Reviewer’s report

Title: Developing and testing an instrument to measure the presence of conditions for successful implementation of quality improvement collaboratives

Version: 1 Date: 8 February 2008

Reviewer: Susanne Hempel

Reviewer’s report:

- Major Compulsory Revisions

The author must respond to these before a decision on publication can be reached. For example, additional necessary experiments or controls, statistical mistakes, errors in interpretation.

1. The authors extracted seven components from a 30 item questionnaire. With such a high number of extracted components in a small item pool follows a typical problem: Two of the seven components become scales that contain only two items. A psychometric scale with two items is controversial. Although the seven factors are distinguishable in the principal component analysis, since the PCA was employed to create scales, scale aspects need to be considered as well. The authors should at least address this problem in the text or state how they will deal with this issue. The possibilities are to accept two item scales, to add similar items to create a longer scale, to discard the two factors, or to critically review the decision to extract seven factors from a 30 item pool. Although there was internal consistency in this sample, there are reduced chances that such a short scale will achieve acceptable internal consistency in another sample.

2. It is not clear how consistent the cross-loading issue was approached. Item 11 and 29 were removed but why were items 27 and 28 left in the questionnaire when they had substantial secondary loadings on the first component?

- Minor Essential Revisions

The author can be trusted to make these. For example, missing labels on figures, the wrong use of a term, spelling mistakes.

3. The implied superiority of an answer mode that contains a middle category instead of a risky 4-point scale is not warranted in my opinion. Using a neutral answer is perfectly legitimate but it should not be overlooked that it comes with its own set of problems. It is no secret that a middle category threatens the underlying interval assumption of a rating scale (3 is more than 1 and less than the score 5) since a middle score can either indicate a middle position on an interval scale or indicate that the rating scale was not used (so being outside the dimension that is assessed because the respondent couldn’t answer)
4. The method section reports some results---the number of missing items and one measure of dispersion characterizing the answer mode. Since this is a legitimate issue when evaluating the quality of an item this may be better addressed in the result section (i.e. items that produce a lot of missing answers are not useful).

5. The measures associated with table 3 are missing in the result section, i.e. the analysis section mentions a measure of dispersion but the result section paragraph is missing.

6. The factor analytic write-up would benefit from adding the scree plot. The decision to extract seven factors from a questionnaire that only contained 30 items could be considered controversial. Although the authors report that they looked at the scree plot, I think it would be easier for the reader to follow that decision were the scree plot included in the result section. Especially since the first 10 or so eigenvalues were not reported so that the reader would have another indication of why seven components needed to be extracted.

7. There are different criteria in what to look for in a scree plot inspection. The authors mention that they looked at where the plot abruptly levels out. This may confuse Cattell’s approach of identifying scree, which does not necessarily show an “abrupt” change of level and the more pragmatic approach of looking for a visible gap in the scree plot. I would suggest the authors clarify this and/or report the scree plot.

8. I think it would be more appropriate not to call the analysis a “factor analysis” in the paper but a component analysis since the employed PCA does not find underlying factors but the contained components that are entailed in the sample; this would apply to all text and tables.

9. It seems the method section does not entirely match the results section. Some of the planned analyses do not seem to have been reported.

10. The order of mentioned analyses in the method and result section do not really match which may make it harder for the reader to follow the paper.

11. The factor analysis approach takes up most of the result section with classic reliability and item analytic considerations being mainly reduced to the convoluted table 6. I personally think it would be worth looking at other reliability measures than Cronbach’s alpha and to report them clearly. The method section suggests that corrected item total correlations will be presented but these are not that easy to find in the result section. It could be argued that table 5 overwelms the reader with information that mainly is not critical (since the loadings for all items on all components are already reported in table 4) and the important information (item-totals (?) for each scale) drowns in the number of results.

12. It may be better to find an alternative wording for ‘corrected for overlap’ in table 5 (is is not entirely clear what that means and how it was achieved).

13. Initially the authors assumed a five factor model. In case the authors tried to replicate these five factors rather than being guided entirely by the exploratory analysis, this should be reported. It may be worth looking into setting the number
of extracted components to five and to evaluate the emerging components conceptually. Especially if the authors insist on the seven component structure it may make a stronger case were other models investigated apart from following entirely the exploratory and data-driven nature of the procedure.

14. This paper addresses the factorial validity of the designed questionnaire. A very important point, however, I think it is worth pointing out that this is not the only criteria a measurement tool has to fulfill. The items were derived from previous research and by consulting experts what do mainly practitioners think what are characteristics for successful implementations of quality improvement interventions. Quality improvement research is only at the beginning in building up a constructive evidence base. There are not many certainties yet, we do not know a lot about what kind of quality improvement intervention or components works. Studies that systematically investigate the variables that predict successful short term and long term implementations are not as numerous yet as that they could have meaningfully informed the initial item pool. To me it seems that the authors demand a leap of faith from the reader that the initial item pool consisted of relevant items and covered all relevant aspects. From the current study we only know that there is empirical evidence that practitioners are able to distinguish the mentioned factors. Whether these are in fact related to successful implementation of quality improvement interventions needs to be seen. The authors address this point as further research, however, it could be made clearer that so far we don’t know whether the questionnaire really measures what it was designed to measure.

- Discretionary Revisions

These are recommendations for improvement which the author can choose to ignore. For example clarifications, data that would be useful but not essential.

Please note that both the comments entered here and answers to the questions below constitute the report, bearing your name, that will be forwarded to the authors and published on the site if the article is accepted.

15 The distribution of answers on the answer mode could be discussed a bit more critically I would have thought. The seven point answer mode was introduced as providing room for distinction. The empirical results regarding which answer category was used indicate rather that for several items the extreme answer was not useful.

16 Although the decision to use a Varimax rotation was understandable and was justified in the text, it might be worth mentioning the conceptual problem here: are the factors really meant to be independent conceptually. It might be worth adding a sentence that the focus was on identifying components that can be distinguished empirically.

What next?

-------
Based on your assessment of the validity of the manuscript, what do you advise should be the next step?

- Accept after minor essential revisions (which the authors can be trusted to make)

Level of interest

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

BMC Health Services Research has a policy of publishing all scientifically sound research whatever its level of interest. However if you choose one of the first three categories below, we may ask the authors if they would like the manuscript considered instead for the more selective journal BMC Medicine.

- An article of importance in its field

Quality of written English

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

As we do not charge for access to published research, we cannot undertake the costs of editing. If the language is a serious impediment to understanding, you should choose the first option below, and we will ask the authors to seek help. If the language is generally acceptable but has specific problems, some or all of which you have noted, choose the second option.

- Acceptable

Statistical review

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

Is it essential that this manuscript be seen by an expert statistician?

If you feel that the manuscript needs to be seen by a statistician, but are unable to assess it yourself then please could you suggest alternative experts in your confidential comments to the editors.

- No, the manuscript does not need to be seen by a statistician.

**Declaration of competing interests:**

I declare that I have no competing interests