Author's response to reviews

Title: Assessing the Safety Attitudes Questionnaire (SAQ), German language Version in Swiss University Hospitals - A Validation study

Authors:

Natalie Zimmermann (natalie.zimmermann76@gmail.com)
Kaspar Küng (kaspar.kueng@insel.ch)
Sandra Engberg (sje1@pitt.edu)
Susan Sereika (ssereika@pitt.edu)
Bryan Sexton (bryan.sexton@duke.edu)
Rene Schwendimann (Rene.Schwendimann@unibas.ch)

Version: 2 Date: 12 July 2012

Author's response to reviews: see over
July 12, 2012

MS: 1342279176701131: “Assessing the Safety Attitudes Questionnaire (SAQ), German language Version in Swiss University Hospitals - A Validation study

Dear editor

We thank you for your message and the thoughtful and reasonable comments from the two reviewers which helped us to improve our manuscript. Please find below our point-by-point responses corresponding with the changes made (yellow marked) in the manuscript.

Sincerely yours
René Schwendimann, corresponding author

Editor’s comments

Please make the following formatting changes during revision of your manuscript. Ensuring that the manuscript meets the journal’s manuscript structure will help to speed the production process if your manuscript is accepted for publication.

1. Competing interests: Manuscripts should include a “Competing interests” section. This should be placed after the Conclusions/Abbreviations. Please consider the following questions and include a declaration of competing interests in your manuscript: Financial competing interests and Non-financial competing interests
Response: We integrated the “Competing interest section” and declared neither financial nor non-financial interests.

Comments of reviewer 1 (JE)

Major compulsory revisions

1) On page 5, a retest of the survey is mentioned. This retest is not mentioned in the abstract nor is the rationale for the retest explained to the reader on page 8. The purpose of the retest needs more discussion. Also, why 6 months between test and retest? Is this an appropriate timeframe?

Response: Retest was mentioned in the abstract as well as the methods, results and conclusions. We agree that the purpose of this test should be more explained in the article. Patient safety culture (PSC) should be a fairly sustained characteristic of units. The idea of doing a test-retest analysis within a time frame of six months was to gather information to determine whether the PSQ can be measured by the SAQ over a sustained period of time. This is important for further investigations about PSC and its relation to adverse events with low incidence rates (exp. falls, nosocomial infections...).
Accordingly, we added additional information about implications of the retest results in the abstract’s
conclusion section and our rationale for performing this analysis was added in the background section of the manuscript at the end of the subheading “measurement of PSC”.

2) For the four items mentioned with low I-CVI levels, what does “no further tests were not carried out”? Does this mean that they were excluded from subsequent analysis? If not, why were they included?

Response: We discussed the four items with low I-CVIs and decided to keep two (15/17) because their values were close to the recommended cutpoint and one (22) was kept despite its low I-CVI because it showed high loading to its respective factor “stress recognition”, a factor which arose in the analysis as a strong unitary factor, detached from all other factors. The fourth item (24) was dropped because it was deemed to be unclear to many participants. Content validity was not re-examined after this item was dropped. This section was revised in the manuscript.

3) In the Internal Structure section, while I understand why you collapsed into three response categories, I am unsure why you did not try to collapse into disagree/neutral/agree. It would be helpful for the reader to see the % breakdown of these five categories so he/she can verify for him/herself that your collapsing is warranted.

Response: We also tried also to collapse into disagree/neutral/agree, but unfortunately, the problem was the same as before: There were not enough responses in the category disagree. The only way to meet the criteria for CFA was to combine the responses as we did. In table 3, readers can see the % of answers “Disagree (categories disagree strongly and disagree slightly). In our opinion, this will be a good indicator for the warranty of the collapse.

4) The collapsed response categories in the Internal Structure section are problematic for the analysis you are conducting. It is one issue to argue that your original five response choice scale could be conceptualized as interval yet an entirely different issue that your collapsed three choices scale could be conceptualized in the same way. Your collapsed scale is clearly ordinal and as a result requires estimation different than ML when conducting the CFA. The authors are encouraged to read Flora and Curran’s 2004 Psychological Methods paper that helps address CFA with ordinal data. Given that I think you need to re-run the analysis using a different approach (as Flora and Curran describe), I cannot evaluate your results because they might change based on this new analysis.

Response: We appreciate this concern. We are aware of the fact that we are working with ordinal data when collapsing the original Likert scale into three point Likert scale. This was in deed a major challenge for our analyses. Opinions regarding the right way to carry out CFA are highly different. We used MPlus for data analysis because this software package is seen as an appropriate method for CFA with ordinal data. We also agree that CFA for ordinal data should not be done using Maximum Likelihood estimation (ML). This is the reason, why we have chosen to work with MPlus and not with AMOS. We did not use ML but the weighted least squares (WLS) estimation method, as recommended in the Flora and Curran Paper (2004). We described this procedure in the methods section of the manuscript (page 7, evidence based on internal structure). Perhaps there was a misunderstanding because we used the ML for the Missing values analysis and this was described only one sentence before we explained the CFA (results section, page 10). Accordingly, we described the estimation method used in CFA more specifically also in the result section of the manuscript.

5) I understand why you include the SOS but isn’t a possible take home message for the reader that the much more efficient SOS, which has acceptable reliability and validity information, correlates well with the SAQ and therefore could be used instead of the SAQ?
Response: This could be considered if one wishes to focus on patient safety related behaviors of clinicians such as nurses. However, using the SAQ will provide more specific information on the perceptions and attitudes of healthcare professionals regarding safety culture dimensions than would be possible by using the SOS alone. This will allow one to identify which dimension patient safety where improvements might be necessary. This is the reason, we don’t recommend using the SOS instead of the SAQ. But the readers have with our data the possibility to argue for using SOS.

Minor essential revisions
1) In table 1, what does the following mean? “There is a moderate to strong correlation between mean values of SAQ and the once of the Safety Organizing Scale distributed at the same time.”

Response: We hypothesized that there would be moderate to strong correlations between the SAQ factor scores and scores on the SOS. We revised the sentence in table 1.

2) Were there differences between nurses and physicians in SAQ scores?

Response: There were some differences in SAQ scores between nurses and physicians in terms of Agree/Disagree percentages, but we did not report them given the relatively small sample size.

3) If culture is a unit-level phenomenon, why do you report only individual-level results?

Response: Since we focused on the psychometric testing and properties of the German version of the SAQ we focused on the individual level results. Data aggregated at the unit level will be reported in the paper mentioned in our response to the previous question.

Comments of reviewer 2 (BH)

Major Compulsory Revisions

1) Important results are not mentioned in the results section, but later discussed in the discussion section, and other data is not listed in the result tables:
   – The items with high missing values: What are the actual missing values and which items are these? In the discussion section (first paragraph) high values are mentioned but they are not anywhere else listed and there is no information on the particular item characteristics.

Response: We added information on the criteria for dropping items due to missing values to the data analysis section and added information on the number missing items and their general characteristics to the results section. In the interest of not increasing the word count too much, we refer readers to Table 3 to see which items were dropped.

   – In table 3 only four factors are listed. What about the missing two? In this table only 21 items are listed, nine items are missing.

Response: All six factors and all 30 items of the SAQ are listed in table 3. Perhaps there was a visibility problem in the PDF file?

   – Is the result of the Swiss analysis a six-factor-model as well? There is no information on this subject, only the fit indices of the confirmatory factor analysis are displayed and only four factors described in table 3. This is irritating.
Response: Yes, the Swiss analysis also showed a six factor model. As above mentioned, we can’t explain why there are only four factors listed in your version. We are sorry for the inconvenience. We will be extremely attentive for the re-submission, making sure that all tables are displayed correctly.

– Table 3 would benefit from the information on CVIs (content validity indices) of items and scales and of the number of responses that could be analyzed. The authors regret the smaller sample size due to some items with higher missing values but there is no information on the actual number of responses that could be incorporated into the analysis.

Response: We agree that the specific CVI-values will be of importance for readers. Consequently, we added the column for CVI-values at table 3, and the title of table 3 as well as a footnote for the CFA columns indicating that it was performed on the 194 subjects with complete data. We also included this in the results section.

– What was the sample size for the confirmatory factor analysis and what are recommended sample sizes?

Response: We conducted the CFA on only participants with full data (n=194) and with all participants (using Expectation Maximization Algorithm to impute missing values). The fit statistics for both models are shown in Table 4 along with the sample sizes. In Table 3, the CFA standardized model factor loadings and standard errors are reported for the analysis using the 194 participants with complete data. We added additional information to the psychometric testing section of the methods, to the results section and to Table 3 to clarify this. There are many different recommendations in terms of minimum sample size in confirmatory factor analysis. In our analysis, we followed the suggestions of Pett et al. (2003), that there is a need of at least 10-15 subjects per item and a minimal overall sample size of 300 cases, following the categorization of Tabachnick and Fidel, (2001).

2) The course of the discussion section seems to be a bit confused

– 2. Paragraph: Why is the comparison of the results with the original questionnaire impaired due to a smaller sample size?

Response: Because we had to conduct the factor analysis with another item-structure than the original. We specified this sentence accordingly in the manuscript.

– 3. Paragraph: Why are items with response means of higher than 3.5 skewed to the negative end? To me it seems that they should have means less than 3 or better 2.5 to be skewed to the negative end.

Response: We agree with the reviewer’s concern and have deleted this section from the manuscript.

– 4. Paragraph: What are factor allocation indices? Where is the data discussed in this paragraph? And what is its meaning?

Response: The factor allocation indices described in this paragraph give information about the mapping of an item to a factor (modification index, cross loadings, residual variance). Data are shown in the results section (page 10, 3. paragraph). However, we agree that this information needs to be discussed in more detail. High modification indices or cross loadings can be considered as a sign for items which can’t be mapped clearly to one factor. We described the Information about factor allocation in more detailed in the result section.

– 5. Paragraph: I did not understand the sentence “It is notable the item “problem personnel” was elicited using the unit versus hospital designation and may have suffered from the same methodological confusion as the perception of management sub scale.”

Response: We absolutely agree with the reviewer. The statement is unclear and since it does not cover essential information we’ve delete the sentence.
8. Paragraph: Why was the decision to collapse three Likert scale categories into one appropriate? Why did it not change the main statement (which main statement?) of the SAQ?

Response: While we recognize the limitations of having to collapse the responses categories, we would argue that combining the neutral response with the slightly and strongly disagree responses is appropriate for this scale because staff, who can’t decide to agree at least slightly, can be considered as more critical toward the situation. We revised the section accordingly.

8. Paragraph, following sentence: Which issues with the factor analysis are meant?

Response: We wanted to say that the above mentioned problems (missing values, sample size, distribution, category collapse) are possible reasons our factor analysis results, which are in some parts different from other studies. We revised the sentence accordingly.

10. Paragraph: What is meant by the last sentence of this paragraph? Why are these results useful? What is the association (of what?) with low prevalence outcomes (patient safety incidents?)

Response: We believe that time.

Minor essentials Revisions

Abstract, result section: Internal consistency is displayed with data that is not displayed in the text or in tables (#=0.878). 0.647 – 0.827 should be displayed.

Response: We corrected the numbers in the result section of the abstract.

Conclusions: It seems to be wise to state that the German language version of the SAQ needs some minor improvements and after these may appear to be a sound instrument. The recommendations in the text do not match with the conclusions in the abstract section.

Response: We agree that readers might draw other conclusions from the main text than from the abstract. We therefore revised the text under the subheading “recommendations” to match with the text in the abstract.

Keywords: I would recommend using established mesh terms

Response: We think that is important to keep the keywords “Swiss hospital setting” and “Safety Attitudes Questionnaire (SAQ). SAQ is a common search term in this field and interested readers will probably add the “Swiss setting”. However, we changed the keywords into MESH terms as follows: Psychometric properties into psychometrics, and patient safety culture into patient safety.

Background section: 3.paragraph: the measurement of patient safety culture had been published in many cases but not well explored (as Rhona Flin concluded in her review on the subject). Some refer to patient safety culture as having the precision of a cloud. Please revise this notion.

Response: We agree, and revised sentence.

Methods section

1. Paragraph: What is the rationale for the sampling procedure?

Response: We realize that rationale was not sufficiently explained and we described the rationale for sampling in the method section.
– 4. Paragraph: Is reference 5 really demonstrating the broad implementation of the SAQ and not only citing other papers?

**Response:** Thank you; this reference does indeed not indicate a broad implementation. We deleted this reference.

– Statistical analysis: What is IQR? Inter quart range?

**Response:** Yes, we define **IQR** in the sentence “(inter quartile range)” and we added a footnote to table 2.

– Psychometric testing: Why was the SAQ compared to the SOS? Is this instrument validated in the same setting and country and language?

**Response:** The German version of the SOS was used because of its good psychometric properties in similar hospital settings. We added information about validity and reliability indices of German version of the SOS including a recent reference.

Result section

– 3. Paragraph: Percentages are displayed but it is not stated the relation to what. What do ceiling effects in five of 30 items (55 to 63 % of ?) mean? Please provide the relation to the other data as well.

**Response:** We agree this is difficult to understand for readers. Additionally, there was a mistake regarding the floor effect: its item number 22 not 21. We described the ceiling effects in more in detail in the result section and corrected the item number.

– 4. Paragraph (content validity): I-CVI was demonstrated to be good, but values between 0.35 and 0.95 reported. This obviously a mistake.

**Response:** Yes, this sentence was not correct; we revised it accordingly.

– 5. Paragraph: How could be confirmatory factor analysis be carried out if normal distribution seems to be a prerequisite and many of the items were skewed?

**Response:** If data are skewed and ordinally scaled, CFA can be done as well, but not using the maximum likelihood estimation, but rather the weighted least squares (WLS) estimation method (as described in Muthen, 2007; Flora and Curran, 2004). This was also the reason why we used Mplus for our analysis.

Discussion section

– Please divide the discussion section in shorter paragraphs. Sometimes it seems to be quite wordy.

**Response:** We revised the structure of the discussion section accordingly.

– 1. Paragraph: A cronbach alpha of less than 0.8 is moderate but cannot be declared to be a sign for a strong reliability (with the other data being mostly less than 0.8).

**Response:** We agree, the word “strong” is- given our results- to “strong” and needs to be damped. We therefore revised this sentence.

– 2. Paragraph: The crucial step of the translation of a questionnaire is the understanding of the intention of an item in the original instrument. This intention is not always clear to everyone and unequivocal. Maybe the authors struggled with some of the translation because the intention of them were not self-evident

**Response:** We had the chance to work with Brian Sexton, the developer of the original scale. So we could ask him and discuss the intention of the individual items. We also discussed the wording and meaning of items in the backward translation procedure with US colleagues (co-authors).
3. Paragraph: The scale stress recognition had already been demonstrated in earlier studies (e. g. Speroff et al 2010) to follow a different pattern than the other scales of the SAQ. Speroff then opened the debate if this scale is an aspect of safety climate at all.

Response: We thank the reviewer for this important information and we added Speroff’s notion in the discussion section.

Table 1: 3. Column, 4, Line: “once” instead of “ones”?

Response: We corrected the term to “Ones”

Table 6: Please provide explanations for abbreviations in the table heading (CVI, MV, CFA).

Response: We inserted a footnote.