

Molly C. Dougherty, PhD, RN
 Editor, *Nursing Research*
 School of Nursing CB# 7460
 UNC- Chapel Hill
 Chapel Hill, NC 27599-7460

Dear Dr. Dougherty,

March 1, 2006

Thank you for the opportunity to revise our manuscript #2005/191 entitled, "Psychometric Testing of the Revised 15-Item Bakas Caregiving Outcomes Scale (BCOS)." We are grateful for the detailed comments and suggestions provided by each of the 4 reviewers, and we believe that their input has greatly improved our manuscript. Because of the number of comments and level of detail provided by each of the reviewers, we have included their comments verbatim in italics, followed by our comments and revisions in bold. Although this makes a rather lengthy cover letter, we believe that it will provide you and your reviewers with the best explanation of the changes that we have made.

Reviewer #1

Ms NUMBER: 2005/191

TITLE: Psychometric Testing of the Revised 15-Item Bakas Caregiving Outcomes Scale
The manuscript focuses on the important concept of caregiving outcomes and the revision of a known measure of life changes resulting from providing care. Such measures are vital to identifying caregivers at risk and in need of intervention. The manuscript is well-written and organized and has the potential to make an important contribution to the field of family care. However, I have several concerns outlined below that I believe detract from the overall strength of the paper.

- *The introduction is well-written and relevant literature and reviews of the original measure are mentioned. I was disappointed that despite mentioning the weaknesses of the BCOS (as identified by two comprehensive reviews), the current manuscript did not address all concerns, in particular responsiveness to change (page 6, paragraph 2, line 20-22).*

Responsiveness to change over time was not measured in this study, but the BCOS has been measured from 1 to 4 months after stroke using the same dataset for this study in another publication (Bakas et al., 2006). We have addressed this in the limitations on p. 20.

- *The rationale for adding the 5 new items is given on page 3, paragraph 3, line 22-23. It appears that the argument is to increase the content validity of the measure, but this is never explicitly stated and content validity is never examined or evaluated in the manuscript (e.g. through expert panels, content validity index). By the end of the manuscript we are still left with this concern regarding the new scale. Additionally, on page 12, paragraph 1, lines 1-4 the authors mention that a former version of the scale originally included two additional items regarding financial outcomes, which were deleted due to low item-total correlations, low variability and conceptual ambiguity. It would have been helpful to have described how those two items differ from the additional 5 items, and whether conceptual ambiguity is still a concern.*

Content validity has been documented using experts for the original BCOS, and we did not feel it was necessary to consult experts again regarding the addition of the new items because they came from our prior psychometric testing results (Bakas & Champion, 1999). We have added detail regarding the new items on pp. 4-5, in more detail on p. 11, and in the discussion on p. 17.

- *The hypotheses as stated are very specific and detailed, particularly hypothesis #1. I am wondering if it would not be better to phrase the hypotheses around the key psychometric properties to be addressed and demonstrated e.g. reliability, validity, item analysis, variability....and leave the details of the analyses to be performed to the analysis plan.* **We have simplified the hypotheses on p. 5.**

- *I think it might be more meaningful to include significant others with spouses when describing the sample as both of these categories are very different from other relationships mentioned – they are romantic/intimate relationships quite different from children or other relatives and other family care research has shown them to experience the care process very similarly (page 7; paragraph 3, line 18-19).* **This has been included on p. 13 under sample characteristics.**

- *I think the authors do a great job of describing the scale and descriptive statistics and how they handled missing data. On page 12, paragraph 3, line 15 they mention that when there are less than 50% responses mean imputation will be used. This seemed like a large amount of missing data, but then the next few lines reveal that there was actually very little missing data. I think the authors may wish to consider how they phrase this section.* **We have reworded this section on pp. 11-12.**

- *The authors did a very nice job of describing the measures used in the analyses.* **We did have to delete a little detail regarding each measure pertaining to transformations, normality, and where caregivers scored overall for each measure in order to make room for us to address other comments. The rest of the measures information was preserved.**

- *I think there needs to be a rationale for why an intra-class correlation as opposed to another type of correlation was chosen to examine test-retest reliability and also why the cut-off of 0.60 was chosen. Currently, these choices are unclear to the reader and just need to be justified.* **References regarding the use of the intra-class correlation and cut-offs have been provided on p. 12 under Data analyses.**

- *Cronbach's alpha almost always increases with more items so the authors should be cautious in the way they discuss this data and perhaps place more emphasis on the fact that the longer scale is still a very internally-consistent measure.* **We worded this more carefully at the top of p. 18 in the discussion section.**

- *There are several places throughout the manuscript where evidence of unidimensionality and factor loadings above .40 are mentioned in the same sentence (page 2, paragraph 1, line 15; page 5, paragraph 1, line 1-2; page 14, paragraph 3; lines 15-19). I am concerned that some readers may assume that loadings above .40 are evidence for unidimensionality, which is untrue.*

It is entirely possible to have several dimensions and still have all loadings above .40. Related to this, the authors did not provide adequate evidence for the unidimensionality of the scale. As this was an important aim of the paper and reason for conducting an exploratory factor analysis, I believe that further evidence is necessary. Currently, the authors provide the factor loadings for a 1-factor solution, the eigenvalue for this factor and percent variance accounted for by this factor. The authors also state that an exploratory factor analysis did not produce any interpretable factors (page 14, paragraph 3, line 17). More information is needed. It is often the case that a 1-factor solution looks the best before rotation – no information is given as to whether the factor analysis solution is before or after rotation, and if applicable the type of rotation used. No information is given regarding the other eigenvalues. I think it would greatly benefit the reader to know how many eigenvalues were above 1.0 and a sense of where the gap (scree) might have been. One main reason for raising these concerns is that when the factor loadings for the 15 items are placed in hierarchical order, it appears that there are 3 items that load between .40-.49 and 1 item at .55; all other items are loading above .62. These are the same items with low item-total correlations. 3 of these 4 items deal with questions regarding time for family, relationship with family and time for social activities/friends. I think if the authors could provide the evidence I outlined above they could better convince the reader that the scale is unidimensional, and that these 3 items are not hinting at a second factor. There just isn't enough information given regarding the factor analysis to conclude unidimensionality. **We are grateful for these comments and have provided more detail. On p. 12 under data analyses, we have added confirmatory factor analysis to further clarify the one-factor model. The findings regarding both the exploratory factor analysis and the confirmatory factor analysis are provided on pp. 14-15. More detail has been added regarding the one-factor and two-factor solutions, including inspection of the scree plot. The initial Eigenvalues are provided at the bottom of Table 2. Also provided on p. 15 are results from the confirmatory factor analysis showing good fit. Further discussion of these analyses is provided on p. 18.**

- *I am concerned about the general item that is used to assess criterion-related validity. This item is worded similarly to the 15 items on the scale but is an overall question regarding life changes as a result of care. I am not convinced that a correlation between the total score for the 15-items and this single-item is a strong case for criterion-related validity, and the high correlation of .67 raises the question of why a 15-item scale is needed rather than a single item. Additionally, concurrent criterion-related validity is never as strong as predictive validity as both occur at the same time and the desire for participants to be consistent can bias results. The authors just need to address these issues.* **We have addressed these issues on p. 20.**

- *I very much liked the comparison between the original scale and the 15-item scale on pages 18, paragraph 1, lines 1-5. I think the manuscript would really be strengthened if this comparison could be used throughout.* **This was also a concern expressed by reviewer #4 who suggested that we capitalize on the opportunity to analyze the 10-item BCOS using the same dataset for comparison. We have done this throughout the manuscript and have included data regarding the 10-item BCOS in Tables 1 & 2. The results from the 10-item versus the 15-item BCOS are compared throughout, with a summary provided in the Conclusions on p. 21.**

- While I recognize the tremendous amount of work it has taken the authors to get to this point, I raise the possibility of moving beyond classical test theory to consider the advantages that generalizability theory (in addition to classical test theory) might add to evaluation of the scale. I commend the authors for the work they have done and the importance of this scale. I think with a little more detail and discussion regarding the analyses, and some revisions outlined above, this will make an important contribution to the field. **We have added a citation and recommendation regarding Generalizability Theory in our section on Limitations on p. 21.**

Reviewer #2 –

Problem statement: (1) adequate but could be improved

In the first sentence of the third paragraph, it isn't entirely clear to me what is meant by "...several conceptual attributes were not adequately developed as described in the methods section." When the authors are talking about the report of the development of the original Bakas & Champion 1999 measure. Please if possible explain why it is necessary to add items to the scale if the scale what value is added – I'm not clear on that from the introduction. **We clarified this in the introduction bottom of p. 4 and top of p. 5. Further detail was provided on p. 11.**

Attention to relevant literature: (2) inadequate but can be fixed

None of the antecedent variables are covered in the literature review (caregiver optimism as a personality factor, dependent care tasks as a situational factor, nor caregiver appraisal as a mediator variable). In addition, the relationships among the variables depicted in the conceptual model are not described or covered in the literature review. This makes the findings more difficult to interpret within the context of the theoretical framework the authors are using to demonstrate construct validity. **We did not have space to do a detailed literature review on all of the proposed relationships in the conceptual model. However, we did provide citations regarding empirical support for the model on p. 5 (Bakas & Burgener, 2002; Bakas & Champion, 1999), and added some findings in the Relevant literature section on p. 6 that showed that self-esteem, dependent-care tasks, and appraisal (using a similar model) explained 48% adjusted variance in emotional distress, 25% adjusted variance in perceived general health, and 45% adjusted variance in caregiving outcomes as measured by the 10-item BCOS (Bakas & Burgener, 2002).**

Theoretical framework: (2) inadequate but can be fixed

Although the Lazarus theory of stress and coping seems a reasonable choice as theoretical framework for this project, no rationale for its selection is provided. Without a strong rationale for the use of this framework, and then a review of literature to support the selection of the independent variables used to test theoretically the relationship of the BCOS to those variables, the construct validity findings are not as strong as they could be. Also, the conceptual model figure is not very clear, and the relationships depicted are not explained in the text. Not all of the relationships depicted in the figure were ultimately tested in the hierarchical regression (because all demographic and model variables were screened for inclusion in the regression equation p.13), and this should be indicated on the figure. **The model has been clarified as much as possible given the length restrictions of the manuscript on p. 5. We have also clarified Figure 1 to include headings that correspond with the theory (Caregiver personality factor, situational factor, appraisal, emotional outcomes, and adaptational outcomes). Also noted**

in Figure 1 are the variables that were not entered into the regression equation because they were screened out as described in the methods section.

Research design: (2) inadequate cant be fixed now, but must be addressed somehow

The authors acknowledge that the cross-sectional design is a limitation, that a longitudinal study will be necessary to document how well the BCOS performs over time with stroke caregivers. But given that the main strength of the original measure is that it is the only reported instrument that measures perceived changes resulting from providing care – it seems crucial that evaluation of the tools validity should include some aspect of its ability to capture change over time, and this cannot be done with a cross-sectional design. What argument can the authors offer that the revised scale should be used in an expensive family caregiving study without being able to address this critical aspect of tool performance. We attempted to address this concern in the limitations section on pp. 20-21.

Data analysis: (1) adequate with some reservations

I was very unclear about how missing data were handled and whether or not the pattern of missing data were examined. This I think needs to be added. I am also concerned about the number of scales that had to be transformed. For me, this makes interpretation of the regression findings difficult. Can we ask a statistician if something needs to be said about this? I am not skilled enough to know. Also on page 13, there is a statement about how variables were screened for inclusion in the regression equation. The rationale for this screening from an authoritative source wasn't offered, I would like to see a citation for this. I think it would also be helpful to have a statistician review the article. We have described on p. 12 that there was very little missing data for the BCOS and for the rest of the scales on p. 14. Given that there was so little missing data, a meaningful comparison of missing vs non-missing data is not feasible. As for the transformation issues, Reviewer #4 also had concerns, therefore those changes are addressed in Reviewer #4's comments to follow.

Reviewer #3 –

The manuscript is a report of an instrument development project, in which the original 10-item form of the BCOS was expanded to include 15 items. The new scale was evaluated for its psychometric properties.

Problem statement

The need for a measure of outcomes following assumption of a stroke caregiving role is well-established. Stroke is common and most survivors are cared for at home by family caregivers. Such a measure could be used in clinical practice to identify caregivers in need of additional resources and in research to evaluate effectiveness of various intervention strategies.

Literature review

The author(s) include most of the relevant resources and identify how the science has evolved recently in this area, thereby justifying the need for a more comprehensive caregiving outcome measure.

Framework

The conceptual model is based on stress and coping theory. The primary outcome variable is identified as Caregiver Adaptational Outcomes, which is proposed to be measured by the BCOS. These outcomes are the “emotional and adaptational outcomes of providing care” (pg 4, line 13).

Design & Analysis

*The problem related to caregiving outcomes and the need to reliably and validly measure these outcomes is clear. Of some concern, however, is the simultaneous presentation of caregiving outcomes as unidimensional, yet with most of the recommendations directed to interpretation, at the level of individual items. Recommended uses of the measure mostly include identification of specific aspects in need of attention, as identified by variation in score of individual items. There is limited discussion about what a score from the total scale might mean, and how it's recommended for use. The implication is that scale interpretation should be a series of interpretations of the 15 single items. **We have mentioned that the BCOS could be used at both the item level and the scale level in the discussion section at the bottom of p. 17, top of p. 18 and in the conclusions on p. 21.***

*In discussion of the hierarchical multiple regression analysis (pg 15, line 11), it looks the percent variance should be 5, to match Table 3. **The regression was rerun without transformations as recommended by Reviewer #4, and the % variance values between the text and the table have been checked for consistency.***

Results

*The first sentence of the discussion probably should refer to stroke caregivers, rather than stroke survivors (pg. 15, line 21). The statement about use of items at the individual level (pg. 16, lines 13-15) implies a much weaker stance regarding this issue, than does the concluding statement (pg. 18, lines 22-23). It would be helpful to identify a clearer position and recommendation regarding this critical point. **We have fixed our typo as indicated, and have also provided a clearer position regarding this point bottom of p. 17, top of p. 18.***

Organization & Style

Well-written manuscript.

Reviewer #4 –

This paper provides a psychometric assessment of a revised version of the Bakas Caregiving Outcomes Scale (BCOS). An article on revision of this scale is of importance and would have sufficiently wide appeal. However, there are quite a few serious issues with the assessment provided in this paper.

| *Serious Issues.*

1. The presentation of the components of the model in Figure 1 and their descriptions in the text on page 4 lacks consistency and clarity. The figure does not distinguish between personality and

situational factors. The reader has to infer that emotional outcomes consist only of caregiver depressive symptoms and that SF-36 GH stands for self-perceived general health. Time is listed as a dimension of caregiver dependent-care tasks in the figure, but is not mentioned in the text. Caregiver and stroke survivor characteristics and survivor impairment are not discussed in this part of the text. It is stated that the original theory supports mediation by coping as well as by appraisal, but Figure 1 and the discussion do not appear to address coping at all. Caregiver optimism is described as a antecedent personality factor, but later on page 9 it is called an antecedent personality disposition. What is the difference between a factor and a disposition?

We have revised Figure 1 to provide added clarity and to try and save some space in the text. The theory headings have been added to Figure 1 (i.e., personality factor, situational factor, appraisal, emotional outcomes, etc.) and the variables that were excluded from testing were also indicated at the bottom of Figure 1. We have also removed any reference to coping, since this was a component that was not tested. Going into a lengthy discussion of why coping was not included is not possible due to page limitations. We have also removed the term disposition to clear up this inconsistency. All the terms throughout the manuscript were double checked for clarity and consistency.

*2. The research hypotheses, and hence most of the paper, address the psychometric properties of the revised 15-item scale by itself and not how it compares to the 10-item scale it revises. There is only a limited comparison at the end of the Discussion Section of results for the 15-item scale and those from a previous study of the 10-item scale. The authors have not capitalized on the special opportunity their current data provides for comparing results for the complete scale development process applied to all 15 items to associated results when that process is applied to the original 10 items from the same data set. **This was a key suggestion that we had not considered with the first submission of the manuscript. We have provided data for the 10-item BCOS in tables 1 & 2, and have made comparisons throughout the manuscript.***

*3. Seventy-one percent of the data used in the psychometric assessment come from a study of caregivers of stroke survivors while the rest come from a second study of aphasic stroke survivors. The authors do not justify combining these two subsamples, do not compare available subject characteristics and outcomes for the subsamples, and do not address whether the results would have been any different if applied to separate subsamples. A related problem is that the test-retest assessment is based on the second, smaller subsample alone. It is possible that those results do not generalize to the larger subsample. This seems to be a serious flaw that could have been avoided through better design, but does not seem to be rectifiable any more. **We have provided a detailed rationale for the two groups of caregivers in this study on p. 7 under Methods, as well as group differences published in a prior study (Bakas et al., 2006). We have also included in the limitations on p. 20. that the test-retest was only completed with the aphasic sample.***

4. The 2 items dropped from the 12-item BCOS to obtain the 10-item BCOS were related to financial outcomes, but one of the new items for the 15-item scale addresses "my financial well-being" (page 12). If consideration of financial outcomes was of little or no value earlier, why consider them again? How differently was this issue addressed that would suggest it might now be valuable given that earlier results suggested it was not? Its loading in Table 2 of 0.41 is the lowest for all 5 new items. Some authors suggest dropping items with loadings below 0.40, so

this one is right above that boundary. Given its marginally acceptable loading and the earlier problems with addressing this issue, why not consider a 14-item BCOS instead? What impact on the results is there to including this item or not? **This concern was also expressed by the other reviewers. We have added detail regarding the new items on pp. 4-5, in more detail on p. 11, and in the discussion on p. 17.**

5. The claim that the items exhibit unidimensionality is not sufficiently supported. Having a high loading on a given factor can justify the use of an item in a summated scale associated with that factor, no matter how many factors have been extracted. The fact that this holds for all item loadings of the 1-factor solution does not justify the use of that solution over those based on higher numbers of factors. What does it mean for a set of items to "fit within" a single factor (page 14)? In what sense did exploratory factor analysis "not produce any additional interpretable factors" (also page 14)? **We responded to this concern for Reviewer #1 as well. On p. 12 under data analyses, we have added confirmatory factor analysis to further clarify the one-factor model. The findings regarding both the exploratory factor analysis and the confirmatory factor analysis are provided on pp. 14-15. More detail has been added regarding the one-factor and two-factor solutions, including inspection of the scree plot. The initial Eigenvalues are provided at the bottom of Table 2. Also provided on p. 15 are results from the confirmatory factor analysis showing good fit. Further discussion of these analyses is on p. 18.**

6. The authors do not seem willing to admit the seriousness of some of the limitations to their results. On page 17, the results quoted on lines 12-13 indicate that the benefit the 10-item BCOS had in being related to task difficulty appears to no longer hold for the 15-item BCOS. Since the original 10 items are still included, perhaps the extra 5 items have masked their relationship with task difficulty. Alternately, maybe the relationship between the 10-item BCOS and task difficulty that was identified in earlier work is not likely to hold in many other studies. We have addressed this in the discussion on p. 19. Further, on page 18, lines 12-14, the referenced work provides supports for the earlier 10-item BCOS not the revised 15-item BCOS and so does not really ameliorate the limitations of the 15-item BCOS listed prior to this statement. Also, while the results provide some support for the conceptual model of Figure 1 as stated on page 17, lines 18-20, it also suggests that major parts of that conceptual model do not hold including the impacts of 3 of the 6 nodes corresponding to subject characteristics as well as personality and situational factors. We have provided discussion on why task difficulty was not a significant predictor, and also why optimism was not significantly related to the BCOS on p. 19. We have also indicated in Figure 1 the variables that were not entered into the regression equation for testing to provide clarity.

7. Results are reported in Table 1 for measures some of which have been transformed. It is not clear, though, whether means, standard deviations, and actual ranges are reported for the measures before or after being transformed. For the reported alpha values, it doesn't seem possible they could be based on the transformed measures since items would need to be transformed not just total scores. Is this the case? Table 3 reports regression results based on these measures, but does not indicate that some have been transformed. A footnote seems necessary in that table identifying the variables that have been transformed and how they were transformed. The outcome measure for this analysis (BCOS score) was not transformed while

five of the predictors were. The issues addressed by transformations are really related to assumptions about the outcome variable of an analysis, not about its predictors. Since a regression analysis is based on a model for the outcome conditioned on observed values for the predictors, the untransformed predictors could be validly used in the analysis. Hence, it appears that the wrong analysis has been conducted. The appropriate analysis should start with a regression analysis of BCOS scores in terms of untransformed predictors. If a certain predictor that was expected to be significant was not, it could be transformed to determine if there was a nonlinear relationship between it and the outcome. Whether the predictors are transformed or not, the residuals generated by the model need to be assessed for compatibility with model assumptions. Even if an outcome is reasonably symmetric when considered separately, it is still important to assess assumptions for a regression model of that outcome. Plots of residuals versus individual predictors could be used to decide which predictors might benefit from being transformed. The transformation needed to make a predictor more symmetric when considered separately need not be the same as the transformation that best improves its value for predicting a specific outcome. **We are grateful for the direction in this area. We have rerun the regression without transformation as suggested, and have provided further information regarding the residuals scatter plot on p. 16.**

8. The reported *F* statistics for Steps 2-4 of Table 3 address the complete model to that stage based on all predictors added so far into the model. These are not the salient *F* statistics. Given that the reported *F* value for Step 1 is significant, the reported *F* statistics at later stages must necessarily be significant as well. What should be tested is whether or not the additional terms added at each stage provide a significant improvement in the *F* value. **We added change statistics to Table 3.**

9. Research Hypothesis 4 was not fully supported by the results of the regression analysis. It was hypothesized that personality, situational, and appraisal variables would be significantly related to BCOS scores following the relationships depicted in Figure 1. While variables of all three types were significantly related to BCOS scores when considered separately, only appraisal variables were significantly related in the final regression model encompassing all three types of variables. For this reason, the statement on page 17, lines 6-8 in the Discussion Section that 35% of the variance in BCOS scores was accounted for by "the constructs of Figure 1" is misleading since only few of those constructs actually accounted for this variance. **We clarified this in Figure 1.** On the other hand, this result does have the positive aspect not mentioned in the paper of supporting complete mediation of the impacts of personality and situational factors on BCOS scores by appraisal factors. **We included this point in the discussion toward the bottom of p. 19.** Another related problem is the inclusion in the analyses of only those predictors which were significantly related to BCOS scores in bivariate analyses. While this is a reasonable strategy in an exploratory analysis setting, it does not exactly address what was hypothesized about the construct validity of BCOS. To truly address relationships hypothesized prior to data collection, the model incorporating all those hypothesized relationships needs to be tested not just those relationships which happen to turn out significant in related data analyses. Of course, after testing that hypothesized model, it is not unreasonable to address modifications of that model to improve it. **We have provided our rationale for the screening criteria for inclusion in the data analyses section bottom of p. 12, top of p. 13. We have also indicated variables in Figure 1 that were not tested because of screening.**

Other Issues.

1. *Research Hypothesis 5 is expressed in terms of the "BCOS criterion-item" and the "SF-36 Health Survey General Health Subscale" but these are not defined until much later in the Methods Section.* **These were taken out of the hypotheses when we simplified them as recommended by Reviewer #1.**

2. *On page 8, the first sentence on lines 5-6 seems improperly stated. Wouldn't the mean BCOS scores differ with the two gender variables as is the case for a two-sample t test? The way it is stated would be appropriate for a logistic regression predicting gender from BCOS scores.* **We reworded this statement as suggested and placed it at the bottom of Table 3 for clarity and to save space in the text.**

3. *On pages 8-9, it is not clear what the difference is between the SSQOL and the SSQOL Proxy. Is the latter the same as the scale based on the 5 domains used in the paper? If not, what justification is there for this adaptation of an existing scale?* **Clarified on p. 8.**

4. *The Method Section includes results which seem more appropriately placed in the Results Section. Instead of presenting a description of how the sample was obtained, sample results are reported in the Method Section for subject characteristics and how they related to the BCOS outcomes.* **We moved subject characteristics to the results section.** *Instead of just describing the various measures used in the paper and published results for those measures, the Method Section also presents results for the current data from Table 1.* **Moved to results section.** *Similarly, instead of just describing the data analysis process in the Methods Section, some of the results for the regression analyses are also included. This is done to justify power for that regression model, but power is more appropriately addressed by considering the number of variables proposed for consideration prior to the analysis rather than the number that turned out to be significantly related to the outcome in the analysis.* **Moved to results section.** **We did not do a power analysis prior to psychometric testing, but did keep this statement in the results section to justify that we had enough subjects given the number of predictors that were selected.** *This appears to be a serious issue. Using the formula provided on page 13, even one more variable (from 11 to 12, for example the time variable that was considered but not significant) would increase the required sample size from 138 to 146. While this is just less than the value of 147 reported on page 13, only 144 observations were available for the regression analysis due to missing data. It appears they have just below the required sample size rather than just above as claimed.* **We did not change the number of predictors, thus our sample was adequate for the regression.**

5. *Comorbidities for caregivers and how long they have been providing care do not seem to have been considered as subject characteristics. Wouldn't these have the potential for impacts on the outcomes? Why not consider them?* **We did not measure these.** **Furthermore, length of time caregiving is often unclear for caregivers to answer since some spousal caregivers have been known to indicate they became caregivers at the time of marriage, sometimes 40 years ago.** **Our sample does consist of caregivers who provide care for stroke survivors**

approximately 4 months after stroke, so we believe that this is a better indication of their length of caregiving.

6. *On page 9, line 21, should "OCBS" be changed to "the OCBS time subscale"?* **Changed as suggested.**

7. *On page 11, lines 5-6, is the additional item also measured on a 5-point scale?* **Changed as suggested.**

8. *On page 12, line 9, rather than describing the item as the "last" one which can be confused with the last one of the previously listed 5 items, why not describe it as something like "an extra item"?* **Changed as suggested.**

9. *Table 3 requires a footnote indicating that the results are based on a subset of 144 subjects so the reader does not have to search for this fact on page 14. The footnote in that table about the two gender variables would be more informative if instead of giving the coding for gender it indicated that the gender effect was for females compared to males, which is the consequence of coding females with a larger value than males.* **Footnotes added to bottom of Table 3.**

10. *On page 16, in the sentence on lines 13-15, the authors claim for their uniqueness in documenting the quality of items seems exaggerated. Are they really claiming that, of all scale development studies in the psychometric literature, there are few that have assessed their items as much as in this study, which can be described in the single sentence beginning on line 5? Such a statement requires that detailed support be provided in the Relevant Literature Section based on extensive literature searches.* **This statement was deleted.**

11. *On page 18, line 15, consider rewording "specificity" since it might be confused with the term as used in the context of a logistic regression.* **The word "specificity" was removed.**

We greatly appreciate the expertise and time spent by the reviewers to provide comments and suggestions for improvement. We have attempted to address each of the comments and suggestions in sufficient detail, while at the same time trying to remain within page limitations for the manuscript. In doing so, we have exceeded the recommended limitations of 14-16 pages of text (excluding title page, abstract, tables, figures, references, etc.) by 1 ½ pages. We do not know where else to cut without jeopardizing the inclusion of the revisions suggested by the reviewers. We have only included 3 tables and 1 figure, so we are hoping that there will be room for the extra 1 ½ pages of text. We are very open to receiving further requests for revisions or suggestions on how to decrease the length of the manuscript if needed. Thank you again for the opportunity to revise our manuscript and to learn from the expertise of your reviewers.

Respectfully submitted,
 Tamilyn Bakas, DNS, RN, FAHA
 Associate Professor