

Dear Reviewers:

Thank you for your thoughtful review of our manuscript. We take your concerns seriously and have addressed them to the best of our abilities. In areas where we do not agree with you, we have stated them with the rationale for our disagreement. We address your comments point by point in blue fonts as follows:

Reviewer Comments (there is no Reviewer #3):

Reviewer #1: This is an important study, and professionally reported. By the time I read the objectives I was excited to read more. I have a few substantive questions/comments to consider, and suggest careful editing for grammar and APA requirements for citations.

Substantive issues:

Either on Page 6, lines 7-8, or Page 7, line 5, I would like one-two sentences and more citations including perhaps Cuéllar who co-authored the ARMSA, to support using language preference as a proxy. While this is valid, it would be good to give wider support to the validity of this decision.

We have added more details and references to support the use of language as proxy for acculturation.

Page 7, starting line 4, I think a sentence or two about how the Spanish translation was developed would make the report of the study stronger. For example, were back-translation, testing for conceptual and language equivalence conducted? (as described by Phillips and colleagues in WJNR in 1996 and elsewhere).

We have added a description of the translation procedures that Bob Roberts used to translate the CESD into Spanish and gave appropriate references.

Page 8, is DIF described/used by scientists other than Zumbo? Perhaps a few other citations.

We've added several more key citations.

Writing/organization:

The first sentence of the abstract is very long, might read better as two sentences.

We have rewritten the background and objective sections of the abstract.

P. 6, line 18, replace relationship with relation

The sentence now reads:

In education testing, the item characteristic curve shows the difficulty of the item relative to a person's knowledge of that topic

P. 7, lines 2-3 and 7-11, sample results I believe should go under "Findings" rather than in the Methods section.

We agree. Demographic description of the sample is now in the Findings section.

P. 10, line 12, needs to be changed because the study is looking at responses to questions; the participant is not making the association of symptoms with depressed mood, correct?

Perhaps something like, "Hispanic does not report these symptoms as associated with depressed mood..."

We rewrote the sentence as follow:

Acculturated Hispanics are more likely to report these symptoms than non-acculturated Hispanics.

Again, editing: I found several editing needs, for example, singular verbs associated with plural nouns, (eg p. 5, line 3).

We have carefully proof read the draft for grammatical and spelling errors. There should be no grammatical or spelling errors in this revised version of the manuscript.

Thank you for the opportunity.

Reviewer #2: This Brief Research Report assessed differences in responses to the CES-D between women who completed a Spanish version of the questionnaire compared to those who completed an English version. The author states that language preference is being used as a proxy measure for acculturation. The topic is very important given the frequency of proxy measures for acculturation in the literature, and the lack of documentation on translation methods and equivalence of instrument forms when two versions are used in a study. This manuscript could make a contribution demonstrating an acceptable method for assessing instruments that will be used in translated forms across cross cultural samples.

There are a few (fixable) problems with the manuscript, however. The manuscript title should include the term "language acculturation" since only that domain of acculturation is considered.

We disagree with this reviewer about the use of "language acculturation" in the title. We do not feel that the term "language acculturation" clarifies the complex issue of measuring acculturation. Language preference is used in our study as a proxy measure of acculturation, and as such, we found it to be a significant predictor of other indicators of acculturation, such as years lived in the U.S. and place of birth. As pointed out in our address to Reviewer 1, many studies have used language as a proxy measure for acculturation, and those studies do not include the term "language acculturation" in their articles. We feel that "language acculturation" could be misconstrued by some readers as learning a new

language rather than using language preference as a reflection of the larger adaptation and adjustment into a new environment and culture over time for a particular study participant. Thus, no changes were made to our original title.

A more serious problem is that, even considering the space limitations of a brief report, the manuscript does not introduce why acculturation is expected to contribute to depression in this population.

The motivation for this article was the findings from recent studies that show the prevalence of depression differed for Mexican-Americans who were born in the U.S. versus those who immigrated here (Escobar et al, 2000; Vega et al., 1998). We do not know why U.S. born Mexican-Americans have a higher prevalence than their foreign born counterparts. One plausible explanation is that the instrument employed to measure depression may be culturally bias. We have stated such in the Background section of the manuscript, with a sentence summing up the rationale for the study:

Whether these prevalence rates represent real differences in population morbidity or the results of cultural differences in the conceptualization and expression of depressive symptoms are not known.

We have made no claim that acculturation contributes to depression. In fact, we have stated in the Background section that the evidence in the literature is inconclusive:

Regrettably, recent studies on depression in Latino groups and Mexican immigrants have been inconclusive on whether immigrant status is protective, especially for those from Mexico (Lara, Gamboa, Kahramanian, Morales, & Bautisita, 2005).

Our manuscript simply noted the differences in the prevalence of depression reported in recent studies, and we investigated whether cultural bias in instrumentation could have explained these differences.

In addition, a strong rationale is not articulated for conducting this study from a methodological or conceptual perspective.  
See our comments above.

The articles reviewed deal with differences in depression by country of birth and length of time in the host country, both of which are frequently used proxies for acculturation--but none of the background articles explain why language as a proxy for acculturation is chosen. Perhaps this study really examines cultural differences or simply language preference, since a strong rationale for equating language preference with acculturation status is not really presented. We have presented additional references to support the use of language as a proxy for acculturation.

The second paragraph on p. 5 (lines 4-9) is not clear in its relationship to the topic at hand.

We disagree. It is highly relevant because cultural bias could influence how people respond to a measurement resulting in different outcomes. Are immigrants protected from depression or do they simply respond to the depression instruments differently, thus resulting in a lower depression prevalence? The literature has not answered this question, and we made that case.

Finally, the author makes a case that the language or translation may affect the way words are interpreted, but then does not relate that to acculturation, which is the apparent purpose of the study. In general, the acculturation and translation issues are not differentiated or clarified as part of the problem statement so the rationale for this study remains a bit murky.

Acculturation and language are closely related in our study. We have used language preference as a proxy measure of acculturation. We did relate acculturation and the way words are interpreted in our manuscript by stating the potential bias in the CESD instrument in the Background Section as follow:

The instrument may be biased for several reasons. First, the translated instrument may fail to capture the conceptualization of depression for the other cultural group; that is, the instrument may neglect certain items in the domain of contents of depression for the other cultural group. Second, the instrument may include items that are highly relevant to the population from which the instrument was originally developed, but those same items are irrelevant to the other cultural group because they are not related to how that cultural group perceives or expresses depression. Third, the *emic*, or meaning behind the translated word, may not be the same between the two groups. Certain idioms, such as “could not get going,” are harder to translate. Finally, social desirability or stigma may also play a role in how people respond to a survey.

In summary, the goal of the article is to examine differences in performance on the CES-D between women who choose to take it in Spanish or English. But a secondary point is establishing conceptual equivalence or validity in the two versions. This measurement issue is not discussed but is in fact an important methodological component of the paper. How do we know that the differences between the two groups were due to acculturation vs. translation, and how might we be able to do so? How might one follow up on this study to make better judgments regarding the adequacy of translations in general or the issue of allowing participants to choose the language in which they take questionnaires without creating a biased sample? More sophisticated interpretation in terms of the measurement issues raised by this study would be useful and make more of a contribution to cross-cultural research.

Differential item functioning (DIF) was used in our study to identify CESD items that operate differently between the English version and the Spanish version. DIF is a method that could help us determine cultural bias for each item in the CESD scale. Translated instrument could use DIF to identify problematic items

(that is, items that does not operate the same way for both groups) for further investigation.

Reviewer #4: Re: "Effects of Acculturation on the Responding of Depressive Symptoms among Hispanic Pregnant Women"

#### Purpose

This study investigated whether acculturation influences the responses to symptoms items and the total scale score of CESD in Hispanic pregnant women. Acculturation was measured by subjects' language preference.

#### Background

The background about the differences in depression between Mexican descent born in the US and Mexican immigrants was clear. But there wasn't enough information about issues of using CESD in the Hispanic population. The rationale for studying instrument validity when translated into another language was sufficient.

DIF is a new method developed by researchers in the education field to study item bias and an instrument's validity. It has not been widely used in nursing. Thus more detailed description of the method is needed. I found the description of the method confusing on page 6. The analogy from education to mental health field did not logically flow. For instance, the DIF in education is based on the assumption that test takers who have similar knowledge (based on total test scores) would perform similarly on individual test items regardless of their characteristics such as sex and ethnicity. Translating this assumption into this study, it meant that subjects with similar depression status (based on total score) should perform in similar way on individual items of CESD regardless of their language preferences. But in the manuscript, the authors related the item response to a person's conceptualization or knowledge of depression. I am not clear how the authors deduced that relation from the education theory. We rewrote the paragraph and avoid using technical terms (e.g. continuum of variation of a latent factor) to present the idea more clearly.

#### Methods

There was not enough detail of the study methods. What were the inclusion criteria? Was the data collected by personal interview or paper-pencil method? If paper-pencil, were all women able to read and understand either English or Spanish? How many clinics were involved in recruiting subjects and who collected the data?

More details were added to the methods section.

I understand that the language preference could be used to measure acculturation. But there were other significant differences between the two groups of women (table 1) that could also explain the differences in CESD scores

other than acculturation, such as length of stay in the US and education. Actually, all demographic variables reported in the manuscript were different between the two groups of women. Given these differences, it is difficult to claim that the differences in DIF were due to language preference unless all other confounding variables were also controlled.

The differences reported in Table 1 support the use of language preference as a proxy measure of acculturation, especially the length of stay in the U.S. Education level and marital status also reflect cultural or societal differences between the two countries. The logistic regression models would be over adjusted if we included these demographic variables.

For the measure of CESD, who translated the questionnaire and what measures were used to ensure the accurate translation? Was higher score indicating higher depression?

These questions are similar to Reviewer 1, which we have addressed above.

For statistical analysis, Zumbo (1999) suggested using  $\alpha=0.01$  as the significance level to control for the multiple tests. Here I suggest the authors to do the same. With 20 tests to perform at each analysis, using  $p<0.05$  would inflate type I error too much.

We did use .01, so we are not clear about the reviewer's comment.

#### Findings

Although both uniform and interaction model (model II and III) were mentioned in the analysis section, there was no mention of model III in the finding section. Thus it is not clear whether there was an interaction between language preference and total score.

Table 3 and table 4 has been changed. Table 3 now includes the correlations and Table 4 now presents the DIF results from the logistic regression models. Appropriate changes to text have been made.

R-squared was used to measure effect size. Again, Zumbo (1999) suggested that  $R^2=0.13$  or 13% as the indication of reasonable effect size. In this study, non of the  $R^2$  reached 13. This should be mentioned in the result section. The chi-square change might be significant, but the effect size was small to medium. There was a typo in table 3 in the column of  $R^2$ .

We made the corrections as suggested.

#### Discussion and implications

In the discussion, I think "impact" or "influence" should be replaced with "related" because there were too many extraneous variables to the observed relationships thus it did not warrant causal inferences.

Changes were made as suggested.

## Summary

This is a study using DIF to test the validity of CESD in a group of Hispanic women. Applying this method in mental health situation is innovative and appropriate. This method offers opportunities for psychometric tests, especially in the instrument development stage. Since this method is relatively new to nursing, more detailed description is needed to understand the findings. There were multiple weaknesses in the method section that weaken the quality of this study. We revised part of the method section and added a figure to clarify the DIF concept and strengthen the paper.