

Author's response to reviews

Title: Assessing an organizational culture instrument based on the Competing Values Framework:
Exploratory and confirmatory factor analyses

Authors:

Christian D Helfrich (christian.helfrich@va.gov)

Yu-Fang Li (YuFang.Li@va.gov)

David C Mohr (david.mohr2@va.gov)

Mark Meterko (mark.meterko@va.gov)

Anne Sales (ann.sales@va.gov)

Version: 2 **Date:** 11 January 2007

Author's response to reviews: see over



DEPARTMENT OF VETERANS AFFAIRS
Puget Sound Health Care System
1660 South Columbian Way Seattle, WA 98108-1597

Dear Dr. Mittman,

Thank you for the excellent feedback and for the opportunity to revise and resubmit our manuscript, *Assessing an organizational culture instrument based on the Competing Values Framework* (#3697413001077852). We believe it is much improved and hope that you will find it so.

Attached on separate pages we have addressed each reviewer's specific comments. I will briefly summarize the two most significant changes.

First, we re-organized and re-wrote our Discussion section based on major validity types (external, internal and construct) in order to better address the internal validity issues raised by Dr. Whalley regarding the modifications to the VHA instrument.

Second, we have confined our analysis to the non-supervisory population following a measurement equivalence / invariance analysis, as suggested by Dr. Etchegaray. We decided not to report the results of the ME/I in the current manuscript because we felt it took us away from our major point, to test the psychometric properties of this instrument and the validity of the CVF subscale structure among non-managers. We also describe supplementary analyses to address potential concerns of further, unidentified heterogeneous subgroups.

These and other, less extensive revisions are detailed on the attached pages. One other minor modification that I will note here is that we refer to "team" culture instead of "group" culture, in order to be consistent with the terminology used in most publications.

Again, we appreciate the opportunity to resubmit our manuscript, and we will be happy to respond to any additional questions or concerns.

Sincerely,

Christian D. Helfrich, MPH, PhD

Attachments

Christian D. Helfrich, MPH, PhD
Implementation Research Coordinator
Ischemic Heart Disease Quality Enhancement Research Initiative
Tel. 206-277-1655
Fax. 206-764-2935
christian.helfrich@va.gov

Response to Reviewer A
Manuscript 3697413001077852

Dr. Whalley raised five major compulsory revisions and three discretionary ones. These highlighted both the need to clarify our language about the analyses, and prompted us to complete some additional analysis.

Our responses to specific reviewer comments follow. We include the reviewers' text followed by our response.

Major Compulsory Revisions

1. There should be more explanation and discussion of the changes that were made to the original CVF instrument. While I would agree in principal with the assertion that a "robust conceptual framework should not be predicated on a single instrument" (page 19), it nonetheless remains that the operationalisation of a conceptual framework can only be as good as the measurement indicators used. Thus, the changes made to the original instrument could have considerable implications for interpreting the findings in relation to the framework. Taking the change illustrated on page 19 as an example, it could be argued that the revised item portrays a picture of collaborative facilitation rather than the hard-driver leadership of the original. Such a change in meaning could alter the way responses align with other items in the instrument, an issue potentially further confounded by the fact that the rational subscale had already been reduced to just three items.

We agreed and have added an extensive discussion of the modification of items (pg. 22), including clearly delineating the types of modifications made (changes to item wording, elimination of two items and use of Likert versus ipsative scales), and cataloguing which items were modified from the version used by Zammuto and Krakower and what those modifications were. We accordingly re-organized the Discussion section by major type of validity (internal, external, construct) in order to better indicate where potential limitations could arise in our study.

We also provide rationales for why the item modifications likely do not account for the problems we observed with the convergent / divergent properties of the CVF subscales in this dataset. For example, we argue that 4 items had potentially significant modifications and that 3 of these were in the Hierarchical subscale; this might then explain the Hierarchical scale's relatively poor reliability, but it wouldn't explain why the remaining three subscales failed to diverge (pg. 22). We also note that items that were unaltered (but for the use of "facility" rather than "institution") had similar high cross-subscale correlations to those that had been modified in the VHA instrument (pg. 23), suggesting that the high cross-subscale correlations were not the result of the item modifications.

Finally, we briefly touch on some potential issues with item wording that owe to the original instrument and were carried over into the VHA instrument (pg. 23), notably the

use of terms such as bureaucratic and innovative, which may carry unintended normative connotations, and the use of item that include multiple statements, to which respondents may have divergent responses but are obliged to answer as a whole.

2. Internal consistency is commonly defined as the consistency with which individuals respond to the items within a scale. While the evaluation of inter-item correlations within a scale is one possible assessment, the analysis described focuses more on the convergent/divergent properties of the items. This approach leads to seemingly counterintuitive findings; for example, the hierarchical subscale is reported to have better internal consistency than the other scales based on the higher internal correlations, even though its Cronbach's alpha coefficient and item-rest correlations (which are the mainstay of internal consistency evaluation) are the lowest of all the scales. The inter-item correlations results reported are more indicative of item discriminant validity than of internal consistency. Indeed, I think the issue of item and scale validity is more in tune with the point the paper is trying to make, although I would suggest that using item-to-scale correlations would make a more parsimonious assessment.

We appreciate this clarification and suggestion. We had been using scale reliability to refer to internal consistency and convergent / divergent properties of the subscales. We have revised our text to distinguish between the two. We have also added item-to-scale correlations as a comparison for the item-rest correlations, which we agree helps make our primary point about item and scale validity.

3. Cronbach's alpha is not an index of unidimensionality and there is a considerable literature to this effect (see for example Green et al 1977; Cortina 1993; Schmitt 1996; Shevlin et al 2000). Thus, it is perfectly possible to get a high alpha coefficient from a multidimensional scale, and indeed a comparatively low alpha with a unidimensional scale. Of particular note is the fact that alpha is influenced by the number of items in a scale. This has important implications when comparing the 12-item humanistic culture scale and the shorter 3-4 item scales. It raises the question as to whether the coefficient of 0.68 for the hierarchical scale is actually as poor as it might seem, particularly as the item-rest correlations are all well above the 0.20 threshold. It certainly means that shortening the instrument may indeed have altered its psychometric properties, at least in relation to Cronbach's alpha. Given this and point 2 above, I do not think that the scale reliability results necessarily indicate a fundamental problem with the traditional CVF structure; rather they suggest the possibility for a simplified structure for the instrument.

We appreciate this correction and have deleted the reference to Cronbach's alpha as a measure of unidimensionality. We have also noted in our Methods (pg. 11) and Results (pg. 15) that the low alpha for the hierarchical scale could equally reflect poor reliability or too few items in the subscale.

4. I am unclear as to why both of the two additional items in the 14-item two-factor confirmatory model were set to load onto both factors. Assuming these two items are

Christian D. Helfrich, MPH, PhD
Implementation Research Coordinator
Ischemic Heart Disease Quality Enhancement Research Initiative
Tel. 206-277-1655
Fax. 206-764-2935
christian.helfrich@va.gov

numbers 13 and 10, only item 10 was indicated as cross-loading in the exploratory analysis.

This was a miscommunication on our part about which EFA model the cross-loadings were based on. In any case, in the revised paper we elect to drop item 13 from the two-factor model due to its low factor loading. Item 10 is the only cross-loaded item (pg 17).

5. The results presented indicate, as the authors suggest, that a two-factor dimensional structure could be a viable and perhaps simplified alternative to the traditional CVF for this version of the instrument. However, given my above comments and the fact that differences in the fit of the confirmatory factor models was somewhat marginal (and still well above the recommended thresholds for the traditional four-factor model), I do not think that the study "raises questions about the aggregation of items from a popular survey instrument based on the CVF" (page 21).

Dr. Whalley raises excellent points. The emergent model has only marginally better fit than the traditional CVF, and there are important potential limitations to our data, notably the modifications to the instrument that was used to collect the data. As a result, we toned down the language regarding the two-factor model as an alternative, and emphasized that our findings illustrate the need to validate instruments in each new setting and population in which they are used.

At the same time, we feel that our results raise non-trivial concerns about the CVF. The convergent / divergent properties of the entrepreneurial, group and rational subscales are not consistent with the CVF, and in our data these subscales are highly collinear, suggesting that they might not represent distinct constructs.

We agree that the modifications to the VHA instrument may have altered the psychometric properties of the instrument, but we argue that this most logically affected the hierarchical scale (pg. 24). There is only one item from the rational, group and entrepreneurial scales that we feel was significantly modified. Yet, two thirds of item-to-scale correlations for these three scales were in excess of item-rest correlations. We also note that items virtually unaltered from the Zammuto and Krakower instrument still have cross-subscale correlations in excess of within-subscale correlations.

Given the widespread use of the CVF instrument in health services research, and the way that it is generally presented as well-validated, reliable and generalizable, we think it is appropriate for us to strike a note of caution. We have also revised our Discussion and Conclusions sections to emphasize that more research is needed on the CVF, and that there needs to be validation in each new setting it is used.

Discretionary Revisions

1. While a threshold of 0.40 for factor loadings is undoubtedly a commonly cited cutoff (and thus perfectly justified), I would argue that it might be a little stringent. It would

Christian D. Helfrich, MPH, PhD
Implementation Research Coordinator
Ischemic Heart Disease Quality Enhancement Research Initiative
Tel. 206-277-1655
Fax. 206-764-2935
christian.helfrich@va.gov

certainly be useful to see if the loadings of items 10 and 13 increased if the exploratory factor analysis was re-run specifying the extraction of 2 factors.

We appreciate this suggestion, but elected to keep the threshold of 0.40. In the original analysis, we used three criteria for retaining factors, including eigenvalues equal to or greater than 1. This analysis retained 2 factors with eigenvalues greater than 1. We also repeated the exploratory factoring analysis but specifying the extraction of 2 and 4 factors. Factor loading of item 10 and 13 were unchanged from the original model and that specifying 2 factors after promax rotation: 0.48 and 0.39 on Factors 1 and 2, respectively for item 10, and 0.15 and 0.36 for item 13. Factor loadings from specifying the extraction of 4 factors were primarily loaded on 2 factors, following the same pattern in the original analysis and with slightly lower but salient factor loadings on these 2 factors.

2. The authors might consider dropping the confirmatory factor models with uncorrelated factors, as the argument for correlated factors was already well-made in relation to the exploratory analysis.

Thank you, we agreed and have done so.

3. An alternative approach to the confirmatory factor analysis of the 14-item two-factor model would be to start with no cross-loadings and use the modification indices in LISREL to see if such additional parameters were indicated. This would have the added benefit of allowing equivalent versions (i.e. 14 items with no cross-loadings) of the 2-and 4-factor models to be compared.

Thank you for this recommendation, and we have reanalyzed our data accordingly. In this revision, for the alternative models, we conducted confirmatory factor analysis starting with a 13-item, 2-factor, no cross-loading model as indicated by the exploratory factor analysis. The largest modification index was for the path from the prescriptive culture to item 10. This indicated that we could expect an improvement in model fit by including this path in the model. This is consistent with findings from exploratory factor analysis where the factor loading of item 10 was close to 0.40 on the prescriptive culture subscale. Following these findings, we then tested three alternative solutions for the two factor model. Details of the analysis and findings are presented on pages 20 and 21, and Table 3.

Response to Reviewer B
Manuscript 3697413001077852

Dr. Etchegaray proposed two major and one minor compulsory revisions, the principal one being a call to do measurement equivalence / invariance analysis based on respondent supervisory level. We have revised the analysis and manuscript based on these suggestions, which we describe in detail below.

Major Compulsory Revisions

- 1) Given the large sample size, the analysis should have been conducted based on type of respondent - perhaps based on supervisory level - so that you can determine whether the results apply similarly to all levels. A measurement equivalence / invariance approach needs to be applied (see Vandenberg & Lance, 2000) to determine the robustness of the findings across these levels.

Thank for this important question and suggestion. In response, we have conducted a measurement equivalence/invariance test between high-level managers and the rest of organizational members. Results from this analysis did not support invariance of factor structures or loadings across groups. Therefore in this revision, we have limited our analyses to frontline employees, since they represent a population that the CVF has not previously been validated in.

We recognize that with 70k frontline employees, there is the continued threat of unidentified subgroups for whom there is measurement non-equivalence. Therefore, we ran bootstrap analyses to determine if the factor loadings and sample means differed across 1,000 100% bootstrap samples (these analyses are available upon request). We contend that if there were a significant unidentified subgroup for whom the CVF data would suggest a different factor structure, we would expect to see significant differences in factor loadings and co-variance matrices among bootstrap samples. For the models tested in this study, we evaluated the distribution of the covariances been fitted, goodness of fit indices and factor loadings. We observed virtually identical factor loadings among the bootstrap samples and propose this is a reasonable indicator of respondent homogeneity among the remaining sample of non-supervisors.

- 2) The discussion of poor internal consistency with respect to the scales was a good start but did not note that one reason why internal consistency might have suffered was due to the fact that the items within each subscale were inconsistent with each other from a content perspective. That is, the translation of the items from the original CVF measure resulted in items within each subscale that are likely to receive different ratings from the participants - when that happens, internal consistency will suffer. While there is nothing you can do to fix this issue now (but hopefully will make

some adjustments in the future), you should note this so that the readers understand why internal consistency might have suffered.

Thank you, we have added a brief note about the differences in content areas addressed by items within the same subscale to identify this as a potential cause of the high item correlations across subscales (pg. 24).

However, the different content perspectives of the items was an intentional variation of the originators of the CVF instrument (not a function of the translation). They anticipated that organizational culture should be evident across major organizational domains. This relates to one of the major theoretical assumptions of the CVF: that culture is an organizational-level construct and a pervasive, common influence.

Minor Essential Revisions

- 1) The CFA results were unclear based on the table presented. What were the chi-square values, degrees of freedom, and chi-square to degrees of freedom ratio? The best model needs to be selected based on parsimony and chi-square, not on the other fit values. The results for the CFA, as presented currently, are confusing to the reader and the rationale for how to select the best model is lacking (and inconsistent with my point in the previous sentence).

We have noted the chi-squares and degrees of freedom in the text (pg. 19) and have added them to Table 4.