered to test convergent/discriminant validity

* I know what they are talking about, but can’t do it without some recollection and research.

1. Why did the authors choose direct oblimin rotation as opposed to any of the other widely available oblique methods (promax or geomin are more common)? A rationale should be provided

* We explored multiple rotaions and confirmed the structure across all of them. I have never seen a choice of rotation being articulated sufficiently clear. In my opinion, no single rotation can provide a clear pictures of the underlying structure.

1. The authors should present all factors loadings in Table 2 (including loadings below 0.20)
2. What about a need for future studies to replicate and confirm the purported factor structure using confirmatory factor analysis?

* Yes. We cannot use the same data we used to extract the structure to confirm the same structure. That’s why additional data should be collected and tested against the factor pattern that was produced in the current study.