Reviewer's report

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

Version: 1 Date: 18 November 2005

Reviewer: Thomas Perneger

Reviewer's report:

General

--------------------------------------------------------------------------------

Major Compulsory Revisions (that the author must respond to before a decision on publication can be reached)

This is an important paper, even overdue, given how much this instrument has been used in recent years. So far, the problem has been that several versions of this questionnaire circulate, and little information is available as to what changes were made where and why. This paper will establish a somewhat shortened version of the questionnaire for future use. This is an important contribution.

The paper is generally well conceived and clearly written. Numerical results are extensive (almost too much so), and will allow potential users to assess the merits of the instrument. However, I had several concerns about the theoretical justification of some of the analyses presented (see below). I know that the authors cannot go back and do things differently, but clearer argumentation would strengthen the paper.

General comments

1. Perhaps the weakest point of the instrument is the evidence about its validity. The authors rely mostly on a clean factorial structure of the final 30-item instrument (but about this, more below). What is missing is a clear theoretical framework of safety culture/climate/attitudes. The authors started with another instrument, the FMAQ, but what this instrument does is not explained. Then, they generated additional items with focus groups of health care providers. What exactly were the focus groups instructed to do? Were they provided with a definition of safety culture? Were they asked to identify whatever may influence patients’ safety? It is important to clarify this. For instance, I would have thought that satisfaction at work is an extraneous variable, that may be associated with a culture of safety (as a pre-condition? as a consequence? both?), but that it is not intrinsically part of a culture of safety or a safety climate.

2. I would not consider most of the items as measuring attitudes or culture at all, to the extent that culture is a set of beliefs and values shared by a social group. Most items ask respondents to rate their work environment. What results is a rating of an organisation by its employees. If the employee moves to another hospital, his/her ratings will change, while their attitudes or beliefs should not. The exception is the “stress recognition” scale, which does probe beliefs/attitudes. This dimension is the only truly psychometric subscale of the lot. So, the authors should state more clearly what they set out to measure and why – the explanation on page 4 is disappointing.

3. Factor analysis is key to this paper, but its use is a bit circular. FA was applied to the initial set of 100 items to identify relevant dimensions (page 5). So these dimensions are empirical, not grounded in an a priori theoretical model. Then FA was used again on the full dataset to verify if the postulated
structure is confirmed by the data. But how could it not be, since the model was developed using the
same questionnaire and the same analytic method? This only tells us that the factorial structure is
stable. The issue of validity should be addressed in the Discussion.

4. Furthermore, whatever discrepancy there was in the confirmatory FA disappeared after deletion of
10 items that did not load as predicted. But does that really make them bad items? I am not so sure;
they may be crucial in terms of content validity (but to know this we need a clear conceptual
model...). In any case, a very unorthodox use of CFA.

5. More explanation and justification is needed about the individual versus area level analyses. What
does one tell you that the other does not? If their results differ, what does it mean? The paper
contains many numerical results that are barely commented upon.

6. This leads to the Discussion. I was hoping for the Discussion to clarify what the results mean, how
they relate to other work on safety attitudes, what the limitations of the study are. Instead, the
discussion is a bit of a promotional text for the instrument (which hardly needs it, given its
widespread use...), disconnected from the results. It should be more closely related to the topic of
the paper, and the speculative content should be reduced.

--------

Minor Essential Revisions (such as missing labels on figures, or the wrong use of a term, which the
author can be trusted to correct)

7. Abstract (p3): avoid unexplained terms such as “executive walk rounds”
8. Background (p4): define safety culture here, or even better, show a conceptual model, perhaps
graphically
9. P5: provide details about the FMAQ, and about the focus groups
11. P9: explain what is gained by computing SEM-based alpha coefficients, instead of simple
coefficients.
12. P10: no discussion nor references should appear in Results.
13. P10: 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.
14. P16: the link does not work.
15. Table 2: are the SDs for area means, or for individual measurements?
16. Table 4: tell us how to interpret the different alphas. I am not sure you can speak about reliability
regarding the between-area alphas. Area means for these items are correlated, but can one say that
the non-shared variance is measurement error?
17. Appendix A. I do not understand what these results are. Loadings or path coefficients? Why
report both standardized and unstandardized results? Do we need confidence intervals? Do we need
these results at all?
18. Appendix B. Again, explanation needed. I am not sure that these appendices are referred to in
the text, and certainly they are not discussed or interpreted.
19. Figure 1: unnecessary, as all items appear in a table.

--------

Discretionary Revisions (which the author can choose to ignore)

20. P8: data analysis: explain concepts behind MLR chi-2 test, spell out MLR
21. Table 1: Floor/ceiling effects are given, but for which scale? There are 6 of them. I’d suggest that
this detail be left out – a global statement about low floor/ceiling effects should be enough.
22. Table 3: round off percentages (eg, 81, not 81.31), the precision is irrelevant. Consider removing
skewness and kurtosis, of little use to most readers.
23. Figure 2: I would find a frequency distribution easier to read than a ranking. But this is interesting for instance, the range in job satisfaction ratings is astounding.

What next?: Unable to decide on acceptance or rejection until the authors have responded to the major compulsory revisions

Level of interest: An article of importance in its field

Quality of written English: Acceptable

Statistical review: Yes

Declaration of competing interests:

I declare that I have no competing interests