Author's response to reviews

Title: Factors influencing success in quality improvement collaboratives: development and psychometric testing of an instrument

Authors:

Loes M.T. Schouten (l.schouten@cbo.nl)
Richard P.T.M. Grol (r.grol@iq.umcn.nl)
Marlies E.J.L. Hulscher (m.hulscher@iq.umcn.nl)

Version: 3 Date: 4 July 2010

Author's response to reviews: see over
Dear editor,

Please find resubmitted our revised manuscript ‘Factors influencing success in quality improvement collaboratives: development and psychometric testing of an instrument’. Based on the reports from the reviewers we have made the revisions in our manuscript. This covering letter provides, point by point, our replies to the comments made by the reviewers and explains how we have dealt with them in our manuscript. We would be very pleased with and are looking forward to a final decision on accepting our revised review for publication.

Yours sincerely,
Loes Schouten & Richard Grol
IQ Healthcare, Scientific Institute for Quality in Healthcare
Radboud University Nijmegen Medical Centre

Corresponding author:
Loes M.T. Schouten
Dutch Institute for Healthcare Improvement
P.O. Box 20064
3502 LB Utrecht, The Netherlands
Telephone: +31 30 2843944
Fax: +31 30 2943644
E-mail: l.schouten@cbo.nl
Reviewer's report I
Title: Factors influencing success in quality improvement collaboratives: development and psychometric testing of an instrument
Version: 2 Date: 16 February 2010
Reviewer: Mireille Van den Berg

Reviewers report:
1. Is the question posed by the authors new and well defined?
I share the view of the authors that there is currently a lack of information on which factors contribute to the success of Quality improvement collaboratives (QIC’s). There is great need for a checklist that guides initiators in starting up a QIC based upon experience from earlier successful QIC’s.

2. Are the methods appropriate and well described, and are sufficient details provided to replicate the work?

Major Compulsory Revision: Literature search
The authors refer to an earlier paper for their searching strategy. I would like to have more information on the strategy they used (search terms, databases inclusion and exclusion criteria) in the method section of this paper. It would give me more insight in the type of papers the authors reviewed. The author identified five papers that met their inclusion criteria, what are they?

A: For the purpose of brevity we did not repeat the search strategy in the paper but referred to an earlier paper (our systematic review published in BMJ; Schouten, 2008), but we agree with the reviewer that the reader may like to have more information on the strategy used. In our revised manuscripts we added a description of the strategy and papers that met our inclusion criteria:

“We searched the Medline, Cinahl, Embase, Cochrane and Psychinfo databases for literature about QICs in the period from January 1995 to June 2006 inclusive. We started with a Medline search for free text terms describing QICs, and we combined the keywords (non-MeSH): “quality and improvement and collaborative” or “(series or project) and breakthrough”. The same steps were repeated for the other databases. We also reviewed the reference lists of the included papers. To distinguish and define determinants of success, studies were included if (1) they gave an overview of key elements or components of QICs applied in Health Care, and (2) were written in English. Titles of articles and abstracts identified in the search were reviewed for relevance by two researchers (LS, MH).

Furthermore it seems a rather small number of studies to base a 50-item instrument on. This can also be concluded from the Cronbach’s alpha of 0.86 and 0.92, which is rather high for internal consistency, suggesting that some items are overlapping and therefore redundant.

A: Our search identified five studies that met our inclusion criterion in giving an overview of key elements of QIC applied in Health Care. These papers provided a list of 128 items (!) of expert opinion based determinants of success.

Although the Cronbach’s alpha of 0.86 and 0.92 is rather high for internal consistency, this seems partly due to the number of factors (three) found in
exploratory factor analyses and the subsequent relative large number of items per factor. To check redundancy or overlap in our 50 item instrument we considered removing items that showed redundancy of measurement by a high correlation ($r>0.85$) with another item. In addition (in an attempt to reduce items) we considered removing items with the following characteristics in our revised manuscript: items with a high proportion of missing responses (>10%); (and; (3) items with skew distribution (items with > 90% of the answers in categories 1 and 2 or 4 and 5). Based on these criteria nine items were removed (see also reviewer 2).

Major Compulsory Revision: Selection of items
How many items came originally from the review and how many were deleted after each selection by the author and by the expert panel

A: The five papers overviewing QICs provided a list of 128 items of expert opinion based determinants of success. The authors (LS and MH) condensed this list to 72 items. After review by the expert panel the list was reduced to 50 items. We included these numbers in the Method section of our revised manuscript.

Major Compulsory Revision: The instrument
The authors state that the proposed instrument can be used prospectively as a checklist to guide initiators, facilitators, and participants of QICs with information about how to carry out or participate in a collaborative with theoretically optimal chances of success. The result looks to me more like a questionnaire for judging the collaborative retrospective. Rephrasing all questions into a checklist requires the same steps for testing construct validity and internal consistency. The result from this study does not provide us with a checklist to guide initiators yet.

A: Although the reviewer is correct that we did perform construct validity and internal consistency testing using the instrument retrospectively, the items in the instrument nevertheless can be used prospectively as well. With a little rephrasing the instrument can be used to inform or guide initiators, facilitators or participants about how to carry out or participate in a QIC:
1. Chairperson of the expert panel is/should be an opinion leader
2. Expert panel gives/should give advice on changes
3. Chairperson of the expert panel is/should be an expert in the topic of the QIC
4. Expert panel has/should have ample time
5. Expert panel gives/should give positive feedback
Etc
Etc

The reviewer is correct that rephrasing all questions into a valid and reliable checklist requires the same steps for testing construct validity and internal consistency. We emphasized this limitation in the Discussion section of our revised manuscript.

3. Are the data sound and well controlled?
The authors give to little information on the search strategies and the selection of the items following. It is therefore not possible to comment on the quality of these actions. The analysis of construct validity and internal consistency seems solid.
A: In response to the reviewers suggestion we added more information about the
search strategy and selection of items in the Method section of our revised
manuscript (see above).

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

5. Are the discussion and conclusions well balanced and adequately supported
   by the data?
   Major Compulsory Revision: The initial purpose of the instrument to be used
   prospectively as a checklist to guide initiators of QICs with information about how
to carry out or participate in a collaborative, seems to get a bit lost in the
discussion and conclusion of this paper. It is not clear at what point of the
process of working in a QIC this instrument can be used and the authors give no
directions on how to apply it.

   A: As indicated by the reviewer the initial purpose of the instrument to be used
   prospectively seems to get a bit lost. We thoroughly reread our Discussion and
   Conclusion and have rewritten some sections in the Introduction, Discussion and
   Conclusion in our revised manuscript to be more clear about the purpose of the
   instrument to be either used retrospectively and prospectively.

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

7. Is the writing acceptable?
   Yes

   Level of interest: An article of importance in its field
   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
Reviewer's report II
Title: Factors influencing success in quality improvement collaboratives: development and psychometric testing of an instrument
Version: 2 Date: 1 April 2010
Reviewer: Cordula Wagner
Reviewer's report:
In general the article is well written and interesting for a broader group of readers.
Nevertheless, there are important questions to be answered.
The authors state in the introduction that they are not aware of any instrument aiming at understanding factors of success of QIC’s. This is not true.
In the discussion they mention the publication of Duckers et al. with the title: Developing and testing an instrument to measure the presence of conditions for successful implementation of quality improvement collaboratives'.
That publication describes a comparable instrument focussing on the same aim as the instrument described in the article under review. Why not mentioning in the introduction that there is already another instrument. It does not make the article under review less interesting or relevant. There is room for more than one instrument.

A: We agree with the reviewer that the sentence in our Introduction may be confusing as we are aware of the presence of other instruments developed or used to measure success in QICs (Duckers et al; Mills; Mills et al; Neilly et al), although the instruments differ (as mentioned explicitly in our Discussion section). In response to the reviewers suggestion we changed these sentences in our Introduction and refer to the existing instruments in the Introduction section of our revised manuscript.

Result section; sample
First sentence, last word: collaborative should be questionnaire?? If the word collaborative is right: what is meant by: ‘completed the collaborative’?

A: The word collaborative is right here. With ‘completed the collaborative’ we imply that all improvement teams participated in the collaborative activities and did finish the collaborative and that there were no drop outs. To be more clear, we explicated this in our revised manuscript: All sites attended the conferences, submitted monthly progress reports and participated in the collaborative activities.

Clarify: 144 participants of 44 teams completed the questionnaire; the response rate is 95%. In the next sentence it is mentioned that teams have 7 members on average. 46 teams x 7 members is 322 possible respondents. Is a sample been taken? Is it representative?

A: The response rate is referring to the number of teams that completed the questionnaire. From the 46 participating teams, 44 completed the questionnaire (response rate 95%). We agree with the reviewer that the next sentence is a little bit confusing here. We did not take a sample. We asked team members to complete the questionnaire at the last conference. In larger teams not all team members (but generally up to four) do attend these conferences.
Last two sentences before ‘construct validity testing’: the authors describe that the topics of the questionnaire showed high scores and that items showed little variation. What does this mean for the instrument aiming at measuring differences between teams, because not all of them will have success. This is an important point for the discussion.

A: The topics of the questionnaire in our study indeed showed high scores and items showed little variation. This may suggest that –at least for these collaboratives—determinants or conditions seemed to be present or fulfilled. These scores, however, are not necessarily applicable to teams in other QICs. The explicit aim of this study was to describe the development and psychometric testing of the instrument.

We highly agree with the reviewer that further research in QIC projects is needed to determine its usefulness in quantitatively explaining the difference of success between teams participating in a QIC. We therefore need studies that link the scores on the instrument to key effect parameters in QICs. Our research program does include these kind of studies, but we are not able to report on this yet. We elaborate on these issues in the Discussion section of our revised manuscript.

Construct validity: the sentence: ‘a cutt-off point of 0.4 ….’ Has already been described in the method section.

A: We indeed mention the cut-off point of 0.4 in our method section. We removed this redundancy in our Result section.

4 items of the questionnaire has been excluded based in their factor loadings. From a theoretical point of few these questions are quit essential for QIC’s. why do they not fit into the factor structure? Should these questions be reformulated and tested again?

A: Indeed four items of the 50-item questionnaire were excluded. These factors were excluded based on their factor loadings (<0.4). Before items were removed, their importance was considered as judged by the reviewers (LS, MH) opinions on their content validity. Two of the items (item 3.29 and 3.30) are about ‘measurements to plan and test changes. ‘Measurement’ is a quite essential component in QICs. But this element is also represented in item 3.28 (my team gathered measurements), 3.31 (my team used measurements to track progress), 3.34 (my team tracked progress continuously) and other items in the 46 items instrument. It can be argued that the items left may be sufficient regarding the ‘measurement‘ topic in the instrument. For items 2.20 (participation enhanced multidisciplinary collaboration in my organization) and 4.47 (competition between teams during the joint working conferences) it may be worthwhile to reformulate these questions and test them again in further studies. However in an attempt to reduce the number of items another selection was retained for further analysis (as described later in this letter)

Intercorrelations : it is confusing that the names of the scales are different. In the last part of the construct validity section new names for the three scales has been introduced. I would expect these names in the ‘intercorrelations' section as well
A: These ‘slip ups’ are indeed confusing. We thoroughly checked our manuscript and tables on correct and uniform use of names/labels.

Table 3: half of the table is double and not necessary; what does ‘number’ mean (last row).

A: We removed redundancy and unnecessary details in table 3.

Table 4: some of the inter-item correlations are very high (0.8; 0.77); this means that some items are measuring quite the same. As the authors have not enough respondents for a factor analysis of 50 items, it would be preferable to reduce the number of items.

A: The inter-item correlations are relative high. This seems partly due to the number of factors (three) found in exploratory factor analyses and the subsequent relative large number of items per factor (7-18). To reduce the number of items we considered removing items with the following characteristics: items with a high proportion of missing responses (>10%); (2) items that showed redundancy of measurement by a high correlation (r>0.85) with another item, and; (3) items with skew distribution (items with > 90% of the answers in categories 1 and 2 or 4 and 5).

Before items were removed, their importance was considered as judged by the reviewers (LS, MH) opinions on their content validity. After the initial item reduction, we used factor analysis to explore whether the selected questions grouped into relevant domains or concepts.

We deleted 9 items from the initial 50-item instrument: 1.3 (chairperson was an expert), 2.10 (general goals of collaborative were clear), 2.11 (team supported collaborative’s general goals), 2.15 (team directly involved in changes), 2.16 (team had relevant expertise), 2.18 (teams were motivated), 2.21 (team focused on patient improvement), 2.22 (team focused on care process improvement), 3.28 (team gathered measurement data) with 90% of the answers in category 4 and 5.

One more item (4.47: competition between teams) was removed because the factor analysis showed it did not fit with any distinct factors representing the different concepts.

Discussion: again the authors state that their study is the first study exploring potential factors of success for QIC’s. This is not true. They refer to three other studies later on. Besides they have based their selection of topics on the literature. There is a much broader literature base on success factors. There only a few instruments. Therefore, the article is still relevant.

A: The reviewer is right in stating that there is a much broader literature base on success factors. However, our instrument was built exclusively on the key components of QICs based on expert literature and expert opinion about QICs. In response to the reviewers comment we changed the tone of our statement in the Discussion section.
Limitations: the authors describe that there are different standards applied for principal component analysis and that 5 to ten cases per item are generally recommended. I agree. Therefore, I think that the authors should reduce their number of items before the actor analysis. They could use a cut-off point of 0.6 for example.

Now, they have 50 items and only 144 respondents, instead of 250 to 500. 144 is not enough for a valid factor analysis with 50 items.

A: Remarks can be made with regard to the sample size of our study (50 items and 144 respondents). As suggested by the reviewer we reduced the number of items in the analysis of our revised manuscript. After item reduction 40 items were retained for further analysis. Factor analysis of these items showed the same pattern (as the initial analysis of 50 items) and did not change the conclusions of internal consistency: items clustered in the same three scales with comparable loadings and chronbach’s alpha’s and (inter)scale correlations. Although our subject-to-variable ratio still may not meet the ‘5 to 10 cases per item’ standard, the findings and absolute magnitude of factor loadings in our analyses seems quite robust. In addition, several authors state that the minimum level of sample size is dependent on other aspects of design, such as the communality of the variables (MacCallum, Widaman, Preacher, and Hong, 2001, p. 636), degree of overdetermination of the factor (or number of factors/number of variables) (MacCallum, Widaman, Zhang, & Hong, 1999), and the size of item loading (Costello & Osborne, 2005, p. 5). Guadagnoli & Velicer explicitly state that if components possess four or more variables with loadings above .60, the pattern may be interpreted whatever the sample size used (Guadagnoli & Velicer, 1988, p. 274). In our analysis respectively 7 out of 7 (factor 1), 10 out of 18 (factor 2) and 9 out of 15 items (factor 3) showed up loadings > .60.

The authors finish the discussion with stating that the instrument is a valid and reliable measure. This conclusion can not been taken yet; based on this first study only.

A: As the reviewer is correct that this is the first psychometric testing of the instrument we changed the tone of our conclusion.

Level of interest: An article of importance in its field
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.