Personality and Social Psychology

Leadership behavior changes following a theory-based leadership development intervention: A longitudinal study of subordinates' and leaders' evaluations

GERRY LARSSON,[1,2] CHRISTER SANDAHL,[3] TERESA SÖDERHJELM,[3] ENDRE SJÖVOLD[4] and ANN ZANDER[1]

[1]*Swedish Defence University, Karlstad, Sweden*
[2]*Hedmark University College, Elverum, Norway*
[3]*Karolinska Institutet, Stockholn, Sweden*
[4]*Norwegian University of Technology and Science, Trondheim, Norway*

Larsson, G., Sandahl, C., Söderhjelm, T., Sjövold, E. & Zander, A. (2016). Leadership behavior changes following a theory-based leadership development intervention: A longitudinal study of subordinates' and leaders' evaluations. *Scandinavian Journal of Psychology*.

The aim was to evaluate effects of leadership courses based on the developmental leadership model at the leadership behavioral level. A longitudinal design was employed with assessments before, one and six months after the leadership courses. The sample consisted of 59 leaders who made self-ratings and were rated by at least three subordinates on each occasion. Leadership behaviors were measured with the Developmental Leadership Questionnaire (DLQ). A limited increase of favorable leadership behaviors and a significant reduction of unfavorable leadership behaviors were found, particularly according to the subordinates' ratings. A cluster analysis yielded three meaningful leader profiles and showed that this pattern was found in all three profiles, irrespective of how favorably they were rated before the onset of the intervention.

Key words: Leadership development, theory-based intervention, longitudinal, Development Leadership Questionnaire (DLQ).

Gerry Larsson, *Department of Security*, *Strategy and leadership*, *Swedish Defence University*, *651 80 Karlstad*, *Sweden*. E-mail: gerry.larsson@fhs.se

INTRODUCTION

Research on leadership is extensive yet studies on the effectiveness of leadership development interventions is sparse and inconclusive (Day, 2000; Dvir, Eden, Avolio & Shamir, 2002). This may be seen as surprising given the amount of time and money spent on such interventions. To illustrate, Grint (2007) estimated that the yearly sum amounted to between $15 and $50 billion world-wide.

One reason for the lack of research may be the complexity of leadership development in working life. Time limits for interventions are often tight, the leaders and their co-workers have numerous job tasks and responsibilities at the same time, antecedent conditions and outcomes at various levels may be difficult to assess, etc. All combined, this means that ideal research designs are often difficult to implement. Such designs are suggested to be theory-based, longitudinal, have random assignment of participants to intervention and control groups, use valid and reliable multilevel measurements of antecedents and outcomes in a variety of organizational contexts, etc. (Day, Fleenor, Atwater, Sturm & McKee, 2014; Reichard & Avolio, 2005).

A relevant question raised by Day *et al*. (2014) in their extensive review of research on leader and leadership development interventions is *what* develops as a function of these efforts? Are some intrapsychic and behavioral aspects easier to affect than others and under which contextual conditions? In a time-limited program with adults, hands-on skills, coping strategies and leadership behaviors tend to be easier to affect, than deeper psychological aspects such as authority and self-confidence in the role of leader (Day *et al*., 2014).

Additionally, are there individual differences regarding developmental trajectories across time? This question indicates that the standard variable approach to research needs to be supplemented by a person approach, where one studies profiles or patterns of, for instance, various leadership behaviors and their antecedents and outcomes across time (Bass, 1998; Nystedt, 1997). Leaders start in different places in their developmental journeys and develop different aspects at different rates. There is no clear evidence regarding who benefits more or less from leadership programs, those who are good leaders already at the onset or those who have much room for improvement (Day *et al*., 2004).

A further basic question concerns differences and similarities between leader development and leadership development. The former focuses on the leader as a person while the latter deals with interactions between several individuals (Day *et al*., 2014). In our experience, the personal development is focused when leaders are brought out of their ordinary work environment to attend a program, although interpersonal skills often also receive attention.

Summing up, the challenges for researchers in the leader and/or leadership development area are considerable and we venture to guess that few, if any, studies fulfill all the aforementioned ideal attributes. In this study, we present a case which is theory-based, longitudinal, uses a well-documented measurement instrument at two levels – the leaders themselves and their subordinates. However, it is a comparatively small sample with no random assignment of leaders to various intervention conditions. Antecedent and outcome measures are also lacking. Given these strengths and weaknesses, we will now summarize the theoretical model on which the study is based, as well as the format and

content of the leader development intervention, and conclude the introduction by presenting three hypotheses.

THE DEVELOPMENTAL LEADERSHIP MODEL AND COURSES BASED ON IT

The present study uses the developmental leadership model (Larsson, Carlstedt, Andersson *et al*., 2003) as its point of departure and the model is presented in Fig. 1.

According to the model (Fig. 1), leadership can be understood against the background of a number of interacting factors. The interplay between leader and contextual characteristics shapes leadership behaviors. This implies that the model rests on an interactional person-by-situation paradigm (Endler & Magnusson, 1976).

Two main classes of leader qualities are identified: basic prerequisites and desirable competencies. The more favorable basic prerequisites a leader has, the greater the potential to develop the desirable competencies and vice versa. The model also implies that a favorable combination of these two characteristics is a necessary condition for successful leadership. However, neither of them is sufficient in itself. They do not constitute a guarantee for successful leadership, because this is also affected by environmental conditions.

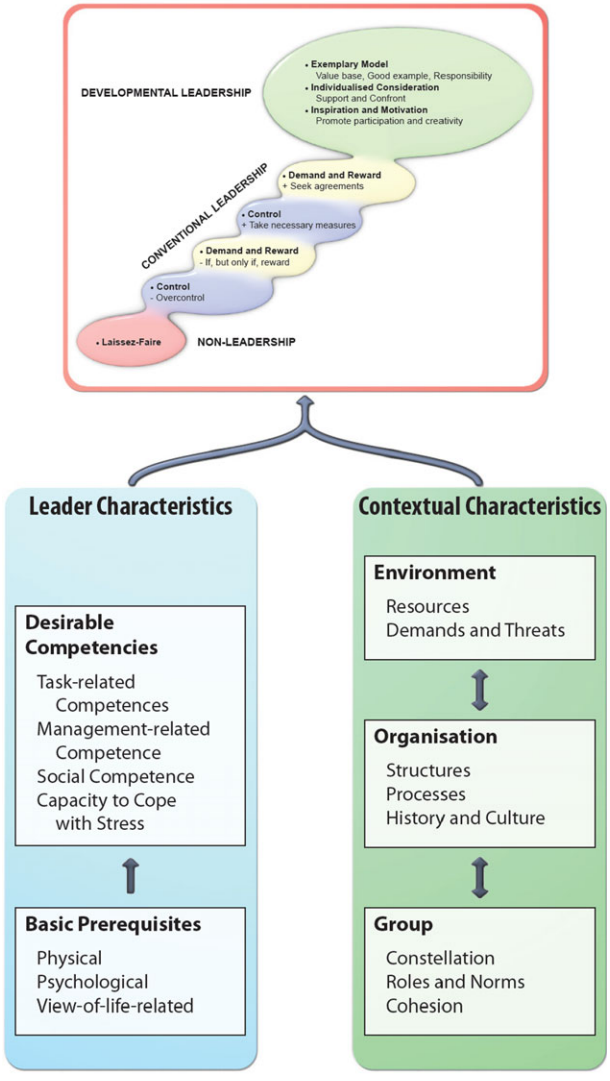

Fig. 1. The Developmental leadership model (adapted from Larsson *et al*., 2003).

The environmental characteristic shown in Fig. 1 should be regarded as examples of these kinds of conditions. Typical illustrations include: (1) environment – degree of predictability of environmental resources and demands; (2) organization – number of hierarchical levels and degree of power centralization; and (3) group – degree of role and norm clarity. The figure shows that groups and organizations mutually influence each other. The same holds true for organizations and the external world. There is a great deal of literature on these aspects and they will not be elaborated on here.

Also, the leader qualities labeled "basic prerequisites" will only be mentioned briefly here, for the same reason. Somewhat simplified, these aspects include individual characteristics such as physical fitness, intelligence, creativity, personality and view-of-life (see for example Yukl, 2005).

The model includes three basic types of leadership styles: developmental leadership, conventional leadership and non-leadership (laissez-faire). This part is heavily influenced by the writing on transformational leadership and the full range of leadership model (Bass, 1998, 1999). However, some alterations have been made to the original American model. These include a reduction of the number of factors in the transformational (or developmental) domain and an elaboration of what Bass (1997, 1998) labels "transactional leadership." Within each of the two factors included in the transactional (conventional) part of the model, both positively toned facets (demand and reward – seek agreements and control – take necessary measures) and negatively toned facets (demand and reward – if, but only if, reward and control – overcontrol) are identified (Larsson *et al*., 2003).

Leadership courses based on the developmental leadership model consist of four steps: (1) a pre-course 360-degree assessment of the leadership behavior and desirable competences domains of the model (see Fig. 1) using the Developmental Leadership Questionnaire (DLQ; Larsson 2006); (2) a 3-day course with 10–15 participants, who normally do not work together, with intensive work on the 360-feedback, detailed analysis of strengths and weaknesses in supervised 3-person groups and the production of a personal development plan; (3) homework assignments including planned meetings with two other course participants targeting experiences of attempted behavior changes including reflections on supportive forces and obstacles; and (4) a 2-day follow-up session about three months later with a focus on personal leadership development experiences, revision of the personal plan and reflections on future leadership development.

AIM

The aim was to evaluate effects of leadership courses based on the Developmental leadership model at the leadership behavior level. The three hypotheses described below were explored.

Hypothesis 1. Course participants will exhibit an increase of leadership behaviors reflecting the leadership styles developmental leadership and conventional-positive leadership.

Hypothesis 2. Course participants will exhibit a reduction of leadership behaviors reflecting the leadership styles conventional-negative and laissez-faire leadership.

Hypothesis 3. All course participants, irrespective of their leadership behaviors at the onset of the course, will exhibit the favorable behavioral changes presented in Hypotheses 1 and 2.

METHOD

Participants

The selection of informants was done as follows. First, leadership instructors, who were authorized by the Swedish Defence University as course instructors (facilitators) of the leadership course developmental leadership, and employed by one of five Swedish leadership consultant companies, were informed about the study and asked if they were willing to contribute. Their role would be to inform their coming course participants about the longitudinal study.

In the second step, the course participants were informed about the study using written information provided by the Swedish Defence University. This information was obtained about one month before the first course meeting alongside the standard 360-evaluation which precedes these courses. At that time, the course participant was asked to select people who could rate him/her. They were advised to choose their immediate manager and 6–10 subordinates. When the raters received information about the course their target person was about to attend, and the request (anonymous and voluntary) to make the rating, they also received information about the longitudinal research project. These pre-course ratings constitute the first assessment occasion.

The leaders (course participants), as well as their raters, could choose to agree to take part or to say no (or simply not respond to the attached web link). All leaders and raters were also informed that study participation was voluntary and anonymous, and that they could leave the study at any time.

The second measurement took place about one month after the end of the second and final course meeting (see course description above). The third and last assessment took place about six months after the course, which is about ten months after the first measurement.

In total, 102 leaders were asked to participate. Among those who said yes, 97 leaders and 777 raters actually responded to the first assessment (before the leadership course). In the present study we chose to focus on the leaders and their subordinate raters and to only accept leaders who had at least three responding subordinates on each measurement occasion. This yielded a final study sample consisting of 59 leaders and 361 raters (see Table 1 for more details. Table 1 shows that most leaders and subordinate raters were women, 30 years old or more and had a university education. The majority of the leaders had middle manager positions.

The leadership consultant companies involved in the study give courses to a broad array of leaders on the labor market. Following from this, the leaders and raters in the final sample represent a mixture of people employed in administration, service professions, industrial production, schools and health care.

Table 1. *Description of leaders and raters*

	Leaders (n = 59)		Subordinate raters (n = 361)	
Variable	n	%	n	%
Sex				
Female	33	56	234	65
Male	26	44	127	35
Age				
<30	0	0	35	10
30–50	35	59	203	57
>50	24	41	117	33
Education				
Basic school	1	2	15	4
High School	58	9	97	27
University	53	89	247	69
Position				
Line manager	67	11		
Middle manager	41	74		
Higher manager	61	11		
Other	2	4		

Measures

Leadership behaviors. The Development Leadership Questionnaire (DLQ; Larsson 2006) was used to assess leadership behaviors at each of the three measurement occasions. *Development leadership* is a measure with 21 items designed to measure the three facets: Exemplary model, Individualized consideration and Inspiration and motivation. Sample item: "Acts in accordance with the opinions he or she expresses." *Conventional-positive leadership* is measured with six items covering the facets demand and reward – seek agreements and control – take necessary measures. Sample item: "Aims to reach agreements on what must be done." *Conventional-negative leadership* is also assessed with six items, now measuring the two facets: demand and reward – if, but only if, reward and control – over control respectively. Sample item: "Keeps a log of other people's mistakes." The non-leadership dimension finally, consists of three items designed to measure *laissez-faire leadership*. Sample item: "Avoids making necessary decisions."

Respondents are asked to judge how frequently the person they are rating engages in the specific behavior described by each item. Each behavior is rated on a nine-point frequency scale ranging from never, or almost never (1) to always, or almost always (9). Scale scores were computed by adding the raw scores of the items representing the scale and dividing the sum by the number of items (scale scores could range from 1 to 9).

Reliability (Cronbach's alpha) was computed on each scale at each measurement occasion within the leader and subordinate rater groups respectively. Among the leaders, all coefficients were higher than 0.74 except on conventional-negative leadership Time 1 (0.64) and conventional-negative leadership Time 3 (0.71). Among the raters, all coefficients were 0.89 or higher.

Statistics

Within-group comparisons across the three measurement occasions were done with multivariate analysis of variance, repeated measures design. Within-group comparisons between each pair of assessment times were performed with the *t*-test (paired samples). Between-group comparisons on each occasion were made using one-way analysis of variance followed by Scheffé tests, the most conservative post-hoc comparison method with regard to Type I errors, or with *t*-tests (independent groups) when only two independent groups were compared.

Cluster analysis was performed to obtain groups of leaders with similar characteristics across the different types of leadership behaviors. The idea was to supplement the variable approach described above with a pattern seeking person approach. The cluster analysis (nearest centroid sorting, Anderberg, 1973) was based on the subordinate raters' evaluations of their respective leader (12 indicators; each of the four leadership behavior scales on each of the three assessment occasions). The ratio between the between-clusters and the within-clusters mean square was tested with analysis-of variance. In addition to the within and between group comparisons described above, chi-square tests were used to test the statistical significance of differences in proportions between the profiles regarding the rated leader's position and the length of time the rater had known the leader, as well as the raters' sex, age, education and organizational level. Statistical significance was assumed at $p < 0.05$.

Ethics

The study was approved by a Swedish Research Ethical Committee (EPN Dnr 2012/1905-31/5).

RESULTS

Drop-out analysis

The final study sample (59 leaders with their 361 connected subordinate raters) was compared to the individuals in a norm group data base at the Swedish Defence University including about 3,000 leaders and 17,000 subordinate raters. The present study group (both leaders and raters) consisted of a significantly higher proportion of women, a lower proportion of persons 29 years old or younger, and a higher proportion of individuals with a university education. A higher proportion of leaders in the study group were middle managers while more in the norm group were line managers. Nobody in the study group was employed in the Swedish Armed Forces or Police, while this was the case for about one third in the norm group.

Within-group comparisons across time

Table 2 shows the mean scores of the leaders and the averages of how they were evaluated by their respective raters on each of the three assessment occasions. Repeated measures analysis of variance showed statistically significant mean differences on all scales across the three measurements within both groups. Looking at the leaders' self-ratings, a significant increase of developmental leadership behaviors was noted between the two last measurements, as well as a significant reduction of conventional-negative leadership behaviors between the first and the last assessment. Turning to the raters' evaluation of their leaders, an increase of developmental leadership, and a reduction of conventional-negative and laissez-faire leadership respectively, can be seen. Four out of five of the differences on the negative scales remained statistically significant after Bonferroni corrections. In summary, Hypothesis 1 was not supported among the leaders themselves and weakly supported within the rater group (only without a Bonferroni correction). Turning to Hypothesis 2, it was weakly supported among the leaders and strongly supported by the raters.

Between-group comparisons on each measurement occasion

Comparisons between the mean scores of the leaders and the raters were done on each assessment occasion using independent *t* tests. The actual mean scores can be seen in Table 2. On the pre-course measurement, the raters gave the leaders significantly higher evaluations (less favorable) on the conventional-negative and laissez-faire leadership behavior scales respectively. On the second occasion, the same difference remained on the scale designed to measure laissez-faire leadership behaviors. No differences were found between the two groups on the last measurement occasion. After Bonferroni corrections, only the mean difference in the laissez-faire scale at the pre-course assessment remained significant.

Leadership behavior profiles

The twelve leadership behavior scales were entered into a cluster analysis (see Table 3). Significant *F*-values ($p < 0.001$) were obtained on all scales (not shown in Table 3). This indicates that there is more variability between the clusters than within them.

Table 2. *Leadership behavior scales – within-group comparisons*

Leaders' self-evaluation

	1. One month before the course		2. One month after the course		3. Six months after the course				Paired *t*-tests[c]		
Scale[a]	*M*	*SD*	*M*	*SD*	*M*	*SD*	Wilks Lambda	p[b]	*1–2*	*1–3*	*2–3*
Developmental leadership	7.15	0.77	7.15	0.69	7.27	0.76	0.007	**0.000**			A
Conventional-positive	7.20	0.86	7.14	0.81	7.19	1.01	0.010	**0.000**			
Conventional-negative	2.08	0.78	1.88	0.93	1.78	0.67	0.081	**0.000**		B	
Laissez-faire	1.63	1.05	1.61	1.02	1.64	1.31	0.209	**0.000**			

Raters' evaluation of their leaders

	1		2		3				Paired *t*-tests		
Scale[a]	*M*	*SD*	*M*	*SD*	*M*	*SD*	Wilks Lambda	p	*1–2*	*1–3*	*2–3*
Developmental leadership	6.98	0.96	7.14	0.80	7.27	0.96	0.010	**0.000**	A	**B**	C
Conventional-positive	6.94	1.03	7.05	1.00	7.10	1.06	0.019	**0.000**		B	
Conventional-negative	2.40	1.00	2.01	0.55	1.81	0.62	0.063	**0.000**	**A**	**B**	**C**
Laissez-faire	2.20	1.05	1.97	0.75	1.85	0.62	0.091	**0.000**	A	**B**	

Notes: [a]Scale scores could range from 1 (little use) to 9 (much use). [b]Significance of *F* of Wilk's Lambda (bold text indicates statistical significance after Bonferroni correction). [c]Paired *t*-tests: A = significant difference time 1–2, B = significant difference time 1–3, C = significant difference time 2–3 (bold text indicates statistical significance after Bonferroni correction).

A solution with three profiles was considered as being meaningful. The leaders in the first profile ($n = 32$) have favorable scores on all leadership behavior scales on all three measurement occasions. The leaders in the second profile ($n = 21$) are rated in between and the leaders in the third profile ($n = 6$) have the poorest ratings on all scales. Compared to the norm group data base at the Swedish Defence University mentioned above, all mean scores of the leaders in profile I on the developmental leadership and conventional-positive scales are above the 67th percentile, and all mean scores on the conventional-negative and laissez-faire scales are below (more favorable) the 50th percentile. The mean scores of the leaders in profile III are below the 5th percentile on the Development leadership and conventional-positive scales and above (less favorable) the 75dth percentile on the conventional-negative scale and above the 90th percentile on the laissez-faire scale.

Within-profile group comparisons across time show weak improvements among the leaders in profile I (only without Bonferroni correction). The leaders in profile II note a significant reduction of conventional-negative and laissez-faire leadership. The leaders in profile III note improvements on all scales across time but it is only statistically significant on the laissez-faire leadership scale.

Between-profile group comparisons were made using one-way analysis-of-variance followed by Scheffé tests. All overall comparisons were significant at the $p < 0.001$ level except for the conventional-negative leadership scale ($p = 0.004$). The Scheffé tests showed that almost all pair-wise comparisons were statistically significant.

Other sub-group comparisons

Sub-group comparisons were performed within the leader group and the rater group respectively, on the variables sex, age, education, the level of the leader's position and the length of time the rater had known the leader. Statistically significant differences were found between male and female leaders (self-ratings by the leaders themselves as well as ratings by their subordinates) on one scale, conventional-negative leadership at the second measurement occasion. In both groups of raters, men received higher (less favorable) scorers on this scale ($p < 0.001$). No other significant differences were found between any of the above-mentioned subgroups.

Table 3. *Comparison of clusters of leaders based on the raters' evaluations*

Favorably rated leaders ($n = 32$)

	1. One month before the course		2. One month after the course		3. Six months after the course				Paired *t*-tests[c]		
Scale[a]	*M*	*SD*	*M*	*SD*	*M*	*SD*	Wilks Lambda	*p*[b]	*1–2*	*1–3*	*2–3*
Developmental leadership	7.56	0.50	7.65	040	7.83	0.43	0.001	**0.000**		B	C
Conventional-positive	7.60	0.43	7.66	0.44	7.76	0.48	0.002	**0.000**			
Conventional-negative	1.82	0.44	1.70	0.39	1.57	0.49	0.039	**0.000**		B	
Laissez-faire	1.54	0.38	1.62	0.52	1.43	0.33	0.028	**0.000**			C

Medium rated leaders ($n = 21$)

	1		2		3				Paired *t*-tests		
Scale[a]	*M*	*SD*	*M*	*SD*	*M*	*SD*	Wilks Lambda	*p*	*1–2*	*1–3*	*2–3*
Developmental leadership	6.67	0.52	6.82	0.50	6.80	0.44	0.003	**0.000**			
Conventional-positive	6.59	0.64	6.80	0.45	6.70	0.64	0.004	**0.000**	A		
Conventional-negative	3.06	0.93	2.32	0.40	2.11	0.62	0.024	**0.000**	**A**	**B**	
Laissez-faire	2.57	0.73	2.09	0.57	2.24	0.48	0.031	**0.000**	**A**		

Unfavorably rated leaders ($n = 6$)

	1. One month before the course		2. One month after the course		3. Six months after the course				Paired *t*-tests[c]		
Scale[a]	*M*	*SD*	*M*	*SD*	*M*	*SD*	Wilks Lambda	*p*[b]	*1-2*	*1-3*	*2-3*
Developmental leadership	4.97	0.85	5.54	0.48	5.91	0.54	0.003	**0.000**			
Conventional-positive	4.69	0.36	4.70	0.51	4.92	0.90	0.004	**0.000**			
Conventional-negative	3.23	1.46	2.61	0.70	2.07	0.76	0.050	**0.019**			
Laissez-faire	4.47	0.46	3.40	0.48	2.72	0.47	0.003	**0.000**	A	**B**	C

Notes: [a]Scale scores could range from 1 (little use) to 9 (much use). [b]Significance of *F* of Wilk's Lambda (bold text indicates statistical significance after Bonferroni correction). [c]Paired *t*-tests: A = significant difference time 1–2, B = significant difference time 1–3, C = significant difference time 2–3 (bold text indicates statistical significance after Bonferroni correction).

DISCUSSION

The aim was to evaluate effects of leadership courses based on the Developmental leadership model at the leadership behavior level. Hypothesis 1 predicted that course participants would exhibit an increase of leadership behaviors reflecting the leadership styles developmental leadership and conventional-positive leadership respectively. Looking at the leaders' self-ratings, the hypothesis was not supported. Among the subordinate raters it was weakly supported (only without Bonferroni corrections). Hypothesis 2 predicted that course participants would exhibit a reduction of leadership behaviors reflecting the leadership styles Conventional-negative and Laissez-faire leadership respectively. This hypothesis was weakly supported among the leaders themselves and strongly supported among their subordinate raters.

Although the longitudinal tendency was in the predicted direction, the lack of clear behavioral changes on the leadership styles Developmental leadership and Conventional-positive leadership was a disappointment. On the other hand, the evident favorable behavior changes noted on the leadership styles conventional-negative and laissez-faire leadership, particularly in the eyes of the subordinates, was promising. Recent studies on the dark side of leadership (Einarsen, Schanke Aasland & Skogstad, 2007; Krasikova, Green & LeBreton, 2013; Larsson, Fors & Nilsson, 2012; Skogstad, A., Hetland, GlasØ & Einarsen, 2014), indicate that destructive leadership may cause more negative consequences than constructive leadership contributes to favorable outcomes. A possible reason for the significant effect on the negative behaviors, and the limited effect on the positive ones, is that the negative behaviors may be more visible and clear-cut, while the positive behaviors such as listening and providing support may be less distinct and more difficult to change.

Hypothesis 3 predicted that all course participants, irrespective of their leadership behaviors at the onset of the course, would exhibit the favorable behavioral changes presented in Hypotheses 1 and 2. The hypothesis was explored in two ways. First, conventional, longitudinal subgroup comparisons based on sex, age, education, the level of the leader's position and the length of time the rater had known the leader, were performed but yielded few statistically significant differences.

Hypothesis No. 3 was also tested using a cluster analysis based on the subordinates' ratings. A solution with three profiles was regarded as meaningful. The leaders in the favorably rated profile received significantly more desirable ratings than the other leaders. They also scored more favorably than a large norm group. This group of leaders improved slightly across time (but only without Bonferroni corrections). The lack of stronger improvements in this group could be due to their favorable ratings already at the beginning of the program, a so-called ceiling effect.

Turning to the leaders in the medium and unfavorable profiles, significant improvements are noted, particularly on the negative leadership behavior scales. The case of the six leaders in the unfavorable profile deserves special attention. Their mean scores in the beginning were extremely poor compared to the large norm group. They noted considerable improvements across time on all leadership behaviors except for the conventional-positive scale. Despite the small size of this group, the change was statistically significant on the laissez-faire scale. The limited group size obviously calls for caution when interpreting the result, but could be seen as a promising hypothesis for future research that "there is hope even for the worst."

A gradual reduction of significant differences between the leaders' self-ratings and the evaluation provided by their subordinates was noted. This can be interpreted as another positive marker of intervention effects. Previous studies have shown that leaders who rate themselves similarly to how others rate them are likely to be more effective leaders (Atwater & Yammarino, 1992).

The almost complete lack of subgroup differences on the variables sex, age, education, the level of the leader's position and the length of time the rater had known the leaders an interesting finding in itself. It suggests that the effects of this kind of leadership course on leadership behaviors cuts across these kinds of background conditions. More studies using larger sample sizes are obviously needed to substantiate this finding.

The longitudinal design is a strength, in the words of Day *et al*. (2014) it was even "true longitudinal" in this case in that it involved the measurement of the same indicators of leadership at three points of time. The theory-based developmental leadership model and the use of the well-established measurement tool DLQ are other strong parts. The use of leaders' self-ratings as well as their respective subordinates' ratings, with a premium put on the latter group in the analyses, is a further strength. Finally, we consider the addition of a person-oriented approach as a complement to the conventional variable-orientation to be a strong part of the study.

We will now turn to weaknesses of the study. First, it is based on a comparatively small sample. The high number of drop-outs obviously constitutes a problem. We believe this can be attributed to three factors. One reason could be the indirect relationship and long distance between the researchers and the participants, the subordinates in particular. The invitation to participate, and the information about the research, was communicated via e-mail. In our experience, it is easier to motivate participation in direct face to face encounters. However, this was not possible to arrange in this study.

A second reason for the high number of dropouts could be a general "questionnaire tiredness" in modern society where one is frequently asked to respond to all kinds of investigations. A third factor is the common and almost inevitable loss of participants in longitudinal studies stretching over a ten month period as was the case in the present study. Taken together, this calls for caution when generalizing the results.

Another weakness of the study is the fact that the subordinates were selected by their leaders and that we did not have a control group. Furthermore, no potential individual antecedent conditions such as personality were used and there was a lack of contextual analysis of the work environments of the participants. Additionally, no real-world outcome measures were used (see e.g., Kaiser, Hogan & Craig, 2008). Thus, we do not know if the observed behavioral changes led to changes in other outcomes such as higher productivity, higher job satisfaction, improved health and well-being, etc. Better handling of these study weaknesses is our main suggestion for further research.

In addition to this, and in particular response to the common problem of high drop-out rates in longitudinal questionnaire-based studies, we recommend the use of qualitative follow-up studies. Although weak in terms of reliability, such studies could add valuable and valid in-depth knowledge on pros and cons of leadership development interventions.

Finally, turning to practical implications, we regard the limited yet positive results of the intervention as promising. In particular, the reduction of the negative leadership behaviors is important and this aspect should be refined during leadership development courses. Similarly, the tendency for even the very poorest leaders to improve is promising for future interventions.

This research was sponsored by AFA Insurance and the Swedish Defence University.

REFERENCES

Anderberg, M. R. (1973). *Cluster analysis for applications*. New York: Academic Press.

Atwater, L. E. & Yammarino, F. J. (1992). Does self-other agreement on leadership perceptions moderate the validity of leadership and performance predictions? *Personnel Psychology*, *45*, 141–164.

Bass, B. M. (1997). Does the transactional-transformational leadership paradigm transcend organizational and national boundaries? *American Psychologist*, *52*, 130–139.

Bass, B. M. (1998). *Transformational leadership: Industry, military, and educational impact*. London: Lawrence Erlbaum Publishers.

Day, D. V. (2000). Leadership development: A review in context. *The Leadership Quarterly*, *11*, 581–613.

Day, D. V., Fleenor, J. W., Atwater, L. E., Sturm, R. E. & McKee, R. A. (2014). Advances in leader and leadership development: A review of 25 years of research and theory. *The Leadership Quarterly*, *25*, 63–82.

Dvir, T., Eden, D., Avolio, B. J. & Shamir, B. (2002). Impact of transformational leadership on follower development and performance: A field experiment. *Academy of Management Journal*, *45*, 735–744.

Einarsen, S., Aasland, M. S. & Skogstad, A. (2007). Destructive leadership behaviour: A definition and conceptual model. *The Leadership Quarterly*, *18*, 207–216.

Endler, N. S. & Magnusson, D. (1976). Toward an interactional psychology and personality. *Psychological Bulletin*, *83*, 956–979.

Grint, K. (2007). Learning to lead: Can Aristotele help us to find the road to wisdom? *Leadership*, *3*, 231–246.

Kaiser, R. B., Hogan, R. & Craig, S. B. (2008). Leadership and the fate of organizations. *American Psychologist*, *63*, 96–110.

Krasikova, D. V., Green, S. G. & LeBreton, J. M. (2013). Destructive leadership: A theoretical review, integration, and future research agenda. *Journal of Management*, *39*, 308–1338.

Larsson, G. (2006). The Developmental Leadership Questionnaire (DLQ): Some psychometric properties. *Scandinavian Journal of Psychology*, *47*, 253–262.

Larsson, G., Carlstedt, L., Andersson, J., Andersson, L., Danielsson, E., Johansson, A. *et al*. (2003). A comprehensive system for leader evaluation and development. *Leadership & Organization Development Journal*, *24*, 16–25.

Larsson, G., Fors Brandebo, M. & Nilsson, S. (2012). Destrudo-L: Development of a short scale designed to measure destructive leadership behaviours in a military context. *Leadership & Organization Development Journal*, *33*, 383–400.

Nystedt, L. (1997). Who should rule? Does personality matter? *European Journal of Personality*, *11*, 1–14.

Reichard, R. R. & Avolio, B. J. (2005). Where are we? The status of leadership intervention research: A meta-analytic summary. In W. L. Gardner, B. J. Avolio, & F. O. Walumbwa (Eds.), *Authentic leadership theory and practice: Origins, effects and development* (pp. 203–223). Amsterdam: Elsevier Jai.

Skogstad, A., Hetland, J., GlasØ, L. & Einarsen, S. (2014). Is avoidant leadership a root cause of subordinate stress? Longitudinal relationships between laissez-faire leadership and role ambiguity. *Work & Stress*, *28*, 323–341.

Yukl, G. A. (2005). *Leadership in organizations* (6th edn). Upper Saddle River, NJ: Prentice Hall.

Received 22 June 2016, accepted 2 October 2016