

Norms for creativity and implementation in healthcare teams: testing the group innovation inventory

MATHILDE M.H. STRATING* AND ANNA P. NIEBOER

Institute of Health Policy and Management, Erasmus University Rotterdam, The Netherlands

*Address reprint requests to: Mathilde M.H. Strating, Institute of Health Policy and Management, Erasmus University Rotterdam, PO Box 1738, 3000 DR Rotterdam, The Netherlands. Tel: +31-10-408-8557; Fax: +31-10-408-9094; E-mail: strating@bmg.eur.nl

Accepted for publication 9 May 2010

Abstract

Objective. To test to what extent the four-factor structure of the group innovation inventory (GII) is confirmed for improvement teams participating in a quality improvement collaborative.

Design. Quasi-experimental design with baseline and end-measurement after intervention.

Setting. This study included quality improvement teams participating in the Care for Better improvement programme for home care, care for the handicapped and the elderly in the Netherlands between 2006 and 2008.

Participants. As part of a larger evaluation study, 261 written questionnaires from team members were collected at baseline (pre-project sample) and 129 questionnaires at end-measurement (post-project sample).

Main outcome measure. Group innovation inventory.

Results. Confirmatory factor analyses revealed the expected four-factor structure and good fit indices. The subscales 'group functioning' and 'speed of action' showed acceptable Cronbach's alphas and high inter-item correlations. The subscales 'support for risk taking' and 'tolerance of mistakes' showed insufficient reliability and validity.

Conclusions. The group functioning and speed of action subscales of the GII showed acceptable psychometric properties and are applicable to quality improvement teams in health care. In order to understand how social expectations within teams working in health care organizations exert influence over attitudes and behaviours thought to stimulate creativity, further conceptualization of the norms for enhancing creativity within health care is needed.

Keywords: innovation, quality collaborative, healthcare teams, creativity, implementation

Introduction

Quality improvement collaboratives (QICs) have received substantial attention as one way to close the gap between best practices and actual practices in health care. To improve a specific subject area of care, temporary teams from different organizations are brought together in a QIC so that learning within and between settings can take place. QICs are expected to enhance quality and efficiency of care by acting as a 'learning laboratory' [1] stimulating and implementing innovations.

West and Farr [2] defined innovation as: '... the intentional introduction and application within a role, group or organization of ideas, processes, products or procedures, new to the relevant unit of adoption, designed to significantly benefit the individual, the group, the organization or wider

society' (p. 9). Innovation success within organizations depends upon a wide range of determinants on the individual, team and organizational levels. Given the increasing use of teamwork in healthcare organizations [3] and given the fact that innovation is often 'originated and subsequently developed by a team into routinized practice within organizations' [4], there is a need for better understanding of how team processes can facilitate or hinder innovation.

In their generally accepted definition mentioned above, West and Farr [2] distinguished idea generation—or creativity—from implementation. Following this distinction, West and Farr [2] argued that mechanisms facilitating or hindering these two aspects of innovation may differ. Determinants of creativity might not be identical to determinants of implementation and might even have opposite effects. The question remains which

aspects of team processes support creativity and which aspects support implementation.

Research by West and Farr [2] suggests that group climate, defined as a set of shared expectations, is key to a group's scope of new ideas and working methods. These researchers further developed this notion into a model that proposes that group innovations are influenced by four factors: degree of agreement upon clear and realistic objectives, participation in decision-making, commitment to achieve the highest possible standards of task performance and support for attempts of innovative ideas. Although West and Farr recognized the dual nature of innovation, the four-factor model does not distinguish between factors influencing creativity and factors influencing implementation. In their attempt to identify which aspects of the climate within teams facilitate creativity and which aspects facilitate implementation, Caldwell and O'Reilly [5] suggested that social expectations of team members—or group norms—may exert control over attitudes and behaviour by representing 'what is' or 'ought to be' in a particular situation. On the one hand, this mechanism might be conducive to creativity, for example by generating social approval when trying new ways of doing things, taking risk and tolerating mistakes. On the other hand, it could facilitate implementation by generating social approval when working together effectively and acting quickly. It would follow, therefore, that 'support for risk taking' and 'tolerance of mistakes' are two important norms for creativity, and that 'group functioning' and 'speed of action' are crucial for implementation.

In line with this concept, Caldwell and O'Reilly developed the group innovation inventory (GII) to assess these four factors. The four-factor structure underlying the 36 items was theoretically consistent with previous research, and predictive validity was shown by significant correlations between the four subscales and rated innovativeness. The items were based on input from more than 2000 managers and tested on a sample of participants in a university-based management development programme and a part-time MBA programme.

The assumptions underlying the GII may not all be valid, however, for health care settings, notably with regard to the norms that are thought to enhance creativity. In health care, the willingness to propose new and creative solutions to problems—with unknown effects and risks—may be problematic in particular. The challenge here is to find a balance between demands placed on professionals, such as responsibility for quality of care and patient safety, and the necessity of constant learning, improving and innovating. The purpose of the present study was to investigate to what extent the concepts of norms for implementation and creativity can be applied to teams participating in a QIC within health care. We tested whether the four-factor structure underlying the GII was confirmed within this setting.

Method

Participants

This study included members of teams participating in the QIC 'Care for Better', a programme for home care, care for

the handicapped and the elderly in the Netherlands between 2006 and 2008. These improvement teams were participating in the following projects: pressure ulcers, eating and drinking, prevention of sexual abuse, medication safety, fall prevention, aggression and behavioural problems and autonomy. As the major instrument to quickly spread evidence-based practices across care organizations and to enable mutual learning across sites, the 'Breakthrough Series' approach developed by the Institute of Healthcare Improvement was used [6, 7]. Although the topics of improvement were different for these projects, the set up of the projects, working with the plan-do-study-act cycle and starting off with small-scale changes first, is the same.

Teams typically consisted of a project leader and four others. As part of a larger overall evaluation study, team members received a postal questionnaire at two time points: two months into the project (baseline) and after 1 year, at the end of each project (end-measurement). For this study, data from two separate samples were used. The first pre-project sample consisted of baseline data for ongoing projects (no end-measurement data available yet). Eighty-six of the 125 project leaders completed the baseline questionnaire (response rate 68.8%). In total, 219 other team members completed the questionnaire. The exact response rate for the other team members cannot be established, since we do not know the size of teams whose project leader did not complete the questionnaire. For the other teams, the average response of team members was 62%. As 44 respondents had not fully completed the GII, a total sample of 261 respondents was left for analysis.

The second sample is used to cross-validate the factor solution. This post-project sample consisted of end-measurement data only, for several projects that had already started before this evaluation study went underway. Thirty-eight of the 83 project leaders completed the questionnaire (response rate 45.8%). This lower response rate may partly be due to the fact that the teams participating in projects on pressure ulcers, eating and drinking and prevention of sexual abuse had not been informed beforehand about the evaluation study. In total, 98 other team members completed the questionnaire. As 7 respondents had not fully completed the GII, a total sample of 129 respondents was left for analysis.

Measures

Group innovation inventory. The 36 original items had been translated into Dutch by two researchers independently. There were no salient differences in meaning between the two translations, and the two researchers agreed upon the final Dutch translation. Each item was rated at a five-point scale ranging from 'strongly disagree' to 'strongly agree', in which higher scores indicate a better or more desirable team climate. Scores for each item in a subscale were summed to determine the subscale score.

Team climate inventory [8]. The short version of the team climate inventory (TCI) [9, 10], a well-validated instrument, served to validate the GII. The TCI consists of four subscales: vision, participative safety, task orientation and

support for innovation. Items included statements such as 'People in this team are always searching for fresh, new ways of looking at problems'. The 14 items were rated on a five-point scale ranging from strongly disagree to strongly agree, in which higher scores indicate a better or more desirable team climate. Scores for each item in a subscale were summed to determine the subscale score. Reliabilities of the four subscales in our study were between 0.77 and 0.80.

Analyses

The psychometric analyses comprised three parts. First, to verify the factor structure of the questionnaire and to test whether the relationship between observed variables and their underlying latent constructs exists, confirmatory factor analysis was executed using the LISREL program [11]. No correlation errors either within or across sets of items were allowed in the model. Based on the four-factor solution found by Caldwell and O'Reilly, each subset of items was allowed to load only on its corresponding latent construct derived from the four-factor theory (boldfaced items in Table 1 on page 507 of Caldwell and O'Reilly) [5]. The 10 items that cross-loaded on more than one factor in the analysis by Caldwell and O'Reilly were allowed to load on all four factors (items 4, 13, 17, 20, 23–26, 33 and 34). The factor loadings of these 10 items were compared and the modification indices were used to investigate of which latent construct the items were indicators. In the second model, the model was improved by eliminating items that cross-loaded on more than one factor or had factor loadings lower than 0.20.

In the second part, item-reduction analysis was performed to develop a short version of the questionnaire that can be used in case the original version is considered to be too long. Items were removed from the original pool following several criteria: (i) items were excluded one by one following modification indices provided by LISREL and the strength of the loadings, (ii) elimination of items was stopped when reliability of each subscale drops below 0.70 and (iii) there should be as few items as possible with a minimum of four, without loss of content and psychometric quality. To test the measurement models, four indices of model fit were used. The cut-off criteria for these four indices were those proposed by Hu and Bentler [12]. First, the overall test of goodness-of-fit assesses the discrepancy between the model implied and the sample covariance matrix by means of a normal theory weighted least squares test. A plausible model has low, preferably non-significant χ^2 values. However, χ^2 is overly sensitive when the sample size is large (anything over 200 [13]), leading to difficulty in obtaining desired non-significant levels [14]. Second, the root means square error of approximation (RMSEA) reflects the estimation error divided by the degrees of freedom as a penalty function. Values on RMSEA below 0.06 indicate small differences between the estimated and observed model. Third, we used the standardized root means square residual (SRMR), which is a scale invariant index for global fit that ranges between 0

and 1. Values on SRMR lower than 0.08 indicate a good fit. As a fourth index of model fit, the incremental fit index (IFI) was calculated. This index compares the independence model (i.e. observed variables are unrelated) to the estimated model. Preferably, values on IFI should be larger than 0.95.

In the third part, internal consistency of the subscales was assessed by calculating Cronbach's alphas. Since we expected the two factors group functioning and speed of action (norms for implementation) to be distinct from the two factors support for risk taking and tolerance of mistakes (norms for creativity), correlations between the four factors were computed. In order to further investigate the validity of the GII, correlations of each subscale with the four subscales of the TCI were calculated. Although the TCI was developed to measure the four factors that together cover the concept of team climate and influence group innovations, previous studies on the TCI do not clarify which of the four factors may facilitate implementation and which may facilitate creativity. Taking into account the content of the items, vision, participative safety and task orientation could be attributed to factors facilitating implementation, whereas support for innovation could be attributed as a factor stimulating creativity.

Results

Sample characteristics

The majority of the team members that filled in the baseline and end-measurement questionnaire was female. Mean age was 44 years (SD 9.8) for the pre-project sample and 43 years (SD 9.9) for the post-project sample. Table 1 lists descriptive characteristics of the two samples of team members. In both samples, more than two-thirds of the team members had been working for more than 3 years within the organization. Furthermore, 158 (60.5%) team members at baseline and 88 (68.2%) at end-measurement worked more than 29 h per week. Teams mainly consisted of nurses and caregivers and management.

Part I: confirmatory factor analysis with 36 items

The factor loadings found in our study showed several differences compared with the results of Caldwell and O'Reilly [5]. Some items had rather low factor loadings on the intended factor. Standardized loadings of the items are shown in Table 2. The indices of model fit also showed that the model fit was insufficient (see Table 3, Model 1). The significant normal theory weighted least square χ^2 statistic is not surprising given its sensitivity to sample size; it was 1742.803. The RMSEA was 0.05 and below cut-off value. IFI was equal to the cut-off value of 0.95 and SRMR was with a value of 0.09 near the cut-off value of 0.08. All indices indicated that the model could be improved.

These results showed that the highest loadings of items 13, 23 and 24 are on speed of action (Factor 2); these items did not load on more than one factor. The items 17, 20 and

Table 1 Total sample characteristics

	Pre-project sample (<i>n</i> = 261)		Post-project sample (<i>n</i> = 129)	
	<i>N</i>	%	<i>N</i>	%
Sex				
Female	213	83.2	96	75
Male	43	16.8	32	25
Works in current position for more than 3 years	212	81.5	105	81.5
Works more than 29 h	158	60.5	88	68.2
Position				
Medical assistants	5	2.0	7	6.0
Nurses and caregivers	57	23.2	17	14.5
Social workers	19	7.7	1	0.9
Medical/social specialists	30	12.2	14	12.0
Management	84	34.1	54	46.2
Health policy and quality staff	30	12.2	10	8.5
Para-/perimedical professionals	21	8.5	13	11.1
Clients	–	–	1	0.9

Pre-project sample consisted of baseline data for ongoing projects; post-project sample consisted of end-measurement data only.

26 had their highest factor loadings on group functioning (Factor 1). Since the wording of these items confirmed these high factor loadings, the pathways to the other latent constructs were eliminated from the measurement model. Item 25 loaded on group functioning and on speed of action. Since this item loaded higher on speed of action (0.62) than on group functioning (0.46) and on grounds of construct uniformity, item 25 should be attributed to speed of action. To improve the measurement model, the other items that cross-loaded (4, 33 and 34) as well as items with factor loadings lower than 0.20 (36, 7, 6, 18 and 32) were eliminated.

Elimination of these items resulted in 28 remaining items, with 10 items measuring group functioning, 9 items measuring speed of action, 5 items measuring support for risk taking (Factor 3) and 4 items measuring tolerance of mistakes (Factor 4). The four indices of model fit showed that the model improved (Table 3, Model 2), RMSEA was 0.05 and below 0.06 and IFI was 0.97, both indicating good fit. The normal theory weighted least square χ^2 decreased to 1116.663 but was still significant ($P = 0.0$). The SRMR index was 0.09, exceeding the cut-off point of 0.08. This indicates that the global fit of the overall model is not yet sufficient, pointing to validity problems that may be caused by the two subscales with lower reliability.

Part 2: item-reduction analysis

Since the subscales support for risk taking and tolerance of mistakes already consisted of only a few items and had low reliability (0.64 and 0.45, respectively), we focused on shortening the group functioning and speed of action subscales.

The results from the stepwise procedure showed that after eliminating items 1, 12 and 23, no additional items could be eliminated, since reliability for 'speed for action' would drop

below 0.70. The reliability of the subscale speed for action with six items was 0.70. The results from the stepwise procedure also showed that items 10, 28, 14, 20, 2 and 8 of the group functioning subscale could be eliminated. With the remaining four items, the subscale had a reliability coefficient of 0.79. Further reduction would reduce the number of items to three.

The overall fit of this final model was further improved (Table 3, Model 3). The normal theory weighted least square χ^2 decreased to 510.252, RMSEA was 0.04, which is far below the cut-off point of 0.06, and the value of IFI was 0.98, indicating that the specified relations between variables are supported by the data. The SRMR index decreased to 0.08, which equals the cut-off point of 0.08 and indicates that the global fit of the overall model is sufficient.

The post-project sample was used to cross-validate the factor solution. In Table 3, model fit of the different models for this sample are presented. These analyses showed similar results with respect to factor loadings and the number of items to eliminate as a result of multiple loading or low factor loadings. The normal theory weighted least square χ^2 started off higher than the one based on the pre-project sample with a value of 2806.60, but decreased in Models 2 and 3 to a value of 564.89. As was shown for the pre-project sample, the RMSEA and IFI based on the post-project sample also indicate good-to-moderate fit. Only the SRMR showed less sufficient fit with a value of 0.14 in Models 2 and 3. Across all fit indices, the post-project sample validates the factor solutions found in the pre-project sample.

Part 3: internal consistency and inter-correlations

The high cross-scale correlations between the shortened and original scale indicate acceptable coverage of the core

Table 2 Standardized loadings of the 36 items in confirmatory factor analysis (pre-project sample $n = 261$)

Questionnaire items	Factor loadings			
	Factor 1	Factor 2	Factor 3	Factor 4
1 Around here, policies and procedures often stop the implementation of good ideas (R)		0.27		
2 People in our group feel that they are all pulling together for a common goal	0.68			
3 Risk taking is encouraged around here			0.49	
4 In our group, we typically try to avoid or smooth over conflicts or disagreements. (R)	-0.32	0.19	-0.18	0.42
5 Management provides rewards and recognition for innovation and trying new things			0.73	
6 Our group has a strong belief in the importance of hard work			0.13	
7 Responsibility for making things happen is delegated to the lowest level possible		0.15		
8 There is a willingness on the part of people in our group to share the credit for successes with each other	0.70			
9 Mistakes are a normal part of trying something new				0.62
10 People in our group encourage one another to understand how people in other groups look at problems	0.48			
11 Successful innovation is important for career success in this organization			0.32	
12 People in this group are willing to cut through bureaucracy in order to get things done		0.43		
13 People have great freedom to act to make necessary changes around here	-0.04	0.70	0.24	-0.26
14 There is a lack of teamwork in our group (R)	0.56			
15 The attitude around here is that when you are trying new things, mistakes are a normal part of the job				0.52
16 In our group, there is a great deal of openness in sharing information	0.80			
17 People in our group encourage each other to try new things	0.74	0.06	0.10	0.01
18 Meeting deadlines is very important in this group (R)				0.21
19 Decisions in our group are made quickly		0.49		
20 There are mixed messages about what is important in our group (R)	0.62	0.11	-0.04	-0.10
21 Management encourages people to try new things			0.87	
22 Members of our group listen carefully to the views of others	0.73			
23 People are given the time and resources to innovate	-0.12	0.41	0.25	0.01
24 In our group, we expect others to take initiative and get things done even if a person is not formally responsible	-0.08	0.55	0.01	0.04
25 Our group is flexible and adapts quickly to new opportunities	0.46	0.62	0.00	-0.29
26 In our group, we try to reach a consensus about important decisions	0.76	0.19	0.07	-0.21
27 People feel that it is important to challenge the status quo				0.35
28 People in our group have a difficult time accepting criticism (R)	0.51			
29 In general, it is better to be safe than sorry around here (R)				0.41
30 Once a decision is made, we implement it quickly		0.59		
31 Our group has sufficient autonomy to implement new ideas without clearance from above		0.48		
32 There is a shared vision about what we are trying to accomplish here	0.14			
33 In general, people have a positive attitude about the need to continuously change	-0.71	0.56	-0.12	0.56
34 Conflict is expected and accepted as a normal part of getting things done	-0.80	0.40	-0.21	1.00
35 The organization invests enough in training and updating people's skills			0.43	
36 Entrepreneurial skills are important if you are going to be successful in this group			0.06	

R, reversed item.

areas of the four-factor theory (Table 4, column 4). The four subscales were significantly and positively correlated (Table 4, columns 5, 6 and 7). Between the subscales group functioning and speed of action, reflecting norms for implementation, a correlation of 0.43 was found and this indicates that the subscales are conceptually related. Correlations between group functioning and speed of action with support for risk taking and tolerance of mistakes ranged from 0.14 to 0.40.

Table 3 Model fit of the model for each of the two study samples

	χ^2 (P-value)	RMSEA	IFI	SRMR
Pre-project sample ($n = 261$)				
Model 1: 36 items	1742.80 (0.000)	0.05	0.95	0.09
Model 2: final full version	1116.66 (0.000)	0.05	0.97	0.09
Model 3: final short version	510.25 (0.000)	0.04	0.98	0.08
Post-project sample ($n = 129$)				
Model 1: 36 items	2806.60 (0.000)	0.05	0.96	0.13
Model 2: final full version	1565.02 (0.000)	0.06	0.96	0.14
Model 3: final short version	564.89 (0.000)	0.07	0.96	0.14

RMSEA, root means square error of approximation; IFI, incremental fit index; SRMR, standardized root means square residual.

As a final step in the validation of the GII, correlations of each subscale with the four subscales of the TCI were analysed within the pre-project sample (Table 5). Except for the correlation between support for risk taking and 'vision', the four subscales of the GII correlated significantly and positively with the four subscales of the TCI. The highest correlation (0.70) was found between group functioning and 'participative safety'. Overall, the correlations between group functioning and speed of action with the TCI subscales were higher than the correlations of support for risk taking and tolerance of mistakes with the TCI subscales. Although support for innovation can be seen as a factor stimulating creativity, correlations with support for risk taking and tolerance of mistakes are moderate.

Discussion

The aim of the study was to investigate the factor structure of the GII within health care. The four dimensions underlying the GII were identified in our study. Group functioning and speed of action—reflecting norms for implementation—were found to be moderately strong scales. This indicates that, also within health care, we can assess the extent to which group norms support cooperation and exchange of information among members of improvement teams, as well as the presence of a shared sense of the need to accomplish things quickly. However, our findings suggest that norms associated with enhancing creativity (e.g. support for risk taking and a willingness to tolerate mistakes) are difficult to conceptualize in health care.

A similar picture arises from the comparison of the four short subscales of the GII with the short version of TCI. Correlations between these subscales showed that the GII

Table 4 Cronbach's alpha and inter-correlations of the four subscales (pre-project sample $n = 261$)

	Number of items	Items	Alpha	Original scale	1	2	3
Group functioning	4	16, 17, 22, 26	0.79	0.86**			
Speed of action	6	13, 19, 24, 25, 30, 31	0.70	0.93**	0.43**		
Support for risk taking	5	3, 5, 11, 21, 35	0.64	1.00**	0.14*	0.32**	
Tolerance of mistakes	4	9, 15, 27, 29	0.45	1.00**	0.40**	0.29**	0.21**

** $P < 0.01$ (one-tailed).

Table 5 Inter-correlations with the four subscales of the team climate inventory (pre-project sample $n = 261$)

	Vision	Participative safety	Task orientation	Support for innovation
Group functioning	0.52**	0.70**	0.61**	0.47**
Speed of action	0.28**	0.44**	0.42**	0.34**
Support for risk taking	0.07	0.19**	0.26**	0.26**
Tolerance of mistakes	0.23**	0.35**	0.39**	0.34**

** $P < 0.01$.

construct of group functioning compares well with the TCI construct of 'participative safety'. Both subscales include items on information sharing and participation in decision-making, thought to be important for enhancing implementation. Still, 'support for innovation' (TCI)—a stimulating factor for creativity—only moderately correlated with the GII creativity subscales support for risk taking and tolerance of mistakes. A possible explanation is the fact that some items in the GII support for risk taking subscale also address issues that tap more individual rather than group goals, such as 'successful innovation is important for career success'.

The results with regard to the norms for enhancing creativity show that further conceptualization is needed to understand how these norms may exert influence within teams working in health care organizations. The assumption underlying these norms was that support for risk taking and tolerance of mistakes are inherent in a social control system that exerts influence over attitudes and behaviour conducive to innovation. This assumption may be less applicable to health care settings, however, with their high degree of work complexity. In generating new ideas and trying out new things, errors are inevitable. Health care professionals are likely to experience these as dilemmas, seeking to find a balance between tolerance of mistakes and patient safety. Especially when embedded in a culture of 'blaming and shaming', health professionals may be inclined to prefer safe working methods [15]. The social expectations of their peers (informal control), as well as formal control systems do not allow health professionals to make mistakes since these can be harmful or fatal to clients.

Especially in health care, 'error' is a contested concept. Terms such as mishaps, mistakes, errors or failures imply that there has been some inappropriate behaviour conveying a negative judgemental meaning [16]. The use of particular terms in the GII items and their connotations may have led to inconsistent responses, which could explain the low internal consistency and validity found for support for risk taking and tolerance of mistakes scales. Robustness and validity of these subscales could be improved by adapting the items in terms of risks and mistakes that are relevant and realistic in health care settings. To assess these norms for creativity within healthcare, we believe that asking health professionals to think about their actual work practices instead of more abstract terms as 'new things' may lead to more valid answers. An example being questions that relate to health professionals' resourceful attempts to satisfy individual needs or wishes of clients.

In addition to the theoretical considerations discussed above, also some methodological considerations of our study should be addressed. On the one hand, professionals participating in quality improvement teams usually already have experience working together as co-workers from the same division; on the other hand, they usually had not worked together before in such an improvement team, which may make it difficult for them to respond to items of the questionnaire on group processes. Therefore, we deliberately sent the baseline questionnaire no earlier than 2 months after start of each project. Especially in the starting phase,

improvement teams meet regularly to think about what their strengths and weaknesses are, what their specific team targets are and which improvement actions should be taken. This should give them enough input to give a valid answer to the items of the questionnaire.

Another methodological limitation is the lower response of team members on our end-measurement questionnaire. Given the dynamics in the field, not many respondents were available for this study.

The testing of theoretical associations between constructs such as group norms can be analysed at the team level taking into account the hierarchical structure of the data for individuals nested within teams. As there is the potential for considerable variation within teams and since the main purpose of our study was to compare the psychometric properties of the GII in quality improvement teams with those from the previous study of Caldwell and O'Reilly, we performed confirmatory factor analyses on the individual level. Ignoring the hierarchical structure of the data may lead to a worse fit of the model [17, 18]. The factor loadings found with the two methods (individual versus team level) will be similar in value.

Conclusion

In conclusion, the full and short subscales group functioning and speed of action showed acceptable psychometric properties. They can be used to assess how quality improvement teams in health care experience norms conducive to implementation. The two subscales support for risk taking and tolerance of mistakes showed insufficient reliability and validity. Therefore, we need to further conceptualize the norms for enhancing creativity in order to understand how social expectations within teams working in health care organizations exert influence over attitudes and behaviours thought to stimulate creativity.

Acknowledgements

The authors thank the participating improvement teams and respondents to the questionnaire.

Funding

The research was supported by a grant provided by the Netherlands Organisation for Health Research and Development (ZonMw, grant number 5942).

REFERENCES

1. Senge P, Scharmer CO. Community action research. In: Reason P, Bradbury H (eds). *Handbook of Action Research*. California: Sage Publications, 2001.

2. West MA, Farr JL. *Innovation and Creativity at Work: Psychological and Organizational Strategies*. Chichester: Wiley, 1990.
3. Heinemann GD, Zeiss AM. *Team Performance in Health Care: Assessment and Development*. New York, Boston, Dordrecht: Kluwer Academic/Plenum Publishers, 2002.
4. Anderson NR, West MA. Measuring climate for work group innovation: development and validation of the team climate inventory. *J Organ Behav* 1998;**19**:235–58.
5. Caldwell DF, O'Reilly CA. The determinants of team-based innovation in organisations. The role of social influence. *Small Group Res* 2003;**34**:497–517.
6. Plsek PE. Collaborating across organizational boundaries to improve the quality of care. *Am J Infect Control* 1997;**25**:85–95.
7. Kilo C. A framework for collaborative improvement: lessons from the Institute of Healthcare Improvement's Breakthrough Series. *Qual Manag Health Care* 2001;**6**:13.
8. Anderson N, West MA. *Team Climate Inventory: Manual and User's Guide*. Windsor: NFER-Nelson, 1994.
9. Kivimaki M, Elovainio M. A short version of the Team Climate Inventory: development and psychometric properties. *J Occup Organ Psychol* 1999;**72**:241–6.
10. Strating MM, Nieboer AP. Psychometric test of the Team Climate Inventory—short version investigated in Dutch quality improvement teams. *BMC Health Serv Res* 2009;**9**:126.
11. Jöreskog K, Sörbom D. *User's Reference Guide*. Chicago: Scientific Software International, 1996.
12. Hu L, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. *Struct Equ Modeling* 1999;**6**:1–55.
13. Hayduk LA. *Structural Equation Modeling with LISREL: Essentials and Advances*. Johns Hopkins University Press, 1987.
14. Bagozzi RP, Yi Y, Phillips LW. Assessing construct validity in organizational research. *Adm Sci Q* 1991;**36**:
15. Rushmer R, Kelly D, Lough M *et al*. Introducing the learning practice—I. The characteristics of learning organizations in primary care. *J Eval Clin Pract* 2004;**10**:375–86.
16. Quick O. Outing medical errors: questions of trust and responsibility. *Med Law Rev* 2006;**14**:22–43.
17. Muthén BO. Multilevel covariance structure analysis. *Sociol Methods Res* 1994;**22**:376–98.
18. Dyer NG, Hanges PJ, Hall RJ. Applying multilevel confirmatory factor analysis techniques to the study of leadership. *Leadership Q* 2005;**16**:149–67.

Appendix

Final short version of the GII

Group functioning

16. In our group, there is a great deal of openness in sharing information.
17. People in our group encourage each other to try new things.
22. There are mixed messages about what is important in our group. (R).
26. In our group, we expect others to take initiative and get things done even if a person is not formally responsible.

Speed of action

13. People have great freedom to act to make necessary changes around here.
19. Decisions in our group are made quickly.
24. In our group, we expect others to take initiative and get things done even if a person is not formally responsible.
25. Our group is flexible and adapts quickly to new opportunities.
31. Once a decision is made, we implement it quickly.
30. Our group has sufficient autonomy to implement new ideas without clearance from above.

Risk taking

3. Risk taking is encouraged around here.
5. Management provides rewards and recognition for innovation and trying new things.
11. Successful innovation is important for career success in this organization.
21. Management encourages people to try new things.
35. The organization invests enough in training and updating people's skills.

Tolerance of mistakes

9. Mistakes are a normal part of trying something new.
15. The attitude around here is that when you are trying new things, mistakes are a normal part of the job.
27. People feel that it is important to challenge the *status quo*.
29. In general, it is better to be safe than sorry around here. (R).