

Reviewer #2: This study is an evaluation of the factor structure of the BDI-II using exploratory and confirmatory factor analyses with data from low-income women. The manuscript is very well-written and easy to follow. The authors provide substantial background regarding previous research in this area, justification for sample size, evaluation of missing data and imputation, rationale for type of rotation used in the analyses, and factor selection approach. The authors do an exemplary job describing the nuances of factor analysis. Although I have one major question regarding the analysis, my critique has more to do with context and meaning than methods and analysis.

My primary concern, that is likely addressable by the authors, is the lack of information regarding the timing of the data used from a randomized trial. If the authors used only baseline data from the RCT, the fix in the manuscript should be easy. However, if the authors used data from participants over time where half of the sample may have been involved in an intervention of some sort during that timeframe (data are from Feb, 2006-Sept, 2008), the authors must control for condition in the analyses or somehow take into account the effect of the intervention on half of the sample. Depending on the type of intervention in the RCT (which is not described), responses to depression could have been altered and make interpreting the findings very difficult.

- *This is an excellent point raised by Reviewer 2 and addressing this was overlooked in the original submission. Data for these analyses were collected between February 27, 2006 and September 24, 2008, all at the baseline / study enrollment visit; thus, there is no confounding with the study condition. This was clarified on p.7.*

Other considerations for clarity of the manuscript include the following:

Title: Is the study really a validation of the inventory itself or a validation of its factor structure? Lines 10-11 on page 3 appear to be more accurate in terms of the study's purpose, and that should be reflected in the title.

- *We agree a slight modification to the title would better reflect the analysis conducted. We have changed the title to “The Beck Depression Inventory-II Factor Structure among Low-Income Women.”*

Abstract: Define EFA prior to first use. R2 should be R2. The discussion could be enhanced with a statement regarding the relevance/importance of testing the BDI-II factor structure in low-income women. Perhaps text from lines 6-8 on page 4 or text from lines 12-14 on page 15 could be used in the abstract.

- *“Exploratory Factor Analysis” has been added before the use of “EFA” in the abstract.*
- *I am not certain what the reviewer is referring to in the abstract in terms of “R2.” We did not use that term in the abstract; however, we did use the symbol for chi-square (χ^2) in the Results paragraph of the abstract. We have superscripted the “2” in this symbol to better represented the ‘square’ intent.*
- *We agree that adding text from lines 6-8 on p.4 would add to the abstract and round out the description of the problem; however, given the space limitations in the abstract instructions, it would require deleting something else, which may recreate a similar problem. Thus, we opted not to delete current portions of the abstract, and were unable to make additions to it.*

On page 8, the description of the selection of factors using eigenvalues compared against random data generation is quite interesting; however, based on Figure 1, does the eigenvalue for factor 2 really exceed the criterion? It looks very close.

- ***The reviewer is correct in this regard: the eigenvalue for Factor 2 is not above 1.0 (the eigenvalue for Factor 2 is reported as 0.78 on line 21, p.7). This was what elicited interest in doing an alternative analysis (parallel analysis), given other relevant indicators (explained on p.7 – namely, cumulative and unique percent of explained variance and prior EFA findings) suggested the BDI-II may be consistent with a two factor structure. Lines 22-23 on p.7, and lines 1-3 on p.8 further explain why factor selection based solely on the Kaiser-Guttman rule (eigenvalues only) can be faulty, and thus our rationale for considering alternative avenues (parallel analysis) to further explore factor selection.***

Lines 22-23 on page 8 are confusing. Are the authors referring to NON-zero loadings? If so, what criteria are used as many are above zero.

- ***Thank you to the reviewer for pointing this out, as this is confusing. In that section, we were describing why the factor structure could not be considered a “simple” factor structure (one that neatly falls into separate factors), given there were seven items that did not load at or ‘near’ zero. The term “non-zero” in this case would not be accurate. Gorsuch (1983) describes a simple factor structure as one in which each items loads “at or near” zero on one of the factors; however, he does not specify an absolute value to be used (e.g., 0.000 or 0.002) to define ‘near zero.’ Thus, we think the explanation added in the text on p.9 in response to this comment improves clarity.***

The authors could help readers understand factor analysis better if they commented on why the G-S-C factor structure is best, but the internal consistency reliability is so high (.94) for a one-dimensional construct.

- ***This too is a very good point, and further explanation was added to the bottom of p.9 and top of p.10. It does seem, at least initially, that a high internal consistency suggests a unidimensional construct. Internal consistency measures, however, estimate how consistently individuals respond to the items within a scale. These measures, however, are not tests of the unidimensionality of a scale. In this case, if an instrument contains a set of questions related to “X” (say, physical symptoms of depression) that correlate highly among themselves, and a second set of items – “Y” – which correlate highly among themselves (say, emotional dimensions of depression), the instrument would have a high Cronbach's alpha anyway, even though two distinct dimensions were present. Considered leading researchers in instrument development, Pedhazur & Schmelkin (1991) are very clear on this point (p.102).***

Minor comments include:

Missing or extra words on lines 2 and 21 on page 3.

- ***Corrected – thank you for pointing these out.***

Which university Health Science Center IRB approved the study? Indicating the university name would not compromise author anonymity.

- ***The official university title of the IRB was added to the manuscript.***

Typically, the order in a manuscript is text, references, all tables, all figures rather than tables and figures mixed followed by references.

- ***This is correct – the original submission was not consistent with this sequence as required by Nursing Research, and has been corrected in the revised manuscript.***

Reviewer #3: Problems Statement. Because nurses work with and study low income populations, because the BDI-II is often used with these populations, and because of the high depression rate among low income women, the establishment of the validity of the BDI-II used with low income women is of importance to nursing. The problem statement is clear and relevant to developing the body of knowledge related to the BDI-II because it fills a gap about how appropriate the BDI-II is for use with low income, female populations. The problem statement appears in the appropriate place in the manuscript, i.e., before the design and methods.

Background Literature. The background literature is appropriate for this paper and provides a context for understanding the need for a study of the validity of the BDI-II with low income women. As stated, no other research has investigated this, and therefore this study fills a gap in knowledge. The literature presented sets the context but also gives crucial background on the BDI-II, as well as current disagreements on the factor structure of the BDI-II.

Theoretical framework. No theoretical framework is used. The concept of "confirmatory factor analysis" was defined. However, the concept of "exploratory factor analysis" could have been defined in the paper to enhance readers' understanding.

- ***Upon further reflection of this point, a conceptual definition of neither EFA nor CFA was explicitly presented, and yet such a definition may assist readers in understanding the theoretical justification for conducting both of these analyses. These were added to the bottom of p.4, and should facilitate understanding related to why both are important.***

Research design and method. While a statistician should review this paper to more specifically address this, the design and methodology of this study seemed appropriate. The sample was gleaned from an ongoing RCT and the study was approved by the IRB. However, we are not informed as to how many and which scales were included in the "battery of standardized" instruments that were completed by the women along with the BDI-II. We are not informed if the women received an incentive for completing the BDI-II or the rest of the "battery of standardized" tests as part of the ongoing study (the RCT).

- ***A general description of instruments within the 'battery' was added to p.5 in the "sample" section, along with a description of the incentives women were provided for participating at each measurement occasion.***

Data analysis. A statistician should be consulted for this, however, it is noted that in Table 1, no values are given for the "mean age of children".

- ***This was an accidental omission and has been added to Table 1.***

Results. The report of the results was extensive, well written, and complete. Factor Analyses are very complex, but the authors of this paper managed to report the results effectively for reader comprehension.

Discussion. The discussion section is relevant to the results and background literature presented. Results are discussed and evaluated in relation to relevant literature. The interpretation of the results is reasonable and insightful. For example, the insight that depressive symptoms may show up in different clusters across groups is helpful and insightful. The suggestion that inquiry about how clusters of items show up in different ways for different groups holds potential to deepen our conceptual understanding of the role of depression on health outcomes. Similarities and differences between results of this study are discussed in relation to

the literature. No study limitations are mentioned, however, since these data are from a larger RCT, it is likely that some study limitations do exist.

- ***Thank you for raising this issue – a section was added to the first paragraph in the discussion section to address this.***

Organization and style of presentation. The organization and writing style of this manuscript is superb. I found no extraneous information. All was pertinent and appropriate.

Summary. The strength of this paper is in the rigor of the methods, the "audience-friendly" language used to describe complex and extensive statistical analyses, and the importance of the work for nurse researchers and others who study depression among low income women. There are no real weaknesses, just some missing parts (e.g., study limitations). Overall, this research is significant for health disparities researchers who seek to use valid measures with low income populations.