

Dear Dr. Dougherty,

We greatly appreciate the very extensive and thoughtful review of the manuscript, Reliability and validity of the Restorative Care Behavior Checklist, and have attempted to address each of the reviewers concerns. In this cover letter we articulate how each concern is addressed.

Reviewer #1:

1. The authors describe U.S. policies that impact nursing assistants' provision of restorative care. As an international reader, I believe that the measure is relevant beyond the U.S.A. The authors might consider clearly indicating that these are U.S. policies but that the philosophy of care is not limited to the U.S. Similarly, stating that the reimbursement system (p. 4, line 9) is a U.S. system, might clarify the context for international readers.

We have added a statement to indicate that the Omnibus Budget Reconciliation Act (OBRA) of 1987 and reimbursement issues are only relevant to those in the United States. In addition, we conclude our introductory paragraph indicating that although mandated in the United States restorative care is relevant to other countries.

2. The "Factors that Influence Restorative Care Behaviors of NAs" section of the literature review describes the Restorative Care Behavior Checklist being based in self-efficacy theory. The literature review clearly links this theory to evidence about nursing assistants' work. It does not include information about the relationship between knowledge and performance of restorative care behaviors - but this is one of the factors in the theoretical model that is used for validity testing.

We have included a sentence to indicate that knowledge can influence behavior through self-efficacy expectations. That is, based on the theory of self-efficacy knowledge about performance and benefits of performance of a specific behavior can positively influence behavior through self-efficacy.

3. The discussion section includes additional factors external to the nursing assistant that may influence restorative care behaviors. If, in the introduction to this section, the authors explicitly state that the checklist was developed based on self-efficacy theory and that this theory is a model of internal factors that influence an individual's behaviors, then the lack of discussion of the external factors in the literature review and the connection between this section and the discussion would make more sense. Otherwise, a more complete literature review would discuss the full range of factors that influence nursing assistants' work.

We have included a paragraph in the introduction to address the relationship of additional factors, specifically job satisfaction, gender and age that may influence performance of restorative care activities. This is relevant for convergent validity testing.

4. The theoretical model that is used for validity testing is summarized in a complicated diagram. I was not able to distinguish between the arrow-heads on the diagram, so the hypothesized relationships between constructs and variables was not clear to me. The hypothesized relationships are not clearly stated in the literature review or the validity testing subsection of the methods section.

What does education mean in the figure?

We have revised Figure 1 to include only significant paths and have separated out the lines so that the individual arrow heads can be seen. Hopefully this clarifies the model. We have also stated the hypothesized relationships that are being tested within the model in the data analysis section. Education is measured in years of traditional schooling. In addition, prior education related to restorative care is included in the model as a separate variable.

5. Clarify whether measures other than the Restorative Care Behavior Checklist are published measures or whether they were developed by the authors for this study. There is no citation for the prior use of the Job Attitude Scale (p. 13, line 3-4).

The reference for the Job Attitude Scale is included in the section describing the measures. This section specifically provides the reference for the reliability and validity testing previously done on the Job Attitude Scale.

6. Include description of the residents in the nursing homes or the residents for whom care was observed as well as what time of the day care was observed. This would help with interpretation of the findings that some types of care were rarely observed. It would also help with interpretation of discussion that the findings might be influenced by the kinds of patients who were observed.

Unfortunately we do not have a description of the residents being evaluated or the time of the day of testing. We have acknowledged this as a limitation in the study. Further we agree with the reviewer that this information would be helpful in the interpretation of why a behavior may or may not have been performed. With regard to restorative care activities, however, we believe that regardless of resident physical status restorative care should be provided during all activities. For example, even a patient who has significant neurological disease, contractures and/or is totally dependent on nursing for activities of daily living should receive hand over hand bathing, dressing and feeding.

7. Why is it not necessary to do any of the psychometric analysis on the summary score of the checklist?

We were not clear what the reviewer was suggesting here. The summary scores were used in validity testing as per model testing and reliability testing based on inter-rater reliability. The summary score of the RCBC is the % of restorative care activities performed / all of the activities that were observed.

8. Why did the authors decide against using kappa to test for item agreement? I would be interested in knowing about agreement on two levels: (1) whether or not the care activity was being provided, the denominator of the summary score; and (2) if the care activity was being provided, whether or not a restorative care approach was used, the numerator of the summary score. This would help with interpretation of the finding that several of the activities were rarely performed. Disagreement about whether an activity is being performed would seem to be less important than disagreement about whether a restorative care approach was being used. It is not clear from the manuscript whether the Rasch analysis could differentiate this.

We agree with the review in terms of use of Kappa and have revised to manuscript to include only the kappas. The observations were done such that the observer indicated whether or not the activity was being observed (e.g. bathing, dressing etc), and then whether or not restorative care was provided during the care. Therefore, we were evaluating the reliability of observations with regard to whether or not the care activity was provided and then whether or not a restorative care approach was done. The total score of the RCBC reflects the percentage of restorative care activities performed when considered against all the observed activities performed by the NA. In this study the average percentage of activities in which restorative care was provided was 63% (evidence of restorative care behavior/all observed behavior). We did not feel that it was relevant to consider just the total number of restorative care activities performed without considering what was observed. We can only speculate why the other activities were not being performed more frequently. This has been noted in the limitations section.

9. Data about education are missing from Table 1.

Table 1 has been revised and the row with missing information removed (this was inadvertently left there from a prior revision). Education was reported based on number of years as shown in the table.

10. There are duplicate lines in Table 2 (Number of restorative care activities performed and Number of restorative care activities not performed, with different rounding for SD of number of restorative care activities not performed). Please explain what these lines mean in the text.

The duplication of these lines was an oversight and has been removed.

11. I don't understand what is meant by "Overall they performed restorative care during 64% of the observed activities (evidence of ....)" (p. 16, line 1 to 2). I assume this is an average over the 386 nursing assistants and that there was variation among the nursing assistants. Please include information about variation between nursing assistants.

The scoring of the Restorative Care Behavior Checklist was done by calculating the number of restorative care activities performed over the total number of activities that were observed. This is a cumulative measure based on observations of all of the 386 NAs. The percentage of restorative care activities performed and the standard deviation

is now provided in Table 2 (63% with a SD of 31%). We have removed the mean and standard deviation of the number of activities performed and those not performed as we believe this has little meaning without taking into consideration what was observed.

12. Explain what "missing" means in Table 3 and clarify the sample size. There are consistently 8 missing and the total observations (not applicable + performed + not performed + missing) is 395 (387 + 8 missing). The reported sample size is 386 (p. 10, line 14). This doesn't make sense.

That column should not have been in the table. I think it was copied from a previous table. It has been deleted.

13. The Validity Testing subsection concludes that the fit statistic for Communication OUTFIT is high but not important enough to threaten fit of the model. This is inconsistent with the description of the Rasch analysis in the methods section. A citation to support this conclusion as consistent in Rasch methods would assist the reader.

OUTFIT statistics are unweighted, being affected more by unexpected responses far from the item or person (e.g., a person of low ability unexpectedly getting a correct score on a difficult item). INFIT statistics are weighted, and are affected more by unexpected responses close to the person or item (e.g., a person of low capacity unexpectedly getting an easy item incorrect). High OUTFIT statistics are generally less of a concern for the validity of a measure (Smith & Smith). This is explained in the data analysis section and addressed again in the discussion section of the manuscript.

14. In the discussion section, I am interested in knowing whether the authors think that the low frequency of providing care related to eating, assistive devices and exercise might have impacted internal consistency. Is it possible that the hypotheses used for validity testing are wrong? It is not clear to me that the study, as reported, provides support for validity of the RCBC (p. 20, line 15). The convergent validity testing results do not support the conclusion of validity.

We agree with the reviewers that the limited observations of eating, use of assistive devices, and exercise may have influenced the internal consistency of the measures in this study. The limitations section has been revised to indicate this limitation.

With regard to the convergent validity of the RCBC the discussion has been revised based on reanalysis of the model. Overall, this study provided minimal evidence for the concurrent validity of the RCBC. The hypothesized direct relationship between self-efficacy and outcome expectations with performance of restorative care was not supported. Self-efficacy expectations were, however, indirectly related to restorative care activities through job attitude. In addition the hypothesized paths between gender and job satisfaction to restorative care activities were supported.

Reviewer #2:

1. Page 4 lines 18-23. First time that an explanation of restorative care behaviors occurs. I would recommend defining restorative care early on in the manuscript for readers that may not be as well versed as the authors in the topic.

We greatly appreciate this suggestion and have revised the manuscript so that restorative care and restorative care activities are defined in the first paragraph. Specifically we explain that restorative care programs focus on having caregivers engage residents in functional tasks and exercise. Restorative care activities provided by caregivers include such things as verbal encouragement to have the resident walk to the dining room, walking with the resident versus pushing him or her in a wheelchair, or arranging the environment so the resident can complete the bathing or dressing tasks that he or she has the underlying capability to perform.

2. I would also recommend that the author(s) spell out nursing assistant and include the abbreviation early in the manuscript. I only saw one way or the other but never together to explain what the abbreviation stood for.

The manuscript has been revised so that Nursing assistant (NA) is spelled out with first use and then referred to as NA throughout the remainder of the manuscript.

3. Page 8, line 16. I would also suggest spelling out the Restorative Care Behaviors Checklist and including the abbreviation the first time it appears.

The manuscript has been revised so that Restorative Care Behaviors Checklist (RCBC) is spelled out with first use and then referred to as RCBC throughout the remainder of the manuscript.

4. Page 9, lines 4-6. Why the heterogeneity in the evaluators? I don't see how this potential impact was addressed in the results?

We have revised the discussion section to reflect the significance of mentioning the heterogeneity of the evaluators. We believe the strong inter-rater reliability noted in this study is particularly significant given that the evaluators were not all nurses and did not have prior experience in nursing home settings working with older individuals.

5. Page 10, lines 7-9. This seems a little vague. What do you mean by "known sources of information"? I would suggest being a little more specific here and providing citations.

The "known sources of information" that influence self-efficacy are those that were described by Bandura in his early theory development. Specifically these include performance of a behavior, verbal encouragement to do the behavior, role modeling or seeing like others do the behavior, and physiological feedback or the affective states

associated with doing the behavior. We have revised the manuscript and eliminated the statement, “known sources of information”.

Based on the theory of self-efficacy, and the factors that influence self-efficacy expectations such as having previously performed restorative care or learned how to provide restorative care, it was hypothesized that: age, gender, years of traditional education, prior training in restorative care, and years of experience as an NA would directly influence knowledge of restorative care, self-efficacy and outcome expectations.

6. Page 10 lines 22-23. How did you take into account the potential effects of 36% of the nursing assistants having prior training on restorative care?

The impact of prior restorative care training was included in the model tested. These paths were not significant however, indicating that prior education did not influence self-efficacy or outcome expectations or performance of restorative care activities.

7. Page 11 lines 13-17. How many total items are on the RCBC? I did not receive the appendix A you refer to.

We apologize that Appendix A was not included for the reviewers. We are not sure if this is something that we did incorrectly during the manuscript submission or related to what was sent to the reviewers. We have included this again in this submission. There are 10 items on the measure.

8. Page 13 lines 15-20. The advantage to using the Rasch Analysis over KR-20 or Cronbach alpha is not clear.

We have provided a brief explanation of the use of the Rasch analysis measurement model over tradition KR-20 or Cronbach alpha in the section on reliability testing. An even more indepth discussion is probably beyond the scope of the paper. Essentially with KR20 or Cronbach alpha (which accounts for error due to content sampling) the estimates are bounded by zero and one and reflect the average inter-item correlation among the item responses. In this context persons with high or low scores have less error variance than persons with scores near 50%. Therefore, the average person variance used in KR20 will always overestimate the error score variance of persons with high and low scores. Furthermore, the item variance for an average person sampled is not the same as an average of the individual person test score error variances. In a Rasch measurement model each person’s ability and each item’s difficulty are on a linear scale. Being on a linear scale, these estimates are suitable for the calculations of means and variances. The Rasch models also provide a direct estimate of the modeled error variance for each estimate of a person’s ability and an item’s difficulty. These standard errors provide a quantification of the precision of every person measure and item difficulty and can be used to describe the range within which each item’s true difficulty or person’s true ability falls.

In the text we note that.... High person separation reliability means that there is a high probability that persons estimated with high measures actually do have higher measures than persons estimated with low measures.

We have also added a statement that the Rasch method using person separation reliability is analogous to the KR-20 or Cronbach alpha so that the reader has an immediate understanding of the results (i.e. a .77 personal reliability is the same as an alpha coefficient of .77).

9. Page 16 lines 21-22. Earlier you stated that exercise was one of the least likely activities to be observed?

We have clarified in the text that exercise was one of the least likely activities to be observed. However, when *it was observed* the NAs were likely to provide restorative care.

10. Page 18 lines 14-15. Total number of items on the RCBC? Could there be subscales identified within the 10 categories of behaviors? Then, you would probably have an increase internal consistency at least for the subscales.

We appreciate this suggestion but each of the 10 items on the RCBC is a single item that reflects a care activity so that subscales would not be appropriate. Moreover, there is support, based on Rasch analysis model testing, to indicate that the RCBC is a unidimensional measure.

11. Figure 1. I would suggest putting only the 8 paths that were statistically significant rather than trying to depict visually all 58 hypothesized paths. It is very difficult to follow as it is currently designed.

We apologize Figure 1 was difficult to read-we were trying to make it look pretty and have all the arrow tips converge. We have revised Figure 1 to include only the 9 paths that were statistically significant when the revised model was tested (see responses to the Stat Reviewer below).

Reviewer #3: Stat reviewer

1. This review is focused on measurement and statistical analysis. In general, the analysis was described well and only some minor suggestions for clarity are listed below, with the exception of the AMOS model, which needs clarification that assumptions were checked, that the sample was large enough, and that the results make sense. Figure 1 cannot be read as there are too many paths. In addition, some suggestions for additional issues to include in discussion are presented below.

We have indicated in the manuscript how the assumptions of structural equation modeling were addressed and provided support for the sample size utilized. It should be noted that we, as suggested by the stat reviewer, decreased the size of our initial

hypothesized model and included a more parsimonious model with only 38 hypothesized paths. The data analysis section now reads:

A sample size of 386 was considered sufficient for model testing of the hypothesized 38 parameter model (Bollen, 1989; Kline, 2004). With the exception of gender all variables were normally distributed or, in the case of self-efficacy and outcome expectations, transformed using a reflect and logarithm transformation for a negative skew. The sample size in this study was not sufficient to perform a multigroup analysis and for males and females to be analyzed separately. Therefore, as is commonly done, gender was analyzed as if it was a continuous variable (Der, 2002).

2. P 10 L 9: It would be useful to present a brief rationale for the predictors of age, gender, and ethnicity as concepts that would influence knowledge of restorative care, self-efficacy, and outcome expectations. In particular, it seems odd to consider gender as a potential predictor of these. Both age and ethnicity had little variation in the data and thus may not meet the assumptions for a predictor variable in the AMOS analysis. This is a minor weakness that could be addressed in a couple of sentences: one to clarify the rationale in this section, and another to state that assumptions for AMOS analysis were met in terms of variability, normality of the variables. See, for example, Kline p 84 who states that each group within a grouped exogenous variable must be screened separately to be sure it meets the assumptions. (Rex Kline, Principles and Practice of Structural Equation Modeling.)

We have added a section into the manuscript to provide the background rationale for the inclusion of age, gender, education and job attitude into the model. Ethnicity was actually not included in the model as there is not consistent evidence to support a relationship between ethnicity and the outcomes (self-efficacy and outcome expectations related to restorative care and restorative care behaviors). The section added is as follows:

#### Additional Factors that Influence Type and Quality of Nursing Care

In addition to self-efficacy and outcome expectations related to caregiving activities, there is some evidence that gender influences the type and quality of nursing care provided to older adults. Repeatedly it has been noted that males tend to have more negative attitudes toward older individuals (Gallagher, Bennett & Halford, 2006; Soderhamn, Lindencrona & Gustavsson, 2001) and thereby are less willing to engage older individuals in health promoting activities. Male nurses also tend to feel less confident in their nursing skills, which may further influence caregiving. Males, however, are more likely to learn new information related to caregiving (Tabet et al., 2006). Years of education was also noted to influence attitudes toward older adults such that those with more education were more likely to have positive attitudes toward older individuals (Gallagher et al., 2006). These positive attitudes can influence the type of care provided (Latimer & Thornlow, 2006; Varkey, Chutka & Lesnick, 2006).

3. P 11 L9: This description of the RCBC items is confusing. Suggest calling "interactions" in line 8 "activities" instead. This may clarify the meaning of "Activity not performed..." in line 9. As it reads now, it is not clear whether activity refers to a restorative care activity or a resident activity. In addition, the use of the term "restorative care activities", as in line 1 of page 12, adds to the unclear distinction between the activity done between NA and resident and the restorative care intervention (or behavior or interaction).

We greatly appreciate this request for clarification and recognize that we were used many different terms that could be confusing. The manuscript has been revised such that we use restorative care activities consistently throughout. Quite simply the RCBC includes 10 restorative care activities.

4. P 13 L10: From the wording in this line, the reader expects to see a Person Separation score and a Person Separation Reliability score. There are two scores in the results, one of which is called the Person Reliability score. Clarifying these terms and using them consistently would strengthen the manuscript and improve understanding of the statistical analysis. In this paragraph, provide the meaning or range of potential scores. Without this information, the result on p 18 that score was 6.63 cannot be interpreted. The score on p 18 for "person reliability" of .77 is presented in the next paragraph as the score for "person separation".

We have revised the manuscript to include only the Person Separation Reliability score. The Person separation score is the score used to calculate the person separation reliability score so it is not critical that both be included. The Person separation reliability score is the score that is analogous to the alpha coefficient.

5. P 14: The INFIT and OUTFIT description is very clear. P 16 L19: Clarify the meaning of "mapping" which was not described in the analysis section.

We have included a sentence in the data analysis section to describe mapping. It is now stated:

Lastly, item mapping is described to provide empirical information about item hierarchy with regard to which activities the NAs are most likely to perform and which they are least likely to perform (Smith & Smith, 2004).

6. P 17: Add a justification for testing such a complex model of 58 parameters with a sample of 386 subjects. Kline (see above) presents a rule of thumb of minimum of 10 subjects per parameter estimated with 5:1 making a model unstable. This would suggest that 38 parameters was the limit. Authors should assure the reader that assumptions for AMOS model were tested (see comment on variability above) and met. There is little discussion about the theoretical rationale for the large number of proposed paths in the model, and perhaps a more parsimonious model could be presented as the initial model, especially as the initial model did not produce good results. Figure 1 is impossible to read with so many intersecting paths. A simpler model, with rationale, would be easier to

present. Alternatively, present assurance that the assumptions for the data were considered and that the sample size was adequate. Presenting the results from page 17 in either a figure or a table would facilitate understanding the significant pathways described in the text.

We greatly appreciate this suggestion and agree wholeheartedly. We have revised the manuscript to reflect the testing of a model that includes 38 parameters. As noted we also revised Figure 1 to include the significant paths of that model only. We would be happy to include a model that shows all of the hypothesized paths as well if the reviewers feel this would be helpful.

7. P 17 L19 State which model is referred to, the original or the trimmed model.

We have revised the manuscript to indicate that in this section we are referring to the trimmed model.

8. P 18 L 15, add a brief explanation of why a wide range of behaviors would influence the score of .77.

We revised the text to explain that there may be a difference between providing restorative care related to communication versus restorative care for walking, bathing or dressing. This wide range of activities can impact reliability testing.

9. Regarding the findings, a weakness of the study that should be noted was the lack of observations of eating, use of assistive devices, and exercise (87-93% not observed) of observations did not include these activities. One would expect that these would be prime examples of opportunities for restorative care interventions and the lack of data should be noted as unfortunate and perhaps a reason given for their absence in terms of methods of the study.

We have revised the limitations section of the manuscript to address the lack of observations of eating, use of assistive devices and exercise and indicated plans for future research to attempt to observe more of these activities.

10. The discussion should deal with the issue of gender being the only predictor of engaging in restorative care. First, gender had almost no variability so it should be acknowledged that this is not a particularly useful finding. Second, the fact that none of the hypothesized predictors had an influence on engaging in restorative care should be acknowledged as unexpected and some reasons why posited. This could be tied in with the social ecological model paragraph. In addition, the finding that knowledge and self-efficacy led to outcome expectations, but not to actual restorative care behavior should be mentioned in the discussion along with possible reasons for this unexpected finding.

The discussion has been revised to address the impact of gender, in light of the minimal variability of this factor. The revised model did support an indirect relationship between

self-efficacy and restorative care behaviors. We also discuss, however, why the hypothesized relationships might not have been supported.

11. In regards to findings on validity testing on p 16, where exercise and eating are included with the most common activities in which NAs perform restorative care, these findings should be acknowledged as not useful when so few of these activities were observed, as shown in Table 3.

We have acknowledged in the limitations section that the findings, particularly with regard to internal consistency and item mapping, may be influenced by the limited number of observations of exercise and eating we actually observed during the study. We believe that it is critical to increase the time of observations to be able to observe more of these types of activities prior to determining that they should be removed from the measure.

---

#### CHECKLIST FOR STYLE

Title Page

Supply affiliation and professional title for each author.

Affiliations have been added to the title page.

References

Other: The Chang & Lin, 2005 reference does not match the in-text citation of Chang, 2005. Please correct.

This has been corrected.