Reviewer’s report

Title: Revisiting the dimensional validity of the Edinburgh Postnatal Depression Scale (EPDS): an empirical evidence for a possible higher-order factor

Version: 1 Date: 10 February 2011

Reviewer: Adam Carle

Reviewer’s report:

In this paper, the authors used confirmatory factor analyses (CFA) and CFA within an exploratory approach to evaluate the internal measurement structure of the Edinburgh Postnatal Depression Scale (EPDS). Currently, clinicians and researchers use a single summary score to evaluate postnatal depression. However, if the EPDS does not have a unidimensional structure, one cannot validly create a simple summary score to evaluate postnatal depression.

The authors have identified an important problem (one that exists throughout far too much of the medical literature): failure of empirical research to establish a valid scoring system for self- or proxy-report scales. This is especially important when evaluating constructs like postpartum depression, given its serious effects on mothers, their children, and families.

Major Compulsory Revisions (except where noted otherwise):

1. While I agree that the discussion does not warrant a detailed discussion of the differences across studies, I think some comment on potential reasons for differences would strengthen the need for the current study. For example, I have not read all of the manuscripts the authors cite. However, in my experience, many “factorial” studies fail to correctly model the ordered-categorical nature of most psychometric data. If this is true for some of the studies the authors cite, the authors might offer this as a reason for the differences in the factor structures found across studies. Additionally, I’ve also found that too many papers state they used confirmatory factor analyses, when in fact they used something like principal components analysis, not at all the same thing. If this is true, the authors might also offer this as a reason.

2. Also in the introduction, given that too many in the medical field have insufficient training in advanced psychometric, I think it would strengthen the manuscript if the authors more fully explained why a multidimensional solution would make it invalid to use a single summary score.

METHODS:

3. In posts to the Mplus website, the Muthén’s have indicated that the WRMR has not performed as Muthén and his colleagues hoped across diverse modeling settings. Although Mplus still reports it, it is not a particularly useful statistic (e.g., sensitive to sample size, etc.). I would suggest the authors exclude this measure of fit.
4. I am unclear on whether the authors first generated polychoric matrices outside of Mplus and used these in their analyses or used Mplus to analyze the raw data. If they analyzed polychoric matrices generated outside of Mplus, why? This leaves important psychometric information aside (e.g., thresholds). Either way, could the authors clarify this?

5. When they discuss acceptable fit index values, the authors should also cite:

   (As an aside, very little literature has evaluated appropriate values for models based upon ordered-categorical data.)

6. When discussing the factorial composite reliability and the average variance extracted, the authors don’t acknowledge that the equations and citation they give were developed for continuous factor models. Thus, this equation describes the reliability of the latent response variates, not the observed responses. The latent response variate has a non-linear relationship to the observed responses, complicating a straightforward correspondence. While the authors make relatively valid points regarding these indices, their use with order-categorical measures isn't clear if the authors wish to describe the observed responses (which presumably they do) relative to the measurement model. I’m not confident the authors can use these indices as strongly as they do to evaluate the model. Could the authors either note the problems with these indices with respect to the current data, better defend their use, or use an alternative approach?

7. A higher order factor model with only three first-order indicators isn’t identified. Thus, it’s not possible to evaluate the fit of this model. I would recommend the authors only describe this as an exploratory analysis in which they cannot provide evidence one way or the other against (if they choose to pursue the higher-order approach after responding to point 8).

8. Why didn’t the authors examine a bifactor model approach to the structure of the data? In several publications Steve Reise has cogently and persuasively advocated bifactor models (e.g., Reise S, Moore T, Haviland M. Bifactor Models and Rotations: Exploring the Extent to Which Multidimensional Data Yield Univocal Scale Scores. Journal of Personality Assessment. 2010;92(6):544-559. Reise SP, Morizot J, Hays RD. The role of the bifactor model in resolving dimensionality issues in health outcomes measures. Quality of life research : an international journal of quality of life aspects of treatment, care and rehabilitation. 2007;16 Suppl 1:19-31.). This approach has many advantages over a higher-order approach, advantages especially relevant to the current context (how should one score the scale?).
RESULTS
9. I leave aside some comments here related to the resolution of my questions/comments above.

10. The authors include no discussion of the thresholds in their results. Why not?

DISCUSSION:
11. (Discretionary Revision) In their first sentence, I would remove “may”. The study clearly adds to our knowledge regarding this instrument.

12. I would argue that the authors have presented no evidence for a higher-order structure because they cannot empirically evaluate its fit. However, they could have empirically evaluated the fit of various bifactor models. Moreover, these bifactor models (or lack of their fit), would shed light on the substantive meaning of the complex structure. Do responses to the scale generally measure one construct, with some grouping or nuisance factors; or do responses to the scale measure three distinct constructs, which in turn may measure a higher-order construct?

13. The authors presented no analyses comparing a simple sum of the ten items to an approach based on the complex factor structure (e.g., factor scores based on the [somewhat indefensible] higher-order factor; a sum of each of the three first-order factors’ score, etc.). Without these analyses, or some other method of actually evaluating scores vs. structure, the authors should not comment on the appropriateness of scoring given the results of the analyses and the complex structure. Or, the authors need a much more rigorous defense of their suggested scoring system, a defense that clearly uses the parameters of the factor model.

14. The authors do discuss (to some degree) the indefensible nature of the higher-order model. While they provide very general citations for their use, they have not provided the exact pages within the books they cite that provide support for their approach. Moreover, several studies indicate that the approach is indefensible. Just because people have taken this approach in the literature (as they authors note) does not make it defensible methodologically and empirically.

15. The authors explicitly indicate (page 13) that the three factors may not quality as independent dimensions. If they believe this, I am even more surprised they chose not to examine bifactor models. A suitable bifactor solution would provide a more defensible empirical argument for their conclusion.

In sum, I think this study has the potential to provide important information about the psychometric properties of the EPDS. However, several aspects need addressing before the authors’ conclusions are fully supported.

Level of interest: An article of importance in its field

Quality of written English: Acceptable
**Statistical review:** Yes, and I have assessed the statistics in my report.

**Declaration of competing interests:**

I declare that I have no competing interests