Reviewer’s report

**Title:** Beyond the Frontiers of Effectiveness: A Framework and a Measurement Instrument for Sustainability of Work Practices

**Version:** 1 **Date:** 9 December 2010

**Reviewer:** Michel Dückers

**Reviewer’s report:**

1. **Is the question posed by the authors well defined?**

   The objective is clear. This paper deals with the sustainability of improvement initiatives in health care organizations; an important topic in quality improvement research. Objective of the study is to conceptualize relevant dimensions like routinization and institutionalization and to develop a measurement instrument.

2. **Are the methods appropriate and well described?**

   Structural Equation Modeling is an appropriate technique to explore and confirm relations between variables measured and latent constructs. Imputation is a technique to overcome problems of missing data. The text largely enables the reader to understand the steps taken and their results.

   **Major Compulsory Revisions**

   On page 13 we find: ‘In addition several possible structures in the data were explored with principal component analyses- these however will not be reported for practical reasons of limitation.’ The wish to reduce the word count and to optimize the size of the paper is appreciated but unnecessary. The advantage of an open access journal is that authors can easily add more extensive appendices. Additional principal component analyses and SEM results allow the reader to determine if items also load (or perhaps even load better) on other constructs.

3. **Are the data sound?**

   **Major Compulsory Revisions**

   This is an aspect the authors are aware. Nevertheless, I must be critical. A great deal of work has been performed. Still, I have major concerns about the quality of the data and the fact that the Structural Equation Modeling is performed using a sample this limited in size. Kass and Tinsley (1979; ‘factor analysis’) recommend five to ten cases for each item. Nunnally (1978; ‘psychometric theory’) is recommends at least ten cases for each item. When it comes to sample size 100 is considered poor, 1000 excellent (Comrey & Lee 1992; ‘a first course in factor analysis’). From the perspective of these rules of thumb the amount of available data in this manuscript is problematic. The questionnaire contains 52 items. The number of cases is said to be 112, although Table 1 suggests that for most items
80-100 values have been collected. Comrey and Lee would definitely view this as poor (if not worse). And then we have not even taken into account the number of cases with no missing data (N = 33). The number of cases for each item in this study is at best 1:2. I seriously doubt whether I would have chosen to perform these analyses, also given the low response rate. The authors should respond to this issue.

4. Does the manuscript adhere to the relevant standards for reporting and data deposition?

Yes, although several potentially important explorations have been left out (comment 2). These can still be added as additional files.

Major Compulsory Revisions*

The authors chose to report Confirmatory Fit Indices (CFI), a typical LISREL fit-measure. CFI scores are most of the time positive. Why did the researchers not report the value of the more advanced Tucker-Lewis Index (TLI)? This is also a common index to report. TLI- and CFI values below .90 indicate that the model can be improved. Values between .90 and .95 are acceptable, Values above .95 are good (Bentler & Bonett 1980; ‘significance tests and goodness of fit in the analysis of covariance structures’).

Minor Essential Revisions**

Page 13: ‘For a good model fit the RMSEA values should be low and are recommended to range between .10 and .05’. Good models have an RMSEA value of equal to or lower than .05, values between .08 and .05 are considered acceptable (Browne & Cudeck 1993; ‘alternative ways of assessing model fit’).

5. Are the discussion and conclusions well balanced and adequately supported by the data?

Discretionary Revisions***

The abstract ends with ‘the excellent psychometric properties of the subscales warrant application of the instrument in the evaluation of sustainability of improvement projects.’ I can imagine that the authors want to invite other researchers to use the instrument. With the difficulties regarding the data in mind, together with the idea that we actually have no idea that the instrument really measures the thing it is designed to measure (this requires additional testing), I want to encourage the authors to be more modest.

6. Are limitations of the work clearly stated?

The fundamental problem with the sample size is mentioned.

Major Compulsory Revisions*

The authors should address that Test-Retest Reliability is not assessed and that – besides psychometric properties - no evidence is provided that the instrument actually measures what it is supposed to do.
7. Do the authors clearly acknowledge any work upon which they are building, both published and unpublished?

Yes, this study is linked to key publications and contains meaningful reflections within the context of routinization and institutionalization.

8. Do the title and abstract accurately convey what has been found?

They do. In relation to the title one could argue whether ‘beyond the frontiers of effectiveness’ is too cryptically or not. This is a matter of taste. The second part of the title is fine.

Discretionary Revisions***

The abstract is easy to follow, although I am not sure if I truly understand what the authors mean by ‘routinization (...) of organizational routines’. They seem to point at a mechanism.

I do not like the use of terms like ‘excellent’ and ‘valuable’. Instead, I prefer ‘is in line with typical psychometric properties as defined in the literature’ and ‘provides an opportunity to test...’

9. Is the writing acceptable?

The manuscript is well written.

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

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

*** Discretionary Revisions (which are recommendations for improvement but which the author can choose to ignore).

**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.