Author's response to reviews

Title: The Safety Attitudes Questionnaire: Psychometric Properties, Benchmarking Data, and Emerging Research

Authors:

John B Sexton (jsexton2@jhmi.edu)
Robert L Helmreich (helmreich@psy.utexas.edu)
Torsten B Neilands (tneilands@psg.ucsf.edu)
Kathy Rowan (kathy@icnarc.org)
Keryn Vella (keryn.vella@icnarc.org)
James Boyden (james.boyden@rchft.cornwall.nhs.uk)
Peter R Roberts (prroberts@xtra.co.nz)
Eric J Thomas (Eric.Thomas@uth.tmc.edu)

Version: 2 Date: 18 January 2006

Author's response to reviews:

Dear Editor,

Thank you for your request to revise the manuscript and we thank the reviewer for the thoughtful comments. All comments are addressed below. The paper is improved and we hope it is acceptable for publication in BMC Health Services Research.

General Comments:

1. The reviewer notes that the manuscript and survey lacks a clear theoretical framework and explanation of the aviation version (FMAQ). We thank the reviewer for giving us the opportunity to clarify these important points. To help explain this we have created a new subsection in methods called "Lineage and conceptual background of the Safety Attitudes Questionnaire (SAQ)".

The first paragraph of this section (pages 5-6) now tells the readers that the FMAQ was developed based upon the analysis of airline accidents. We now say, "The FMAQ was created after researchers found that most airline accidents were due to breakdowns in interpersonal aspects of crew performance such as teamwork, leadership, communication, and decision making. The FMAQ measures crew member attitudes about these topics."

The process of translating the FMAQ into the SAQ was guided by two theoretical frameworks/conceptual models: Vincent's framework for analyzing risk and safety and Donabedian's model of quality. We now clarify this by saying, "We relied upon two conceptual models to decide which items to include: Vincent's framework for analyzing risk and safety (8) and Donabedian's conceptual model for assessing quality (18). This generated a pool of over 100 new items covering four themes: safety climate, teamwork climate, stress recognition, and organizational climate." These themes are explicitly noted in Vincent's model and are related to Donabedian's structure-process-outcomes model for assessing quality.

The reviewer asks how job satisfaction became part of the SAQ. There were some job satisfaction items on the FMAQ. They were initially part of our scale "organizational climate". The results of subsequent factor analyses led us to divide this scale into perceptions of management, working conditions, and job satisfaction (this is noted on page 6). In addition to being part of the FMAQ, we considered job satisfaction to be consistent with Vincent's "work environment" domain (although he does not specifically mention job satisfaction in the article we cite).

We used the term "focus groups" in the original version but have deleted it in this version. Instead, we say that we had discussions with healthcare providers and subject matter experts. For some readers, "focus groups" will imply that we audiorecorded, transcribed, and formally analyzed the data. We did not do that. The providers and subject matter experts were asked to help us identify factors that, in the reviewer's words, "influence patient's safety". We ended up with items that were both "pre-conditions" and "consequences" of safety. We wanted to be inclusive during these conversations. Subsequent inclusion of items was directed by Vincent's and Donabedian's models and our factor analyses.
2. The reviewer makes an important point regarding measurement and definition of culture. Throughout the paper we use two terms, safety climate and safety culture. The beginning of the paper makes it clear that this survey does not measure all aspects of culture. For example, the first sentence of the second paragraph in the introduction says, "Healthcare provider attitudes about these and related factors are one component of an organization's safety culture." However, later in the paper we do use the term safety culture. Unfortunately, this mixed message is consistent with the use of these terms in the literature and in daily practice by funding agencies (eg AHRQ) and leaders in healthcare quality (eg Institute for Healthcare Improvement). Nevertheless, to avoid confusion in the revised manuscript we have decided to use the term safety climate instead of culture and we now define these terms in the Methods section. We have created a new subsection titled "terminology" (top of page 5).

3. The factor analysis is circular and not based upon a theoretical model. As noted above, we did rely upon two theoretical models to identify the initial themes, and generated a hybrid framework of risk and quality to guide our refinement efforts. There was not just a list 100 items chosen without theoretical guidance. We think our comments and revisions related to item 1 are also responsive to this item.

4. The reviewer says it is unorthodox to use CFA to help identify items to be removed. Deletion of items was guided by both the CFA results and the theoretical models noted above. Again, our clarification of the theoretical model (item 1) also responds to this comment.

5. More explanation and justification is needed about the individual versus area level analyses. Thank you for highlighting this issue. We added language in the discussion (pages 15-16) regarding clinical area level analyses:

   The SAQ is a psychometrically sound instrument for assessing six safety-related climate domains by systematically eliciting input from front-line caregivers. Evidence here and elsewhere [28] indicates that researchers should consider hospitals comprised of clinical areas to resemble corporations comprised of organizations, because the clinical areas appear to resemble what are typically considered organization-like unique climates. The multi-level model demonstrated that there is more variability between clinical areas than within clinical areas. In other words, context of care assessments appear to be more robust, meaningful, and interpretable at the clinical area level. Climate at the clinical area level is important as many clinical and operational outcomes are tracked at the clinical area level (e.g., catheter related blood stream infections in a specific intensive care unit), and it is easier to target clinical area level improvements than hospital wide improvements. We see the focus on clinical area level climate as a way to acknowledge the complexity of the systems in which we work, rather than assuming monolithic hospital climates that lack diagnosticity of clinical area level issues.

   See also item 10 below.

6. The reviewer asks us to reduce the speculative content of discussion and remove promotional text.

The following promotional text was deleted, "The SAQ can measure safety climate, and can be used to measure the effectiveness of interventions to improve patient safety. For example, in the US large healthcare organizations like Kaiser Permanente, Ascension Health, and Memorial Hermann Healthcare System (totaling over 100 hospitals) use the SAQ hospital-wide. The Volunteer Hospital Association uses the SAQ hospital-wide in regional collaboratives and in their Transforming the ICU program. The Institute for Healthcare Improvement advocates use of the safety climate scale from the SAQ as an organizational benchmark prior to quality improvement initiatives (32)." We also deleted the last paragraph, "To date there are over 175,000 copies of the SAQ in circulation, and many of these are being used in longitudinal quality of care investigations. As our understanding of healthcare climates and contextual factors evolves, we become better equipped in our quest to improve quality of care in medicine."

We still describe the use of the SAQ in other studies, but the purpose is to convey to readers how the survey is being used by others, and provide some data about convergent and construct validity.

The reviewer said that he hoped the discussion will state how the SAQ relates to other work on safety attitudes. There was a full paragraph on this topic in the original draft of the manuscript and it remains in the revised version (page 15, first full para). Please let us know what other information would be helpful.
7. The term "executive walkrounds" has been deleted.

8. Safety culture is now defined on page 5 (see response to item 2)

9. We have clarified background of the FMAQ and rewritten the sentences about focus groups (see response to item 1 for details).

10. page 5 - explain and justify respondent and area level analysis. In our Terminology section of the methods, we explain: The SAQ elicits a snapshot of the deeper safety culture through surveys of frontline worker perceptions. When using questionnaires to study group-level perceptions, the most appropriate term to use is climate (e.g., safety climate, or teamwork climate). Here we use clinical areas (a.k.a., work units, patient care areas, nursing units) as the group-level of interest. By analyzing the psychometrics of the SAQ at the individual level and the clinical area level, we can test the appropriateness of conceptualizing patient safety issues at the clinical area level, because clinical areas are generally associated with managers, geographical locations, and specific clinical and operational outcomes.

See also item 5 above.

The reviewer provided helpful suggestions for parsing down the results presented here. As the original psychometric paper for the SAQ, we see this manuscript as necessarily detailed, so that other researchers have the ability to evaluate and replicate our approach. Future manuscripts should not need to go into the detail used here, which will make them more readily interpretable and accessible to caregivers and researchers who are unfamiliar and uninterested in the psychometric standards and tests we use for this paper.

11. page 9, explain SEM based alphas.

Coefficient alpha is a biased estimator of composite reliability because it makes the restrictive assumption that all loadings for a given factor are equal. Raykov's latent variable-based approach does not make this assumption, yielding a better estimate of the true composite reliability of a scale. We have clarified this point in our explanation of our selection of Raykov's rho coefficient to estimate composite reliability for the SAQ (last paragraph on page 10).

12. The comments about reliability in the results section have been deleted.

13. The reviewer stated "you report that percent that agreed or disagreed with an item ranged from 0 to 50%, or 9 to 100%, within an area. What does that mean? The clinical area is the same, yet people cannot agree."
The original phrasing was poor, so we simplified it to, "In total, for example, one out of five respondents reported that it is difficult to speak up if they perceive a problem with patient care, but at the clinical area level, the percent of respondents who agree ranged from 0% to 50%." In other words, zero respondents reported difficulty speaking up in some clinical areas, while in other clinical areas, half of the caregivers reported difficulty speaking up. See top of page 12.

14. The link now works. It should have been ".org", not ".edu"

15. Table 2: are the SDs for area means, or for individual measurements? Thank you for this catch - your attention to detail is very much appreciated. We clarified this in the table title: Table 2. SAQ Factor Descriptives: Overall Means (Minimum and Maximum); and Overall Standard Deviations (Overall score sample sizes come from n returned in Table 1)

16. The reviewer questioned the usefulness of our between-area alpha coefficients. The reviewer also commented in general that the amount of results presented was substantial, almost too much. The computation of composite reliability for complex analysis scenarios such as the multilevel factor analysis model is still evolving. At the time we submitted our original manuscript for review, there was no literature specifically addressing the computation of composite reliability in multilevel designs, so we applied the existing methods for individual level analysis to each subscale within each level of our study. In hindsight, we agree that this is too much information to present to the reader. Readers will typically use a scale in its entirety and want to know if the scale is reliable before using it. Readers are usually not interested in the reliability estimates of isolated subscales removed from the greater context of the overall scale.

We are pleased to report that the December 2005 issue of Structural Equation Modeling published an
article by Raykov and du Toit that demonstrates how to compute a composite reliability index value for an entire scale in multilevel designs. Using their method, we computed the composite reliability value for the SAQ and have reported it in the revised manuscript (page 13). We believe this coefficient, which applies to the whole scale rather than to specific subscales, more fully captures the spirit in which the SAQ will be used. That is, it will be most often used as a whole scale rather than as a collection of subscales. As an added benefit, switching to the composite reliability index allows us to eliminate an entire table from the resubmitted version of the manuscript, thereby reducing the number processing burden of the reader.

17. The reviewer commented that Appendix A, the table that displayed factor loadings, was not sufficiently labeled so that the reader could determine whether the results presented were factor loadings or path coefficients. The reviewer also questioned the need for unstandardized coefficients and confidence intervals. This comment is consistent with the reviewer's earlier, more general point that too many numerical results are presented, many without comment. In the revised manuscript we strove to reduce the amount of numerical results presented, retaining the most important results, simplifying their presentation, and offering more extensive commentary on the remaining results. In this spirit, we removed Appendix A from the revised manuscript and now report standardized factor loadings at the between-area and within-area levels in Table 3. We have dropped the unstandardized loadings and confidence intervals, which we judged to be of less use to most readers. With respect to Table 4, we added:
Examination of Table 4 shows that the 6 factors have lower correlations at the clinical area level than at the individual respondent level, indicating that the 6 factors are more diagnostic (share less variance with each other) when used at the clinical area level.

18. The reviewer raised similar questions about Appendix B, which displayed the interfactor covariances, confidence intervals of covariances, and correlations. Consistent with our approach outlined in response to comment 17, we eliminated the unstandardized covariances and confidence intervals from the table. It now displays factor correlations at the between-area and within-area levels. We have recast this table as Table 4 rather than an appendix because a table of correlations is more properly treated as a table rather than an appendix.

19. The reviewer commented that Figure 1 is unnecessary because all survey items appear in Table 3. We have found that it is useful to have scale definitions for readers and we would prefer to leave Figure 1 in the manuscript. However, by following the reviewer's other suggestions, we have eliminated two appendices saving considerable space in the new manuscript draft and, hopefully, clarifying the important results that remain.

20. The reviewer requested an explanation of the Mplus MLR test statistic be added to the data analysis section of the manuscript and that MLR be spelled out. We have added an explanation of MLR and spelled it out (page 9).

21. Regarding floor/ceiling effects, we clarified their use:
Response rates and floor/ceiling effects for each scale are presented in Table 1 by administration.

22. The reviewer suggested that percentages be rounded off in Table 3, which we have done. Also, the reviewer suggested that we eliminate the skewness and kurtosis information as these distribution statistics are not of interest to most readers. We agree, and have replaced the skewness and kurtosis columns with standardized between-area and within-area factor loading values, respectively.

23. No changes requested.