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: 6 February 2008

Reviewer: Dimitris Niakas

Reviewer's report:

The paper develops an instrument for measuring the presence of conditions that could potentially lead to a successful implementation of quality improvement collaboratives (QICs) in Netherlands. It reverts to literature review and expert opinion to form a questionnaire consisting of factors thought as affecting the success of QICs. The psychometric properties are then tested in sample of 101 leaders by forming multi-item scales with Principal Components Analysis and assessing their internal consistency reliability with Cronbach's alpha coefficient and convergent and discriminant validity via multi-trait methodology. The authors are generally careful with the methods and the results are in general satisfactory. It is well planed and interesting paper in a vital area of health care management.

However there are some major weaknesses. These are described bellow.

1. When creating a new instrument it is easier to remove items rather than add new ones, which would require a new sample to test them. Data should be collected on at least five but fewer than 10 items for each dimension theoretically postulated to define the construct. This recommendation is based on Guilford's rule that at least three items are needed to define a factor, balanced with what is known about reliability estimation. There are numerous formulas that demonstrate that the more items on a measure, all other things being equal, the more reliable the measure. Two of the factors identified in this instrument are comprised of only two items.

2. Although the instrument was demonstrated to be reliable and evidence of content validity was provided, the only evidence of construct validity is inter-item and inter-factor correlations. There is no statistical evidence that the measure is positively related to other constructs that should be positively related (e.g., conditions for successful implementation). How do we actually know that this instrument will achieve its main purpose, which is to assist in adopting quality improvements? To assess this, collection of additional data (measurable outcomes) is required which can be associated with the surveys and subsequently analysed.

3. The factor structure is questionable and confidence in the instrument is low. One of the main factors for this uncertainty is the small sample size. Comfrey and Lee (1992) suggest that the adequacy of sample size might be evaluated
very roughly on the following scale: 50 â## very poor; 100 â## poor; 200 â## fair; 300 â## good; 500 â## very good; 1000 or more â## excellentâ##. Gorsuch (1983) recommends a minimum subject to item ratio of at least 5:1 in exploratory factor analysis (EFA), but notes that higher ratios are generally better. There is a widely-cited rule of thumb from Nunnally (1978) that the subject to item ratio for EFA should be at least 10:1 In the case of this study it is 101:30 or 3.4:1, which is well below the recommended standard. Bryant and Yarnold suggest there should be at least 10 cases for each item being used, the subjects-to-variables ratio should be no lower than 5 and the total number of cases used in EFA should not be less than 200 (Bryant and Yarnold, 1995).

4. Another problem is the diversity -combined with small sub samples- of the professional groups used for testing the instrument. Although all subjects were regarded as project leaders, their different professional backgrounds imply different perspectives and opinions. Although the authors acknowledge this as a potential limitation, it still remains a problem. Furthermore, the authors appear to easily overcome the problem by stating that â##despite these differences, it is likely that the instrument is suitable for â#¦..â##. On what arguments is this assertion based?

5. The title of the paper suggests that an instrument is developed and tested that measures the presence of conditions for successful implementation of Quality Improvement Collaboratives. Indeed, a questionnaire is developed which depicts whether several processes, support, etc. are present. But it does not show that the scale scores achieved affect success and to what extent they can predict it. In fact, the authors leave this for future research. Therefore, isnâ##t it necessary to show that the literature presented on pages 4-5 indeed proves their success empirically and, in addition, whether evidence exists that the same conditions can be transferred to another setting internationally?

6. On page 8, (line 9) it is mentioned that 101 questionnaires were returned. On the same page (next paragraph) it is also said that 3 respondents had filled in less than half of the total number of items and were excluded from further analysis. Yet on page 9 (Results Section) again 101 leaders are mentioned, whilst adding up 25 physicians, 20 advisors, 15 unit heads, 6 nurses, 3 pharmacists and 31 other positions gives us a total of 100. Which figure is the correct one (101, 98 or 100)?

7. On page 11 the authors suggest they removed items 11 and 29 from further analysis when it is obvious from Tables 3 and 4 that items 11 and 30 have been in fact removed. A more close examination of items should take place since other errors seem to exist. For instance, in Table 5 item No 28 is mentioned twice (within factor 1 and 7).

8. Table 6 could be simplified, since the range of correlations between items and other factors seems to give redundant information (already given in Table 5). Also, some of the footnotes of the Table give the same information as can be found in the main text.
9. The authors state that no items were excluded based on the percentage of missing responses, although a missing value of about 15% was observed. Why? Isn’t this a high percentage that might imply problems with the clarity of the question?

10. The scale of item 23 is said to have been inverted (Table 3, page 21). What does this mean? Typically, Likert scales allow one to interchange positively and negatively worded items in order to minimise any positive biases. Has this been done in only one question here?

11. Why hasn’t test-retest reliability been assessed? This should be mentioned in the limitations section.

A suggestion which the authors could consider is to use this work, which is otherwise well planned and interesting, as an initial pilot test of the instrument from which the results would guide a more extensive field-testing phase, while adhering to the sample-size rules mentioned previously.

References


What next?: Reject because scientifically unsound

Level of interest: An article whose findings are important to those with closely related research interests

Quality of written English: Acceptable

Statistical review: No, the manuscript does not need to be seen by a statistician.

Declaration of competing interests:

I declare that I have no competing interests