Leadership in Health Services

A randomised study of leadership interventions for healthcare managers

Caroline Lornudd David Bergman Christer Sandahl Ulrica von Thiele Schwarz

Received 18 June 2015
Revised 15 December 2015
Accepted 22 March 2016

A randomised study of leadership interventions for healthcare managers

Caroline Lornudd, David Bergman, Christer Sandahl and Ulrica von Thiele Schwarz
Department of Learning, Informatics, Management, and Ethics, Karolinska Institute, Stockholm, Sweden

Abstract

Purpose – The purpose of this paper was to assess two different leader development interventions by comparing their effects on leadership behaviour and evaluating their combined impact after two years, from the viewpoints of both the participating managers and external raters.

Design/methodology/approach – The study was a longitudinal randomised controlled trial with a cross-over design. Health care managers (n = 177) were first randomised to either of two 10-month interventions and a year later were switched to the other intervention. Leadership behaviour was rated at pre-test and 12 and 24 months by participating managers and their superiors, colleagues and subordinates using a 360-degree instrument. Analysis of variance and multilevel regression analysis was performed.

Findings – No difference in effect on leadership behaviour was found between the two interventions. The evaluation of the combined effect of the interventions on leadership behaviour showed inconsistent (i.e. both increased and decreased) ratings by the various rater sources.

Practical implications – This study provides some evidence that participation in leadership development programmes can improve managers' leadership behaviours, but the results also highlight the interpretive challenges connected with using a 360-degree instrument to evaluate such development.

Originality/value – The longitudinal randomised controlled design and the large sample comprising both managers and external raters make this study unusually rigorous in the field of leadership development evaluations.

Keywords Evaluation, Healthcare, Leadership, Development, RCT, 360-instrument

Paper type Research paper

Introduction

Being a health-care manager appears to be an increasingly complex occupation. In many Western countries, the growing pressure to increase efficiency and lower costs has led to considerable structural changes in health-care organisations in recent decades (Berntson *et al.*, 2012). Furthermore, managers in public organisations have to balance conflicting political and operational demands, as well as constraints on autonomy

The authors wish to express gratitude for the financial support from AFA Insurance, grant no. 100073, which made this study and analyses possible. The authors are also grateful for support from Mikael Ohrling, CEO at Stockholm County Council (SCC). Further, they would like to thank Therese Wahlström, Karolinska Institute, for her help with data collection. Finally, the authors would like to thank all respondents for taking their time to answer the questionnaires.

Leadership in Health Services
Vol. 29 No. 4, 2016
pp. 358-376

1751-1879
DOI 10.1108/LHS-06-2015-0017

(Rainey and Chun, 2005). Most scholars would certainly agree that managerial leadership in health care is difficult to handle and achieve (Plsek and Greenhalgh, 2001). However, the ability to lead is generally seen as a skill that can be developed (Day *et al.*, 2014). Although such development has been shown to have a positive impact (Avolio *et al.*, 2009), there is still a shortage of well-designed studies of leadership interventions that can illustrate long-term effects.

Leader development programmes generally produce moderately positive effect sizes but with great variation between studies (Avolio *et al.*, 2009, Collins and Holton, 2004). Further, leader development research is largely based on short interventions (Avolio *et al.*, 2009, Kelloway and Barling, 2010). In a meta-analysis, Avolio *et al.* (2009) found that the median length of interventions was 3-6 h, and only 9 per cent of the included studies had investigated interventions lasting longer than seven days. A review by Kelloway and Barling (2010) indicated that short-term interventions typically comprise workshop training activities, aimed at enhancing a pre-determined leadership skill. Nevertheless, short interventions targeting proven approaches to solve known problems might be ill-suited for addressing the demands facing leaders today (Day *et al.*, 2014), and positive findings have been reported by investigations of leadership development programmes with other approaches, for example, art-based leadership programme (Romanowska *et al.*, 2011). In the health care context, it is plausible that the complex challenges that managers must handle are neither easy to frame nor simple to solve by focusing on practicing specific leadership behaviours. The outlined conclusions identify the need to evaluate a long-term intervention targeting complex leadership challenges for health care managers, which was the aim of the present study.

Leadership is often regarded as an ability that is primarily acquired by continually learning from experience (Jackson and Parry, 2008, Yukl, 2012). However, to transform experience into conscious knowledge, it is necessary to reflect (Kolb, 1984, Lewin, 1951). In the theory of experiential learning (Kolb, 1984), knowledge is built through the transformation of personal experience via reflection, abstract conceptualisation and active experimentation. Recent promising evaluations have indicated that an approach based on experiential learning may effectively support leadership. In backstage groups (also known as dialogue groups), the participants (e.g. managers) meet regularly over a longer period of time in a setting that exploits their ability to learn from experiences (Bergman *et al.*, 2009a, 2009b, Sandahl *et al.*, 2007). Such studies suggest that backstage groups can broaden the perspective of the leader role, strengthen the understanding of the overall purpose of their organisation and enhance the participants' self-confidence (Bergman, 2009).

However, in previous studies evaluating backstage groups, participants also requested more theoretical content to support their personal insights (Sandahl *et al.*, 2007). In leader development, the theoretical content normally includes leadership theory and/or theories tied to other relevant themes, such as communication and group psychology. Although there is an ongoing trend towards creating leader development embedded in a work context (Day, 2001), theoretical models may still play an important role by facilitating conceptual understanding of complex phenomena (Jackson and Parry, 2008). Thus, two different methods of supporting leader development can be discerned: groups emphasising reflection on experience and leadership training with theoretical content as a cornerstone. Little is known about the comparative and

combined long-term effects of these two strategies. Therefore, the leadership intervention programme for the present study was designed to enable a comparison of the two approaches and evaluation of the long-term effects of a combination of both methods.

Traditionally, the model designed by Kirkpatrick (1975) has been used to evaluate the effectiveness of training programmes (Salas *et al.*, 2012) with outcome measures on different levels:

- the participants' reactions to the programme;
- achieved learning outcomes;
- degree of transfer of behaviour from training setting to workplace setting; and
- indirect organisational outcomes.

Considering backstage groups, the first level (reactions) and the second level (learning outcomes) have been explored in a few studies, showing positive results in participants' self-ratings concerning ability to understand their own emotions and reactions, to manage up-coming situations and to clearly define their managerial role (Bergman *et al.*, 2007, Bergman *et al.*, 2009a, 2009b, Björklund, 2009). However, the third level (behaviour transference to workplace) would provide an indication of the effectiveness of the interventions beyond the participants' attitudes. A potential difference between pre-test and post-test ratings of leadership behaviour can be subjectively appraised by the participants themselves. Nevertheless, the ultimate focus of leader development normally includes indirect effects on the organisation, for example, the well-being or efficiency of subordinates (Nielsen *et al.*, 2010). Thus, it is also of interest to assess external raters' perceptions of changes in leadership behaviour because such observations should precede potential changes in well-being and motivation (Kelloway and Barling, 2010). One way of investigating these outcome levels is to use a 360-degree (multi-source) instrument that assesses leadership behaviour from both the perspective of the managers and the perspective of external raters.

For a leader development intervention to be deemed effective in health care, it should preferably affect behaviour related to the managerial responsibilities – the strategic, administrative and personnel objectives (Wikström and Dellve, 2009). These objectives are largely captured by the a three-factor structure of leadership behaviour in the change, production and employee (CPE) leadership model, which has been validated in managerial samples across countries (Ekvall and Arvonen, 1991, Yukl *et al.*, 2002) and across branches, including health care (Larsson and Vinberg, 2010, Sellgren *et al.*, 2008). Hence, we chose to use a validated 360-degree instrument based on the CPE model to evaluate a leader development programme for health care managers.

The objective was to determine whether the ratings of each leadership behaviour orientation made pre-test differed from the following:

(1) ratings at 12-month post-test, also considering any differences between the part of the programme emphasising theory and the part emphasising reflection on experience reflected in the ratings made by:
   - the participating managers, or
   - superiors, colleagues and subordinates; and

(2) ratings at 24-month post-test, including both the part of the program emphasising theory and the part emphasising reflection on experience, as indicated in ratings made by:
- the participating managers, or
- superiors, colleagues and subordinates.

Method

This longitudinal experimental study was conducted within the Stockholm County Council (SCC), which runs nine organisations that provide institutional and non-institutional care in a catchment area with a population of 2 million. The study was approved by the Regional Ethical Review Board (reg. no. 2010/979-31/5).

Procedure

In September 2010, January 2011 and September 2011, the SCC invited all their health care managers ($n = 589$) to participate in a two-year leader development programme that was to be evaluated in collaboration with Karolinska Institutet. The managers that accepted the invitation received information by email about the study before the onset of the programme. They were also informed that they would be randomised to either of two types of interventions (theoretical seminars or backstage groups) during the first year and to switch to the other type of intervention after one year in a cross-over design; thus, by the end of the second year, all the included managers were to have participated in both interventions. Participation in both the development programme and the study was voluntary, and managers could take part only in the development programme if they wished to do so.

Confidentiality and anonymity were guaranteed, and informed consent was obtained. In addition to recruiting managers, for each manager, a superior, up to five colleagues and up to 15 subordinates were also asked to participate in the study by rating the managers. The reliability of a 360-degree instrument is positively correlated to the number of raters included (Hensel *et al.*, 2010); therefore, we chose up to 15 subordinates and up to five colleagues for each manager. The randomisation was performed by the SCC. For each manager, a list of the subordinates' surnames in alphabetical order was entered into a computer and randomised by use of Microsoft Excel 2010; thereafter, the first 15 names on the randomised list were selected. The colleagues were selected in the same manner.

The participating managers and their superiors, selected colleagues and subordinates were invited via email to respond to a web-based survey, including the CPE questionnaire ("Measures" section). The participating managers and their subordinates also completed the Webb-QPS questionnaire ("Measures" section). Data were collected at three time points: at pre-test (Time1) and at post-test; before the start of the second intervention period (Time2) and at a second post-test; and after conclusion of the programme (Time3). The response rate at Time1 was 93 per cent for participating managers, 90 per cent for superiors, 69 per cent for colleagues and 64 per cent for subordinates. The mean number of subordinates responding per manager was 7.3 (range 1-15), and mean number of colleagues per manager was 2.5 (range 1-5). Any dropouts among the external raters were replaced with new respondents recruited by the SCC. The two interventions in the development programme (theoretical seminars and backstage groups) were held concurrently at SCC facilities.

Participants

Of the 225 managers who initially volunteered to participate, 177 completed the full two-year development programme and took part in the present evaluation (Figure 1). In all, 47 of the managers dropped out: 22 before the onset of the programme and an additional 25 during the interventions. For 13 of those managers, the reasons for non-participation were as follows: high work pressure (7), parental leave (2), resignation (2), unclear managerial responsibility (1) and familiarity with another participant (1). For the remaining 34, no information was available regarding reasons for drop out. For 22 of the 25 who left the interventions, demographic data were available for analysis, and χ^2 and Mann–Whitney tests showed no significant difference in age, gender, tenure or educational level between those dropouts and the managers who fulfilled the programme. The 177 who completed the programme represented 30 per cent of all managers in the organisation; 83 per cent of them were women, and the mean age was 50.5 years (range 30-63 years). The SCC managers who declined participation in the programme ($n = 386$) had a mean age of 55 years, and 76 per cent were women. Figure 1 presents a flowchart illustrating the participation throughout the study.

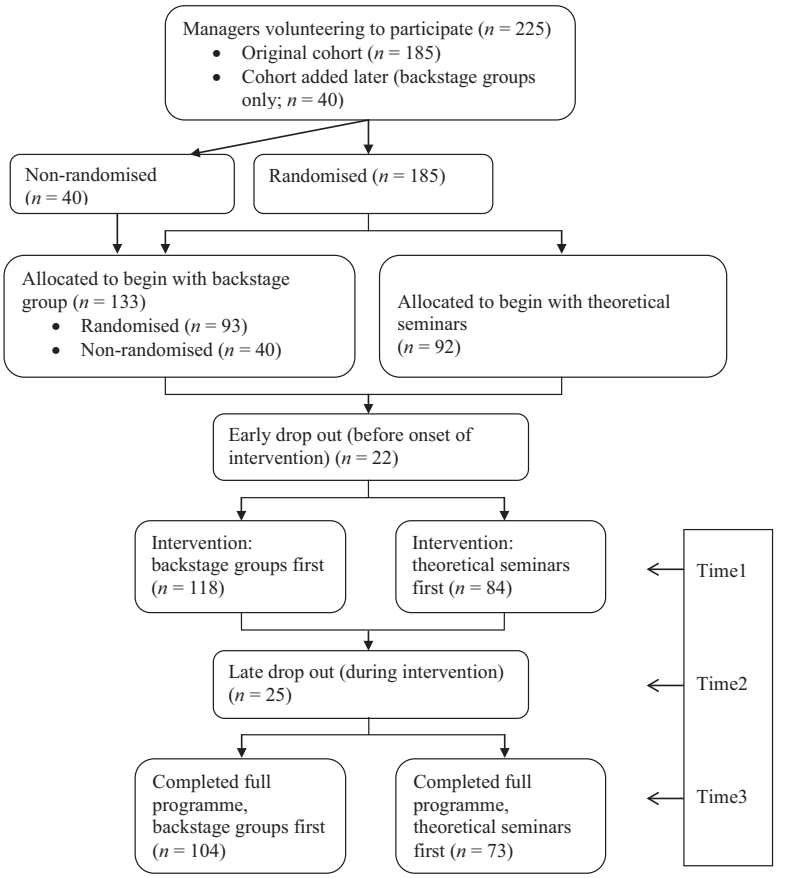

Figure 1.
Flowchart of participating managers and indication of data collection time points; Time1, Time2 and Time3

Panel attrition (i.e. loss of cases between measurement time points) was possible because of the longitudinal design of the study. Table I shows the dropout pattern for the questionnaire respondents, considering all rater sources. To detect systematic differences in the subordinate sample (i.e. the sample of external raters for which these data were available), we created a variable indicating whether a subordinate responded at both Time1 and Time2 ($n = 707$) or only at Time1 ($n = 389$). Next, we performed χ^2 in a cross-table (for gender) and the Mann–Whitney test (for age, tenure and educational level), which satisfactorily indicated no significant difference between the two groups.

Interventions

Here, "backstage group" refers to what Goffman (1959), in his discussion of the role concept, described as how human beings in their presentation of themselves sometimes need to find a place and time to reflect "backstage" on how well they succeed in their different roles. This alludes to the situation in which, after concluding a performance, actors and their director need to reflect on and discuss the performance backstage.

In our study, each backstage group consisted of 7-9 managers and a group leader. The group leaders were experienced first- and second-line managers internally recruited from the SCC. They attended a six-day course on the backstage model and regularly received coaching during the intervention. Each backstage group met for three-hour sessions about once a month for ten months. During the sessions, the managers discussed problems they experienced in their everyday work. The role of the group leader was to maintain the structure and time limits and to facilitate discussions (Sandahl *et al.*, 2012). After the first introductory session, all subsequent sessions followed the same protocol. A session started with a general discussion about what had happened since the previous session and identification of a problem to work on. The person with the problem described it in detail with the help of another group member who interviewed the "problem owner". Then, all the group members except the problem owner discussed how they perceived the dilemma, whereas the problem owner had to remain silent and listen carefully. After some corrections and complimentary information, the next phase entailed sub-groups formulating hypotheses about how the problem might be interpreted. These hypotheses were subsequently introduced to and discussed with the problem owner. Finally, the members of the sub-groups suggested actions to solve the problem and the problem owner chose one solution or discovered another alternative during the process. The session ended with a reflection on what was learned at the session and with meta-reflection on the process of learning.

Each seminar group had 40-60 members, all of whom received a book covering topics related to emotional and ethical challenges in the role as manager (Sandahl *et al.*, 2010). A seminar started with a 45-minute lecture concerning one chapter in the book, and the lecturers were co-authors of that book. The themes of the seminars were as follows:

Table I. Number of respondents of *change, production, employee (CPE)* leadership behaviour ratings by rater source and time-point

Rater source	Time1 *n*	Time2 *n* (no. of respondents left from Time1)	Time3 *n* (no. of respondents left from Time1 and Time2)
Participating managers	168	154 (148)	117 (112)
Subordinates	1096	1,050 (707)	798 (362)
Colleagues	327	281 (216)	222 (122)
Superiors	159	160 (139)	121 (90)

- to be a manager – an emotional challenge;
- feelings in working life;
- the meaning of work;
- group dynamics;
- leadership;
- feelings and ethics;
- to be professional in the role of manager; and
- to develop sense and sensitivity in the role of manager.

A lecture was followed by a 20-30-minute small-group discussion, and the seminar ended with conclusions and action points shared by the whole seminar group.

Measures

Leadership behaviour orientation. Leadership behaviour orientation (i.e. the inclination to behave according to a certain pattern; see below) was measured using the CPE questionnaire, which comprises three orientation subscales: change, production and employee. This instrument is based on the CPE model of leadership behaviour, which was developed using data on 4,000 Swedish managers and 6,400 of their subordinates (Arvonen, 2002). Leadership behaviour was rated by the subordinates on a six-point scale, ranging from "Do not agree at all" (1) to "Agree fully" (6). A previous confirmatory factor analysis of the same baseline data (Lornudd *et al.*, 2015) showed an acceptable fit of a short version of the scales (13 items), and, thus, the same version was used in the present study. The reliability of the subscales and included items were as follows (Cronbach's α was calculated based on data from Time1): *change* 0.92 ("Shares thoughts and plans about the future", "Consistently pushes for development", "Initiates new projects", "Experiments with new ways of doing things" and "Offers ideas about new and different ways of doing things"); *production* 0.86 ("Makes a point of following rules and principles", "Follows and controls work closely", "Plans carefully", "Gives clear instructions"); and *employee* 0.90 ("Relies on his/her subordinates", "Is considerate", "Is just in treating subordinates" and "Shows respect for other people"). The missing data analysis showed a low level of missing internal data (< 6 per cent) on the CPE items at Time1, Time2 and Time3. Indices for each of the three subscales at each time point were computed based on all cases with responses to at least 75 per cent of the items per subscale.

Demographic variables. To enable comparisons of demographic variables between groups (intervention groups and dropouts), we used demographic items from the short web-based version of QPS Nordic (Webb-QPS), which measures health and psychosocial work environment (Hasson *et al.*, 2008).

Statistical analyses. To examine the effects of the interventions, repeated measures of analysis of variance (ANOVA) and multilevel analysis were conducted using SPSS 21. For managers and superiors, repeated measures ANOVA was performed to investigate potential differences between the two interventions between Time1 and Time2. Because of the nested structure of the ratings from colleagues and subordinates (several raters evaluated the same manager), a multilevel regression analysis was performed. In addition to taking into account the dependency of data ratings related to the same

manager, such assessment offered the advantage of using all available information from each time point because it did not require list wise deletion (Hox, 2002). First, all predictors were grand mean centred (Hofmann and Gavin, 1998, McKee *et al.*, 2011) to create a meaningful and interpretable intercept term. Thereafter, a null model was constructed for each outcome variable, which was used as a baseline for comparison with the contribution of predictors, and then an unconditional model was assessed for each outcome by allowing the intercept term to vary by leader (i.e. the level 2 variable). To detect changes over time, instead of absolute levels at Time2, leadership ratings at Time1 were entered as the first predictor. Finally, the type of intervention was entered.

To investigate the combined effect of the two interventions between Time1 and Time3, repeated measures ANOVA was performed for all rater sources, respectively. For colleagues' and subordinates' ratings, this was also done specifically to investigate overall time effects between Time1 and Time2 (because this was not estimated in the multilevel model). To account for the nested structure of the data for subordinates and colleagues, managerial affiliation was controlled for by insertion as a between-subject variable.

Results

Table II shows correlations, means and standard deviations for the outcome variables. The correlations at each time point (Time1, Time2 or Time3) ranged from 0.55 to 0.67. There were strong correlations of 0.69 to 0.76 between different time points for each leadership behaviour orientation (change, production or employee). The correlation coefficients for outcome variables were on average 0.15-0.25 higher at the same time point compared to the other two time points.

Comparison of Time1 and Time2: type of intervention

Repeated measures ANOVAs (for all raters to examine the effect of time) and multilevel regression analyses (for subordinates and colleagues to examine the effect of intervention type) were performed to address the first research question. Table III presents the results of the repeated measures ANOVAs of changes in managers' and superiors' CPE ratings between Time1 (pre-test) and Time2 when one group of managers had attended backstage groups and the remaining managers had attended theoretical seminars. There was no interaction effect for self-ratings, indicating no mean difference in the three leadership behaviour orientations between the two interventions. However, the main effect for time showed a significant improvement in all behaviour orientations between Time1 and Time2 (Change: $F = 8.298$ [1; 144], $p = 0.005$; Production: $F = 5.327$ [1; 143], $p = 0.022$; Employee: $F = 16.444$ [1; 144], $p = 0.000$). When leadership was rated by the superiors, again there was no interaction effect between time and intervention group. A main effect for time emerged in the employee orientation, although not in the expected direction (the overall mean was lower at Time2 than at Time1; $F = 4.437$ [1; 135], $p = 0.037$). No time effect was found regarding the superiors' ratings of the other two behaviour orientations. The results of the repeated measures ANOVAs of the subordinates' and the colleagues' ratings are shown in Table IV. Similar to the superiors, the subordinates rated a lower level of employee-oriented behaviour at Time2 compared to Time1 ($F = 12.294$ [1; 541], $p = 0.000$). No other significant time effect emerged.

Table II.
Correlations and mean values of *change, production, employee (CPE)* leadership behaviour ratings from all rater sources

Variable	1	2	3	4	5	6	7	8	9
1. Change Time1									
2. Change Time2	0.737***								
3. Change Time3	0.691***	0.776***							
4. Production Time1	0.616***	0.466***	0.425***						
5. Production Time2	0.470***	0.667***	0.524***	0.704***					
6. Production Time3	0.400***	0.470***	0.664***	0.694***	0.760***				
7. Employee Time1	0.583***	0.479***	0.370***	0.553***	0.449***	0.294***			
8. Employee Time2	0.475***	0.613***	0.443***	0.419***	0.607***	0.384***	0.771***		
9. Employee Time3	0.463***	0.481***	0.584***	0.338***	0.463***	0.549***	0.724***	0.767***	
Mean	4.35	4.37	4.43	4.47	4.50	4.53	4.74	4.73	4.78
SD	0.96	1.01	0.97	0.88	0.92	0.88	1.01	1.05	0.95

Note: ***$p < 0.001$ (two-tailed)

Table III.
Changes in ratings of leadership behaviour between Time1 and Time2, according to rater source

Outcome variable	Backstage groups Time1 Mean (SD)	Backstage groups Time2 Mean (SD)	Theoretical seminars Time1 Mean (SD)	Theoretical seminars Time2 Mean (SD)	Repeated measures ANOVA Time effect (df1; df2[a])	Repeated measures ANOVA Group × time (df1; df2[a])
Change						
Managers	4.56 (0.85)	4.72 (0.84)	4.67 (1.01)	4.76 (1.00)	$F = 8.298$ (1;144) $p = 0.005$	$F = 0.678$ (1;144) $p = 0.412$
Superior	4.60 (1.11)	4.55 (1.23)	4.45 (1.32)	4.49 (1.46)	$F = 0.002$ (1;136) $p = 0.968$	$F = 0.655$ (1;136) $p = 0.420$
Production						
Managers	4.41 (0.88)	4.55 (0.83)	4.39 (1.03)	4.47 (0.97)	$F = 5.327$ (1;143) $p = 0.022$	$F = 0.379$ (1;143) $p = 0.539$
Superiors	4.84 (0.86)	4.82 (0.92)	4.80 (1.02)	4.72 (1.10)	$F = 1.397$ (1;135) $p = 0.239$	$F = 0.390$ (1;135) $p = 0.533$
Employee						
Managers	5.08 (0.64)	5.22 (0.64)	5.09 (0.76)	5.23 (0.76)	$F = 16.444$ (1;144) $p = 0.000$	$F = 0.131$ (1;144) $p = 0.718$
Superiors	5.18 (0.80)	5.14 (0.82)	5.23 (0.91)	5.09 (0.97)	$F = 4.437$ (1;135) $p = 0.037$	$F = 1.393$ (1;135) $p = 0.240$

Notes: [a]The number of degrees of freedom vary within rater sources between leadership orientations because of occasional missing values; SD = standard deviation

Multilevel regression analysis was performed separately on the subordinates' and colleagues' ratings to discern any difference between the interventions. First, a null model was calculated for each outcome variable (CPE orientation at Time2). Intra-class correlation, ICC (1), was computed based on the unconditional model and showed that individuals were indeed nested within managers: the manager (second) level of analysis accounted for a substantial part of the variance, with ICC (1) at Time2 explaining between 8-28 per cent of the variance for the colleagues and 24-27 per cent for the subordinates. Multilevel analysis was therefore considered appropriate. Tables V and VI show the results of multilevel analysis for colleagues and subordinates separately. In Step 1, leadership ratings at Time1 were added as a predictor to control for initial ratings at Time1. This substantially improved the model fit -2 log likelihood (-2LL) for both subordinates and colleagues, indicating that leadership ratings at Time1 explained the majority of the variance of the ratings at Time2 for both rater sources. In Step 2, intervention type was added as a level-2 predictor, which yielded low β coefficients and improved the model fit in only one case: at Time2, the colleagues rated a larger increase in production orientation for the managers in the theoretical intervention.

Comparison of Time1 and Time3: combination of the two interventions

The combined effect of backstage groups and theoretical seminars was investigated by comparing the leadership ratings from Time1 (pretest) with ratings from Time3 (when all the managers had participated in both interventions). The results of the repeated measures ANOVA of CPE data from all the rater sources at Time1 and Time3 are presented in Table VII. Self-ratings showed significant time effects for all leadership behaviour orientations between Time1 and Time3 (Change: $F = 4.768$ [1; 114], $p = 0.031$; Production: $F = 7.543$ [1; 114], $p = 0.007$; Employee: $F = 9.728$ [1; 115], $p = 0.002$.). For the leadership ratings by the superiors and the colleagues, there were no significant differences between Time1 and Time3. However, the subordinates rated a significantly lower employee orientation at Time3 compared to Time1 ($F = 8.454$ [1; 294], $p = 0.004$).

Table IV.
The effect of both interventions on ratings of leadership behaviour between Time1 and Time2, according to rater source

Outcome variables	Time1 Mean (SD)	Time2 Mean (SD)	ANOVA repeated measures Time effect (df1; df2[a])
Change			
Colleagues	4.43 (1.04)	4.38 (0.89)	$F = 1.098$ (1;79) $p = 0.298$
Subordinates	4.29 (1.04)	4.27 (1.07)	$F = 0.303$ (1;531) $p = 0.582$
Production			
Colleagues	4.75 (0.71)	4.65 (0.78)	$F = 3.787$ (1;79) $p = 0.055$
Subordinates	4.44 (0.96)	4.42 (1.01)	$F = 0.369$ (1;525) $p = 0.544$
Employee			
Colleagues	4.90 (0.84)	4.83 (0.90)	$F = 1.924$ (1;85) $p = 0.169$
Subordinates	4.66 (1.10)	4.55 (1.20)	$F = 12.294$ (1;541) $p = 0.000$

Model	Change	Production	Employee
Null (−2LL)	3,027.77	2,907.51	3,249.03
Unconditional (−2LL)	2,887.96	2,785.78	3,133.73
Step 1 (−2LL)	1,471.72	1,373.82	1,486.21
Step 2 (−2LL)	1,473.55	1,377.08	1,488.52
Model coefficients step 1 (Unstandardised β)			
Leadership Time1	0.739***	0.735***	0.826***
Model coefficients step 2 (Unstandardised β)			
Leadership Time1	0.740***	0.735***	0.827***
Type of intervention	−0.089***	−0.042***	−0.075***

Notes: Number of groups: 166. −2LL = −2 restricted log likelihood; ***$p < 0.001$ (two-tailed)

Table V. Predictors of change in leadership at Time2, subordinates' ratings

Model	Change	Production	Employee
Null (−2LL)	704.26	623.38	720.30
Unconditional (−2LL)	683.99	622.33	716.50
Step 1 (−2LL)	324.71	338.10	358.59
Step 2 (−2LL)	326.30	332.84	361.11
Model step 1 (Unstandardised β)			
Leadership Time1	0.739***	0.739***	0.812***
Model step 2 (Unstandardised β)			
Leadership Time1	0.741***	0.751***	0.812***
Type of intervention	0.108	0.271**	–0.047

Notes: Number of groups: 145. −2LL = −2 restricted log likelihood; ***$p < 0.001$ (two-tailed)

Table VI. Predictors of change in leadership at Time2, colleagues' ratings

Discussion

The object of this longitudinal randomised controlled trial was to compare the effects of two different leader development interventions and to evaluate the combined effect of these interventions after two years. Overall, few differences in improvements in leadership behaviour orientation were found between the interventions. The reported effects over time differed between rater sources, with a similar pattern after one year (regardless of which intervention) and after two years (a combination of both interventions). To summarise, at both post-tests, the managers' self-ratings indicated improved change-, production- and employee-oriented leadership behaviour, whereas the superiors, colleagues and subordinates did not perceive any difference in production- or change-oriented leadership behaviour. However, employee-oriented leadership behaviour was rated lower by superiors (at Time2) and subordinates (at both post-tests).

Below, we suggest some explanations for the mixed results. First, we address the finding that the CPE scale indicated no substantial difference between the backstage groups and the theoretical seminars regarding improvements in leadership behaviour orientations. Thereafter, we consider the results in terms of differences between pre-test

Outcome variables	Time1 Mean (SD)	Time3 Mean (SD)	ANOVA repeated measures Time effect (df1; df2[a])
Change			
Managers	4.63 (0.65)	4.74 (0.65)	$F = 4.768$ (1; 114) $p = 0.031$
Superiors	4.58 (0.84)	4.55 (0.94)	$F = 0.162$ (1; 92) $p = 0.688$
Colleagues	4.42 (1.03)	4.49 (0.97)	$F = 1.007$ (1;50) $p = 0.320$
Subordinates	4.35 (1.06)	4.34 (1.06)	$F = 0.017$ (1;289) $p = 0.898$
Production			
Managers	4.44 (0.63)	4.57 (0.61)	$F = 7.543$ (1; 114) $p = 0.007$
Superiors	4.83 (0.66)	4.86 (0.73)	$F = 0.269$ (1; 92) $p = 0.605$
Colleagues	4.77 (0.61)	4.75 (0.67)	$F = 0.111$ (1;50) $p = 0.740$
Subordinates	4.51 (0.92)	4.46 (0.92)	$F = 1.336$ (1;291) $p = 0.249$
Employee			
Managers	5.12 (0.50)	5.25 (0.48)	$F = 9.728$ (1; 115) $p = 0.002$
Superiors	5.19 (0.61)	5.14 (0.73)	$F = 0.896$ (1;92) $p = 0.346$
Colleagues	4.82 (0.83)	4.85 (0.71)	$F = 0.434$ (1;53) $p = 0.513$
Subordinates	4.70 (1.00)	4.58 (1.06)	$F = 8.454$ (1;294) $p = 0.004$

Note: [a] The number of degrees of freedom within rater sources varies between leadership orientations because of occasional missing values

Table VII.
Changes in ratings of leadership behaviour between Time1 and Time3, according to rater source

and post-tests (i.e. at 0, 12 and 24 months) without making any distinction between the two interventions.

Comparison of the backstage groups and the theoretical seminars showed no substantial difference in improvements of the three leadership behaviour orientations. Although participants in both interventions were guided through all phases of the learning cycle (Kolb, 1984), there was a clear emphasis on practical dilemmas and reflection in the backstage groups, whereas theory was emphasised in the theoretical seminars. However, it is possible that the contents of the interventions were fairly similar and therefore influenced the participants in essentially the same way. Another potential explanation is that both interventions contained other elements that stimulated change: the participants in both interventions may have felt empowered because they experienced valuable support by meeting colleagues and discussing the conditions for performing leadership in health-care services. It is also possible that changes did occur in three indices of behavioural orientation but only in certain sub-groups. Some researchers argue that individual characteristics of the participants account for the main part of the variability in training outcomes (van der Klink *et al.*, 2001). In this interpretation, a potential explanation for the non-significant differences in change between the two interventions is that individual disparities between managers could have had a greater impact than the dissimilarity in programme content.

Did changes occur but not in the behavioural orientations?

According to the evaluation of the two interventions combined, the external raters felt that there was no improvement (or in some cases even deterioration). However, it is possible that the managers did progress but not in terms of the three behavioural orientations. For example, previous studies of interventions based on reflection have

indicated a self-rated development towards a more elaborate way of thinking about the leadership role (Bergman *et al.*, 2009b, Sandahl *et al.*, 2007). Moreover, in another investigation in the present project, selected subordinates of managers that had participated in the full two-year development programme were individually interviewed about their perceptions of changes in their manager's leadership over time (Palm *et al.*, 2015), and a majority of those interviewed reported improved leadership. This finding supports the notion that the full programme actually contributed to development that was also noticed by subordinates. Such development was almost consistently described as the managers demonstrating improved self-confidence in the leadership role, reflected by altered behaviours in relation to decision-making, handling of conflicts and supervising work meetings.

Did changes in behavioural orientation occur but were not captured by the measurement criteria?

In our study, a 360-degree feedback instrument was used for evaluation at 0, 12, and 24 months without the participants receiving feedback on their ratings. The use of a 360-degree instrument for assessment of interventions has previously been questioned by Jellema *et al.* (2006), who concluded that in most cases external raters do not perceive positive behavioural changes in managers after they have undergone training. How can this be explained? For instance, during an intervention, a participating manager might discover a new way to deal with a certain situation and subsequently be able to perceive his/her role as a leader differently (e.g. with feelings of increased coping capacity), with an *intention to behave* when such a situation arises. However, situations that require specific and observable leadership behaviour seldom occur in everyday practice (Larsson and Lundholm, 2010) and, naturally, it would be impossible for external raters to detect an intention to behave before the behaviour is actually displayed. This could explain why the present managers perceived a behavioural change in production- and change-oriented behaviour, whereas the external raters did not. Consequently, external raters need to be more extensively exposed to a manager to register changes, as indicated by research showing that honest feedback ratings are predicted by opportunity to observe (Smith and Fortunato, 2008). In as much as leadership concerns relationships and interaction (Yukl, 2012), it might be easier to observe the presence of employee-oriented behaviour. In line with that, perhaps the present superiors and subordinates had ample opportunity to observe employee-oriented behaviour (which decreased significantly from Time1 to Time2 and Time3) but not change- and production-oriented behaviour. Generally speaking, a decrease in employee-oriented leadership might not necessarily be a negative outcome. A person with a tendency to be highly focused on relations might not be perceived as the most "leader-like" person. The link between agreeableness and employee-oriented leadership behaviour was recently supported by a study showing that managers' self-rated agreeableness significantly predicted ratings in employee-oriented leadership made by external raters (Bergman *et al.*, 2014). Although the personality trait "agreeableness" has been demonstrated to be weakly correlated with leadership (Judge *et al.*, 2002), assuming a curvilinear relationship (leadership might be associated with moderate levels of agreeableness but not with low or high levels) instead of a linear would make sense of our finding. Thus, paradoxically, it is possible that lower ratings of employee orientation may mirror a display of greater confidence and more leader-like behaviour in managers, which would

be in line with previous research indicating that participants in backstage groups learned to more clearly define their managerial role (Bergman *et al.*, 2009a).

Were there no changes at all?

A third interpretation of our inconsistent results is that the relationships detected were spurious and that some factor other than the interventions produced the significant differences between the managers', superiors' and subordinates' ratings. The participating managers gave themselves higher ratings after a development initiative compared to pre-test, and this might be explained by the theory of cognitive dissonance (Festinger, 1957), which indicates that humans strive for an alignment between attitudes, values and actual behaviour. If a busy manager attends an extensive leader development programme, she/he would probably describe it as a valuable experience to justify the time spent. It is also possible that the observed behavioural changes were either β changes (subjective alteration of the measurement scale) or γ changes (altered perception of the construct being measured) (Golembiewski *et al.*, 1976). However, the possibility that the managers did not improve after attending a two-year leader development programme is contradicted by the qualitative study by (Palm *et al.*, 2015), based on interviews with a sample of subordinates.

Limitations

In general, the evaluation criteria used should conform to the content of a development programme. In the present study, learning outcomes based on the results of previous explorative studies of the effects of backstage groups (Bergman *et al.*, 2009b) would have been more closely correlated with the content of backstage groups. For the theoretical seminars, learning criteria for leadership theory would have better corresponded to the content. Nevertheless, to go a step further and investigate the effectiveness of the interventions, we decided to analyse the potential impact on leadership behaviour perceived by external raters. It is however possible that external ratings of leadership behaviour were too distal and too non-specific an outcome measure in relation to the content of the programme. A more feasible approach might have been to predict learning outcomes (such as self-confidence in the leadership role) and to measure more distal, yet specific, leadership behaviours (e.g. decision-making behaviour) that were theoretically linked to these learning outcomes.

No data on the content or process of the development programme were included in the study, which limits comprehension of the results. However, the focus was not on exploring the effects of specific mechanisms of the backstage groups or theoretical seminars but rather to investigate whether the interventions had any impact at all on the three leadership behaviour orientations. Notwithstanding, our study could not show whether the programme had a buffering effect in the sense that the managers who participated received valuable support and, hence, in contrast to the non-participating managers, maintained a high level of leadership behaviour. Thus, future research should aim at identifying what aspects of backstage groups as an intervention that are appropriate for whom and when.

Implications for research and practice

The need for longitudinal studies of leader development has been stressed (Riggio and Mumford, 2011). The present investigation had a longitudinal design, and it also included a large sample comprising both managers and external raters which enhanced

the reliability. Despite the potential limitations of our study, we believe that the improved self-ratings – which did not reveal any significant effect of participating in the backstage groups or the theoretical seminars – raise questions about the mechanisms that drive development. Two development interventions, with fairly different pedagogical approaches, led to the same outcome regarding leadership behaviour orientation. Perhaps characteristics of the participants accounted for the main variance in training outcomes, as has previously been suggested (van der Klink *et al.*, 2001). Or there may be other factors that influence development in programmes of this length (two years), such as repeated monthly social support from colleagues and a positive feeling that the organisation is investing in its workforce.

Conclusions

Our results suggest that improvement in self-rated leadership behaviour orientations can be achieved to the same extent by holding regular sessions with manager colleagues to discuss leadership-related themes as by conducting sessions in smaller groups aimed at finding practical solutions to difficulties in the role of manager. The inconsistent leadership ratings from different rater sources raise questions about how changes in leadership in an organisation can be noticeable. This study also highlights the interpretive challenges connected with using a 360-degree instrument to evaluate leader development.

References

Arvonen, J. (2002), "Change, production and employees – an integrated model of leadership", Doctoral dissertation, Stockholms Universitet, Stockholm.

Avolio, B.J., Reichard, R.J., Hannah, S.T., Walumbwa, F.O. and Chan, A. (2009), "A meta-analytic review of leadership impact research: experimental and quasi-experimental studies", *The Leadership Quarterly*, Vol. 20 No. 5, pp. 764-784.

Bergman, D. (2009), "Leadership development: a comparative evaluation of short-term and long-term programmes in Swedish health care", Doctoral dissertation, Karolinska Institutet, Stockholm.

Bergman, D., Arnetz, B., Wahlström, R. and Sandahl, C. (2007), "Effects of dialogue groups on physicians' work environment", *Journal of Health Organization and Management*, Vol. 21 No. 1, pp. 27-38.

Bergman, D., Fransson-Sellgren, S., Wahlström, R. and Sandahl, C. (2009a), "Healthcare leadership: impact of short-term intensive and long-term less intensive training programmes", *Leadership in Health Services*, Vol. 22 No. 2, pp. 161-175.

Bergman, D., Lornudd, C., Sjöberg, L. and Von Thiele Schwarz, U. (2014), "Leader personality and 360-degree assessments of leader behavior", *Scandinavian Journal of Psychology*, Vol. 55 No. 4, pp. 389-397.

Bergman, D., Savage, C., Wahlstrom, R. and Sandahl, C. (2008), "Teaching group dynamics– do we know what we are doing? An approach to evaluation", *Medical Teacher*, Vol. 30 No. 1, pp. 55-61.

Bergman, D., Stotzer, E., Wahlström, R. and Sandahl, C. (2009b), "Learning from dialogue groups-physicians' perceptions of role", *Journal of Health Organization and Management*, Vol. 23 No. 2, pp. 225-239.

Berntson, E., Wallin, L. and Härenstam, A. (2012), "Typical situations for managers in the Swedish public sector: cluster analysis of working conditions using the job demands-resources model", *International Public Management Journal*, Vol. 15 No. 1, pp. 100-130.

Björklund, M. (2009), "Backstagegrupper för chefer – ledarutveckling genom erfarenhetsutbyte och reflektion [Backstage groups for managers - leader development by exchange of experience and reflection]", Master thesis, Stockholm University, Stockholm, available at: www.diva-portal.org/smash/record.jsf?pid=diva2%3A200362&dswid=8948 (accessed 10 June 2015).

Collins, D.B. and Holton, E.F. (2004), "The effectiveness of managerial leadership development programs: a meta-analysis of studies from 1982 to 2001", *Human Resource Development Quarterly*, Vol. 15 No. 2, pp. 217-248.

Day, D.V. (2001), "Leadership development: a review in context", *The Leadership Quarterly*, Vol. 11 No. 4, pp. 581-613.

Day, D.V., Fleenor, J.W., Atwater, L.E., Sturm, R.E. and McKee, R.A. (2014), "Advances in leader and leadership development: a review of 25 years of research and theory", *The Leadership Quarterly*, Vol. 25 No. 1, pp. 63-82.

Ekvall, G. and Arvonen, J. (1991), "Change-centered leadership: an extension of the two-dimensional model", *Scandinavian Journal of Management*, Vol. 7 No. 1, pp. 17-26.

Festinger, L. (1957), *A theory of Cognitive Dissonance*, Row, Peterson, Evanstone, IL.

Goffman, E. (1959), *The Presentation of Self in Everyday Life*, Doubleday, Garden City, NY.

Golembiewski, R.T., Billingsley, K. and Yeager, S. (1976), "Measuring change and persistence in human affairs: types of change generated by OD designs", *The Journal of Applied Behavioral Science*, Vol. 12 No. 2, pp. 133-157.

Hasson, D., Gustavsson, P. and Sandahl, C. (2008), "Webb-QPS – Utveckling och prövning av en webbaserad metod för mätning av psykosocial arbetsmiljö och hälsa (Webb-QPS – Development and testing of a web-based method for assessing psychosocial work environment and health)", *Report C 2008:1*, Karolinska Institutet, Stockholm.

Hensel, R., Meijers, F., van der Leeden, R. and Kessels, J. (2010), "360 degree feedback: how many raters are needed for reliable ratings on the capacity to develop competences, with personal qualities as developmental goals?", *The International Journal of Human Resource Management*, Vol. 21 No. 15, pp. 2813-2830.

Hofmann, D.A. and Gavin, M.B. (1998), "Centering decisions in hierarchical linear models: implications for research in organizations", *Journal of Management*, Vol. 24 No. 5, pp. 623-641.

Hox, J. (2002), *Multilevel Analysis. Techniques and Applications*, Lawrence Erlbaum, Mahwah, NJ.

Jackson, B. and Parry, K. (2008), *A Very Short, Fairly Interesting and Reasonably Cheap Book About Studying Leadership*, SAGE Publications, London.

Jellema, F., Visscher, A. and Scheerens, J. (2006), "Measuring change in work behavior by means of multisource feedback", *International Journal of Training and Development*, Vol. 10 No. 2, pp. 121-139.

Judge, T.A., Bono, J.E., Ilies, R. and Gerhardt, M.W. (2002), "Personality and leadership: a qualitative and quantitative review", *Journal of Applied Psychology*, Vol. 87 No. 4, pp. 765-780.

Kelloway, E.K. and Barling, J. (2010), "Leadership development as an intervention in occupational health psychology", *Work & Stress*, Vol. 24 No. 3, pp. 260-279.

Kirkpatrick, D. (1975), *Evaluating Training Programs*, Tata McGraw-Hill Education.

Kolb, D.A. (1984), *Experiential Learning: Experiences as the Source of Learning and Development*, Prentice Hall, Englewood Cliffs, NJ.

Larsson, J. and Vinberg, S. (2010), "Leadership behaviour in successful organisations: universal or situation-dependent?", *Total Quality Management*, Vol. 21 No. 3, pp. 317-334.

Larsson, M. and Lundholm, S.E. (2010), "Leadership as work-embedded influence: a micro-discursive analysis of an everyday interaction in a bank", *Leadership*, Vol. 6 No. 2, pp. 159-184.

Lewin, K. (1951), "Field theory and learning", in Cartwright, D. (Ed.), *Field Theory in Social Science: Selected Theoretical Papers*, Social Science Paperbacks, London.

Lornudd, C., Tafvelin, S., von Thiele Schwarz, U. and Bergman, D. (2015), "The mediating role of demand and control in the relationship between leadership behaviour and employee distress: a cross-sectional study", *International Journal of Nursing Studies*, Vol. 52 No. 2, pp. 543-554.

McKee, M.C., Driscoll, C., Kelloway, E.K. and Kelley, E. (2011), "Exploring linkages among transformational leadership, workplace spirituality and well-being in health care workers", *Journal of Management, Spirituality & Religion*, Vol. 8 No. 3, pp. 233-255.

Nielsen, K., Taris, T.W. and Cox, T. (2010), "The future of organizational interventions: addressing the challenges of today's organizations", *Work & Stress*, Vol. 24 No. 3, pp. 219-233.

Palm, K., Ullström, S., Sandahl, C. and Bergman, D. (2015), "Employee perceptions of changes in managers' leadership over time", *Leadership in Health Services*.

Plsek, P.E. and Greenhalgh, T. (2001), "Complexity science: the challenge of complexity in health care", *BMJ: British Medical Journal*, Vol. 323 No. 7313, p. 625.

Rainey, H.G. and Chun, Y.H. (2005), "Public and private management compared", in Ferlie, E.L.L. and Pollitt, C.H. (Eds), *The Oxford Handbook of Public Management*, Oxford University Press, Oxford.

Riggio, R.E. and Mumford, M.D. (2011), "Introduction to the special issue: longitudinal studies of leadership development", *The Leadership Quarterly*, Vol. 22 No. 3, pp. 453-456.

Romanowska, J., Larsson, G., Eriksson, M., Wikström, B.M., Westerlund, H. and Theorell, T. (2011), "Health effects on leaders and co-workers of an art-based leadership development program", *Psychotherapy and Psychosomatics*, Vol. 80 No. 2, pp. 78-87.

Salas, E., Tannenbaum, S.I., Kraiger, K. and Smith-Jentsch, K.A. (2012), "The science of training and development in organizations: what matters in practice", *Psychological Science in the Public Interest*, Vol. 13 No. 2, pp. 74-101.

Sandahl, C., Bäckstrand, M. and Edenius, J. (2012), *Backstage Groups for Managers. Guidance for Group Leaders*, Karolinska Institutet, Stockholm.

Sandahl, C., Edenius, J., Gustafsson, H. and Wahlström, T. (2007), *Självförtroende och Handlingskraft genom Stockholmsmodellen för Ledarutveckling (Self-Confidence and Decisiveness Through the Stockholm Model for Leader Development)*, Karolinska Institutet, Stockholm.

Sellgren, S.F., Ekvall, G. and Tomson, G. (2008), "Leadership behaviour of nurse managers in relation to job satisfaction and work climate", *Journal of Nursing Management*, Vol. 16 No. 5, pp. 578-587.

Sandahl, C., Falkenström, E. and von Knorring, M. (2010), *Chef med Känsla och Förnuft (Manager with Sense and Sensibility)*, Natur och Kultur, Stockholm.

Smith, A.F. and Fortunato, V.J. (2008), " Factors influencing employee intentions to provide honest upward feedback ratings", *Journal of Business and Psychology*, Vol. 22 No. 3, pp. 191-207.

van der Klink, M., Gielen, E. and Nauta, C. (2001), "Supervisory support as a major condition to enhance transfer", *International Journal of Training and Development*, Vol. 5 No. 1, pp. 52-63.

Wikström, E. and Dellve, L. (2009), "Contemporary leadership in healthcare organizations: fragmented or concurrent leadership", *Journal of Health Organization and Management*, Vol. 23 No. 4, pp. 411-428.

Yukl, G. (2012), *Leadership in Organizations*, Pearson, Boston, MA.

Yukl, G., Gordon, A. and Taber, T. (2002), "A hierarchical taxonomy of leadership behavior: integrating a half century of behavior research", *Journal of Leadership & Organizational Studies*, Vol. 9 No. 1, pp. 15-32.

Corresponding author
Caroline Lornudd can be contacted at: caroline.lornudd@ki.se