

Review #4 (statistical review) - anonymous

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 an 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?
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.
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 sub samples, do not compare available subject characteristics and outcomes for the sub samples, and do not address whether the results would have been any different if applied to separate sub samples. A related problem is that the test-retest assessment is based on the second, smaller sub sample alone. It is possible that those results do not generalize to the larger sub sample. This seems to be a serious flaw that could have been avoided through better design, but does not seem to be rectifiable any more.
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?
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)?

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. 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.

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 a regression result 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.

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.

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. 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. 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.

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.
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.
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?
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. 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. 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. 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.
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?

6. On page 9, line 21, should "OCBS" be changed to "the OCBS time subscale"?
7. On page 11, lines 5-6, is the additional item also measured on a 5-point scale?
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"?
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.
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.
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.