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: 12 February 2011

Reviewer: rochelle tractenberg

Reviewer's report:

Manuscript Review, BMC Medical Research Methodology

"Revisiting the dimensional validity of the Edinburgh Postnatal Depression Scale (EPDS): an empirical evidence for a possible higher-order factor" by Michael E Reichenheim, Claudia L Moraes, Alessandra S Dias and Gustavo Lobato.

This manuscript describes an examination of the dimensionality of a 1-item assessment scale that was originally intended to be summarized with a single score but has subsequently been found to comprise possible/putative subscales. The examination involved cross sectional responses from over 800 individuals to whom the scale’s administration was appropriate (i.e., they are from the target population for whom the scale was devised), which supports the validity of estimates and modeling based on responses from this sample.

Following are Minor Essential Revisions, except where specified as “major compulsory”

1. Title: the title is confusing and tends to contradict the abstract.
   A. This reviewer thinks “dimensionality”, or “dimensional structure” (from the abstract) are preferable to “dimensional validity”, which is more used in a mathematical than psychological context.
   B. The authors do report a higher order factor, and although they do not replicate their analyses in an independent sample, the discovery of a higher order factor is reported as “results” in the abstract. Therefore, the use of “possible” in the title does not make sense.
   C. The first word after the colon (“an”) should be removed.

2. Abstract. The respective fits of the 1F, 3F, and 3F/1higher order F models should be presented (i.e., respective CFI/TLI).
   A. In the Results section, the authors give a “range” of values, but do not state what these values represent. Please state what the values represent and what the “range” came from, e.g., is it a range of results from several models? Is it the 95% confidence limits for some unspecified estimate? (Major Compulsory)
   B. In the Results section, the word “again” in the statement that “…an E/CFA again uncovered a three dimensional solution” should be removed or better, since the “uncovered” result is from the exploration (E of E/CFA), “again” should be changed to “initially”.
C. In the Results section, “A new CFA” could be clarified because it wasn’t a simple reset of the software, it was a new plausible model to be fit, namely, a higher order model with 1 overall and 3 first order factors, e.g., “A higher order CFA was fit.”

D. The Conclusions could be reworked, in the sense that this is a single study, albeit based on a large sample. The authors appear to have uncovered evidence explaining both the utility of this scale in its original/intended one-dimensional representation *and* the reason why some studies have found evidence of more than one dimension. This is a crucial tie of these results to those of others. So “Although the EPDS contains three first-order factors, they may not qualify as independent dimensions” should become, “Our evidence suggests that the EPDS should be scored as if it has one dimension. Its items include the three key symptom types in depression as first order factors, but these are in turn strongly associated under a single higher order factor. The total score represents the overall depression symptomatology more validly and defensibly than subscores can.” Or something to that effect.

E. The evidence in this study is very strong against the idea that the first order factors are at all independent- in fact, with N=800, the 95% CIs for the correlation estimates among these three first order factors will be quite small and very far from zero. Therefore, these authors can conclude that the first order factors do not (not “may not”) qualify as independent.

3. The data analysis section (which this reviewer was specifically invited to evaluate)

A. This section is written very clearly and specific details are provided. The information about fit and its evaluation is particularly well written. Oblique rotation is crucial in this application and was clearly described.

B. In the fourth full paragraph under “data analysis”, a reference for the use and definition of “factorial composite reliability” should be included.

C. In the pentultimate full paragraph under “data analysis”, first sentence, the phrase “…model was stepped up” does not make sense. Also, in CFA, one either does or does not specify the presence of a higher order factor- one cannot “tentatively specify” a higher order factor. It is specified and its robustness (fit to the data) is tested. If the higher order factor was actually included in this CFA, strike the word “tentatively”.

D. A technical note on the estimation of relationships outlined in the pentultimate full paragraph under “data analysis”: these two path values are only multiplied together to obtain the estimate if the variance of the higher order factor was set to 1, meaning that none of the first order factors is defining the scale of the higher order factor. This should be clarified (namely, that none of the first order factors defined the scale of the higher factor) here and throughout, unless in the CFAs the authors did use one of the items to anchor/give scale to the factor. Simply stating “Factor variances were set =1 so as to permit estimation of all loadings” (or something like that) sorts this right out.
Following are all in the category of Major Compulsory revisions, except where identified.

4. Results.

A. Please provide the internal reliability coefficient for the one-factor solution (i.e., the 1F model fits poorly, how does the internal consistency look? (Minor Essential)

B. The use of MI poses a challenge to the interpretability and generalizability of these results. Namely, 1) it is not possible to discern just how many models were fit to this data “several” could be anywhere from 3-10 models.

C. The literature review (and 1F CFA results) do not appear to support any 2-F models. The fact that some (unspecified number) 2F models were fit lend these results the flavour of a fishing expedition. While all the methods are well-described and results presented clearly (apart from a few comments as noted), this reviewer concludes that many, many models were tested and this testing was driven by the data (i.e., MI values), and not by theory. This is a *significant limitation* of this study and must be clearly stated as a limitation in the discussion. These results cannot be depended upon until they have been independently validated, due to the data-driven nature of the results. A table that includes small circle-and-square figures representing every model that was tested, together with its fit index values, would go a long way towards “full disclosure”, by both fully specifying all models that were fit and by showing how completely this model space was explored in these analyses. The figures don’t have to be labeled, because the investigators did not go beyond examining fit (i.e., to the substantive interpretation of the factors stage) for most of the models that WERE fit. The legend would simply remind readers that circles are factors (latent and causal) and squares are items, and that (eg), one square represents 2 items. That will make the figures small with immediate interpretability (eg, 2F, one with 8 items, one with 2) and will convey how little beyond the fit of the solution was gleaned by that particular MI-driven model fitting.

D. in a related point, the final full paragraph of the Results section, the authors refer to “the interim E/CFA”—implying that only one exploratory, followed by one confirmatory, analysis was done and described in the paragraphs preceding it. This might reinforce the impression, in a casual reader, that only theory-based models were explored/fit in this study. This text must be modified to specify how many interim models were fit before the higher order factor model was tried.

E. As noted under point 3.C, in CFA, one either does or does not represent the presence of a higher order factor- one cannot “tentatively represent” a higher order factor. It was specified in this CFA and its robustness (fit to the data) was tested. Strike the word “tentatively”. Also, instead of “again showed adequate fit” in the sentence that begins, “Parametrically identical to the previous model…”, change “again” to “naturally” or “as expected”. Not all readers will appreciate that, if 2 models are parametrically identical then their fit statistics will also be. A reference to equivalent models and their testing and interpretation.

F. The authors should read, and make sure their results are/report is consistent
with the criteria shown in, Table 8.1B, in The Reviewer's Guide to Quantitative Methods in the Social Sciences Edited by Gregory R. Hancock and Ralph O. Mueller. This reviewer owns that book but the table was visible via Amazon.com, http://www.amazon.com/Reviewers-Quantitative-Methods-Social-Sciences/dp/041596508X/ref=sr_1_1?s=books&ie=UTF8&qid=1297531686&sr=1-1#reader_041596508X.

5. Table 1. It might be helpful to include the ranges of “goodness” for WRMR, RMSEA, CFI & TLI, and to specify that the 95% CIs are all bootstrapped, all within the table legend. Also, as is done in Table 2, adding the subheader “Factor correlations” to the phi values would be helpful to the less-technical reader. (Minor Essential)

6. Table 2. Legend states that the values are from “the previous E/CFA model”, which does not make sense, even when the reader is referred to the text. Is this the result of the exploratory model? The “previous” model might be completely misconstrued by a naïve reader to be either the EFA or the CFA. Please specify which model this is. It might be useful for the authors to include figures showing (circle/squares, with arrows) the 1F CFA, the EFA result, the CFA based on EFA, and the higher order CFA models. This could clarify the thinking of readers unfamiliar with implications of “higher order” factors on lower order factors, for example, particularly when the authors conclude that one score summarizes a test that appears to have 3 or even 4 (to the uninitiated) factors.

7. Figure 1 needs more details, and particularly for those readers who refuse on principle to remember what the phi and square-root rho values actually represent, the figure conveys no information. The legend states that relationships are shown, but what do they mean? It is trivial to put actual names on these variables and put the greek representations in parentheses for the undoubtedly fewer-in-number technically-inclined readers who use LISREL notation. (Minor Essential)

8. Figure 2 shows that the anhedonia factor is locally underidentified. Please discuss in the text.

9. Figure 2 shows no paths from the higher order factor to any item, but Table 3 shows the factor loadings of all items on the higher order factor. Table 3 suggests that there are both direct (estimates shown in Table 3) and indirect (estimates shown in other tables) effects of the higher order factor on all items. That is, arrows both directly from the higher order factor to all items plus indirectly, through the first order factors, to all items. Which representation is correct? Table 3 and Figure 2 must be reconciled. Did you estimate, eg, i1<- *F1 + *F4 where i1 is item 1 and F1 is the anhedonia factor and F4 is the higher order factor? Or did you only estimate i1<-F1, and then later, F1<-F4? Please clarify—in the text, in Table 3, and in Figure 2.

Level of interest: An article whose findings are important to those with closely related research interests

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.