An Evaluation of Sparebanken Vest's Leadership Development Program

Alexander Grande & Julie Synes Nordvik

MAPSYK345

Master's Program in Psychology

Field of Study: Work- and Organizational Psychology

at the

UNIVERSITY OF BERGEN

FACULTY OF PSYCHOLOGY

SPRING 2022

Academic Supervisor: Magnus Odéen, Sparebanken Vest

Academic Co-supervisor: Iselin Reknes, Department of Psychosocial Sciences

Abstract

Leadership development is one of the world's most popular organizational interventions. This thesis is a practical evaluation of Sparebanken Vest's semi-annual leadership development program. It aims to assess whether the program yields positive behavioral change amongst the participants regarding Hunger & Leadership, and Feedback. Hunger is understood as drive for organizational results and innovation, while feedback is understood as the employees' ability to give and receive constructive criticism. The data was generated from two quantitative surveys – the first one answered by 64 of 86 participants in November and a post-survey answered by 45 participants in February. The goal was to compare the means between our two points of measurement. The total mean percentage of female respondents was 33.75%. We present two hypotheses: i) whether the program was overall effective when measuring *hunger & leadership* and *feedback*, and ii) whether the program was as effective for women as it was for men. The analysis for hypothesis one showed no significant change in mean levels of hunger & leadership (MD = .056, t = .487) and feedback (MD = .992, t = -.01). Hypothesis 1 was therefore not supported. However, the results showed that the participants had high levels before the program, indicating a possible ceiling effect. Our findings further indicated no significant difference in the program's effectiveness for men (MD = .068, t = .470, MD = .085, t = .728) and women (MD = .038, t = .198, MD = -.156, t = -.977) in the categories. The only significant difference we found was already present before the program. Hypothesis 2 was thus supported.

Keywords: Leadership development, leader development, leadership, gender.

Sammendrag

Leder- og ledelsesutvikling er ett av verdens mest populære organisatoriske intervensjoner. Denne avhandlingen er en praktisk evaluering av Sparebanken Vest sitt halvårlige ledelsesutviklingsprogram. Avhandlingens mål er å undersøke positiv atferdsendring har forekommet som følge av programmet, spesifikt innenfor kategoriene Sult og Ledelse, og Tilbakemeldinger. Sult forstås her som driv for resultater og innovasjon, mens tilbakemeldinger omhandler hvor godt medarbeiderne kan gi og motta konstruktiv kritikk. Dataen var generert av to kvantitative spørreundersøkelser – den første ble besvart av 64 av 86 deltagere i november, og en post-spørreundersøkelse besvart av 45 deltagere i februar. Målet var å sammenligne gjennomsnittet på tvers av måletidspunktene. Totalt gjennomsnitt av kvinnelige respondenter var 33.75%. Vi presenterer to hypoteser: i) om programmet generelt var effektivt ved måling av *sult og ledelse* og *feedback*, og ii) om programmet var like effektivt for kvinner som det var for menn. Analysen for hypotese én viste ingen signifikant endring i gjennomsnittsnivåene til sult og ledelse (*MD* = -.06, *t* = -.06) og (*MD* = 0, *t* = -.01. Hypotese 1 var dermed ikke støttet. Resultatene viste imidlertid høye nivåer før programmet, noe som indikerer en mulig takeffekt. Våre funn viste videre at det ikke var noe signifikant forskjell i effekten for menn (*MD* = .068, *t* = .470, *MD* = .085, *t* = .728) og kvinner (*MD* = .038, *t* = .198, *MD* = -.156, *t* = -.977) i kategoriene. Den eneste signifikante forskjellen vi fant var allerede til stede før programmet. Hypotese 2 var dermed støttet.

Nøkkelord: Ledelsesutvikling, lederutvikling, ledelse, kjønn.

Acknowledgements

Our time as master's students is soon over. We wrote this thesis in the third and fourth semesters of our master's program in Work- and Organizational Psychology, spanning eight months. Writing such a long and methodological thesis is an arduous yet rewarding task. Writing a practical thesis is more than new to us as we have primarily been writing theoretical theses these last couple of years. However, due to a common interest in the role of gender in leadership development programs, we chose to work with each other. We would first and foremost thank our academic supervisor Magnus Odéen. Thank you for choosing us. We would have been completely lost were it not for your expertise and patience. We appreciate that you also included us in SPV's training program and let us gain a first-hand look at how these programs are conducted. You have been a constant help and have replied quickly by e-mail and phone when we were lost. We would also like to thank our co-supervisor, Iselin Reknes. You saved us much trouble when we had methodological issues in our thesis and were always available for a conversation. We would also like to thank Tina Ødegård at SPV for making this whole thing possible and greenlighting the entire project. Last but not least, we would like to thank each other. Not only are we best friends, but each other's best sparring partners. Though we had established this relationship long before this thesis, these last eight months have brought our friendships to new levels. We would not have been able to do this without each other. Thank you.

Bergen, April 21st, 2022.

Alexander Grande and Julie Synes Nordvik

Table of Contents

An Evaluation of Sparebanken Vest's Leadership Development Program

The current thesis aims to evaluate the effectiveness of a leadership development program administered by Sparebanken Vest ("SPV") by measuring behavioral change with self-rated surveys. Hence, this thesis is a practical case study. The program's goal is to induce positive behavioral change by improving participants' initiative, drive, and feedback skills. The themes for the program were *hunger*, inspired by the book "Hunger in paradise: How to Save Success from Failure" (Ankersen, 2016), and *feedback* inspired by the article "Find the Coaching in Criticism" (Heen & Stone, 2014).

Paraphrasing from Ankersen's (2016) examples, hunger is understood as initiative and drive to stay on top. On the other hand, feedback is understood as being able to give, receive and work with constructive criticism (Heen & Stone, 2014). As such, we generated questions that focused on these two topics. SPV chose to focus on these areas due to strategic reasons. We did a pre-and post-analysis of the results from the self-reporting surveys to discover a possible behavioral change. A post-analysis is necessary as behavioral change takes time (Day et al., 2014; Prochaska & DiClemente, 1982).

Leadership development is one of the most popular organizational interventions, aiming to attract, retain, and develop leaders. As effective leaders wield, exert, and perform leadership daily, it becomes evident that leadership should be continuously developed. It is thus believed that effective leaders equal organizational success (Cummings et al., 2015b; Hrivnak et al., 2009).

Organizations invest part of their resources in external or internal leadership development programs. The tendency to invest considerable sums in these programs may be because the organizational results connect to how well the leaders perform, which is why it is essential to retain good leaders. By offering developmental opportunities, the organization would not only keep its leaders but might also increase the leaders' organizational

commitment (Cardy & Lengnick-Hall, 2011; Herman, 2005; Kroon & Freese, 2013). Furthermore, developmental opportunities could also increase employer attraction (Backhaus & Tikoo, 2004). Consequently, a lucrative business has sprawled; organizations worldwide spend billions of dollars on training management and leaders (Crawford & Kelder, 2019; Riggio, 2008; Suvatne, 2020).

The demand for effective leaders has arisen due to competition in the industry. Organizations have become more aware of the need to attract and retain effective leaders to ensure a competitive advantage (Amagoh, 2009). Effective leaders are believed to creatively address challenges, motivate employees, respond to changes in the environment, and maintain high organizational performance (Vardiman et al., 2006). Nevertheless, the successful development of leaders rests upon the administration of a successful leadership development program. The literature shows that leadership development initiatives have either positive effects or no outcome (Avolio et al., 2009; Collins & Holton III, 2004). Hence, it may be vital to follow and invest in a systematic leadership development approach to develop effective leaders, potentially yielding satisfactory long-term results.

A successful systematic leadership development program depends on multiple factors, such as competent facilitators, delivery method, and utility of a given program (Lacerenza et al., 2017). To investigate the utility of a given program, the organization should conduct a needs assessment. This may be equivalent to a diagnosing framework in organizational development. Cummings et al. (2015a) explain diagnosing as a continuous process of understanding how the organization is functioning in its current state. This further provides information that may be vital to either design or tweak a given intervention. By not diagnosing the program's utility, the intervention may be presented at the wrong place or at the wrong time, consequently racking up the program's cost without a need for such an intervention.

A leadership development program reaches its objectives and has utility when individual leaders successfully learn new skills. These skills lead to a financial return for the company. Additionally, the company improves its ability to teach new skills through training (Van Velsor et al., 2010). Such *returns* may be interpreted as satisfactory long-term results.

A framework for measuring the general effectiveness of a leadership development program, and its potential returns, is Kirkpatrick's (1979) four steps of evaluation:

1) *reactions* towards the program, 2) *learning* after the program, 3) the transfer of learning into *behavior*, and 4) *results*.

By following Kirkpatrick's four steps, we will assess the effect of the current program, focusing on behavioral change post-program. Furthermore, we include gender as a variable in the analysis. As SPV believes in gender equality measures, though with a gender-neutral focus, they aim to prioritize competence, developable attributes, and a diverse talent pool. Evidence of this may be found in their top leadership. The board constitutes six women out of ten individuals – 60 %, while the corporate leadership consists of three women out of eight individuals – 37.5 % (Sparebanken Vest, 2020).

These numbers contrast with the general statistics of Norwegian companies. In general, women constitute 11 % of board leaders, 14 % of CEOs, 25 % of top leadership groups, and 29 % of boards (CORE - Centre for Research on Gender Equality, 2021). Although SPV has a relatively representative leadership group, one might find men in top leadership positions more frequently.

Research finds little difference when comparing men's and women's leadership styles and effectiveness. Women are thought to be more charismatic, change-oriented, and somewhat more democratic in their leadership style than men. Male leaders are thought to lead more task-oriented than female leaders. However, there are no overall differences in the perceived effectiveness between female and male leaders (Eagly & Johnson, 1990; Halvorsen

& Johansen, 2013). Recognizing that men and women are more similar than different, some differences do exist (Hopkins et al., 2008). Research shows ambiguous results regarding the efficacy of mixed leadership development programs versus women-only leadership development programs. Some argue that men and women have different approaches to information processes, motivation, and responses to stress (Ruderman & Ohlott, 2005). Others suggest that men and women differ in the definition of success. Women tend to define career success as a desire for self-development and intrinsic rewarding roles. At the same time, men view success as achieving status and a high salary (Sturges, 1999).

These arguments raise the question of the effectiveness of mixed leadership development programs. However, Ely et al. (2011) argue that both types of programs may be used to achieve different goals. Additionally, Burke and Day (1986) found managerial training moderately effective in increasing knowledge and performance for men and women. For these reasons, it is natural to include gender as a variable to investigate whether SPV´s leadership development program positively affects both genders.

Theoretical Framework

What is Leadership?

To understand leadership development, one must understand the term *leadership*. However, no universal definition has been agreed upon (Northouse, 2019), likely due to different fields of focus throughout the times and the term's complexity.

The early 20th century emphasized traits that inspired *Great Man Theories.* These theories asserted famous leaders' characteristics as a recipe for success and were based on Carlyle's (1841) ideas. This focus on personal traits further inspired Stogdill (1948) to identify characteristics that may signify leaders. He identified the following: alertness, insight, intelligence, self-confidence, and sociability as some of the traits that signify leaders. However, he introduced the importance of situational factors. For example, a leader might

have certain traits that might be good in one situation and less in another. Stogdill's findings clarify that leadership is more complex than individual characteristics alone.

Due to several failed attempts at identifying common traits among effective leaders, perspectives that purely favor traits have fallen into obscurity (Nawaz & Khan, 2016). Scholars have since realized that leadership is a far too complex concept to fit one standard definition (Northouse, 2019). Still, Northouse has identified four central components. This thesis will thus be operating with Northouse's (2019, p. 5) definition:

"*Leadership is a process whereby an individual influences a group of individuals to achieve a common goal.*".

Northouse's definition interprets leadership as a *process* influenced by the characteristics of *leaders*, *co-workers*, and a *given situation* (Skogstad, 2015). Consequently, effective leaders may lead to success, indicating that investing in leadership development programs might be vital for organizational achievement. Exerting leadership may therefore be a prerequisite for satisfactory organizational results.

What is Leadership Development?

The challenge of coherently defining leadership also affects the definition of leadership development. This is illustrated in how the literature uses *leader* – and *leadership* development interchangeably. However, there is a distinction; *leader development* refers to developing the individual and their traits, skills, knowledge, and abilities as a leader (Day et al., 2014; Lind, 2007; Tronsmo, 1996). In other words, *leader development* is an investment in human capital, aiming to increase their ability to lead by improving their skills, self-awareness, and motivation to lead (Day, 2000; Lind, 2007; Tronsmo, 1996).

Meanwhile, *leadership development* operates on a group level in a social context (Day et al., 2014), further concentrating on the relationships between a given leader and subordinates - also referred to as social capital. Opposite to human capital, social capital

underlines building relationships that enhance teamwork and resource exchange, potentially leading to innovation (Bouty, 2000; Tsai & Ghoshal, 1998).

There are several reasons why the terms will be used interchangeably as "leadership development." Firstly, there is less research on leadership development than on leadership in general. Consequently, there is little to no agreement on a coherent definition (Day et al., 2014). Secondly, leadership development is a comprehensive and widely used term. For example, in 1988, Henrik Holt Larsen presented 50 different approaches for leader - and leadership development (as cited in Bergheim et al., 2007). This shows that dozens of various interventions may be grouped as "leadership development." Day (2000) highlights that either approach is incomplete by itself. In developmental programs, the focus tends to lie on both the individual and the social context and organizational objectives.

Hence, we base our understanding of leadership development as systematically tailored to promote leadership knowledge, - skills, and - abilities (Day, 2000). Therefore, all forms of leader -, managerial -and supervisory programs and workshops aiming to develop knowledge, skills, and abilities will subsequently be included in this understanding. This flexibility permits individual and collective development with different interventions, depending on what is to be developed. The flexibility may also provide an opportunity to systematically tailor a program to the organization's needs. This does not refer to a set-in-stone approach but merely a framework. Scholars argue that using a conceptual framework allows organizations to identify key considerations in designing, developing, delivering, and evaluating a systematic approach to leadership development programs. Indeed, the effectiveness of any development program depends on how well it is designed (Hrivnak et al., 2009).

How to construct a leadership development program using a framework

A framework for constructing a leadership development program may include; *Needs assessments, setting instructional objective*s, *design*, *delivery*, and *evaluation* (Cummings et al., 2015b; Hrivnak et al., 2009)

Needs assessment. The first step of a given intervention is to perform a thorough needs assessment and align the program with these needs (Arthur Jr et al., 2003; Riggio, 2008). A needs assessment seeks to develop clear, measurable objectives for the program. It may ensure that the framework is connected to the overall business strategy, thereby justifying its appliance. If done thoroughly and effectively, the assessment might yield many benefits, such as a return on investment. Research further indicates that tailoring a program based on a needs assessment may yield higher behavioral change and learning (Arthur Jr et al., 2003; Lacerenza et al., 2017; Riggio, 2008).

Develop the objectives and design of the training. Organizations should then establish the outcome objectives for the program. This includes describing the results expected from a competent leader and how those results would be achieved. After that, considering the objectives, the design of the training program should be determined. The design process includes which techniques the program should use to train the participant and determine the duration of the program (Cummings et al., 2015b; Midtun & Jordahl, 2002). For example, suppose the overall objective for the program were to improve self-awareness and competence development in the leaders. In that case, one could use a 360-degree feedback review as a technique (Day, 2000). A 360-degree feedback review is when a manager gets anonymous and confidential feedback from their boss, peers, and subordinates and then reflects with other managers afterward.

Deliver the training. The program can either be facilitated by internal or external consultants (Dalakoura, 2010). After implementing the program, the participants complete the activities included in the design and return to work (Cummings et al., 2015b).

Evaluate the training. After completing the program, the consultant assesses whether the program has met the desired objectives. The goal is to measure if the given program has a form of a return of investment or return of experience. Many methods, specifically Kirkpatrick's four steps, are used as a general framework for such evaluations (Cummings et al., 2015b; Kirkpatrick, 1979).

Kirkpatrick's techniques for evaluating training programs

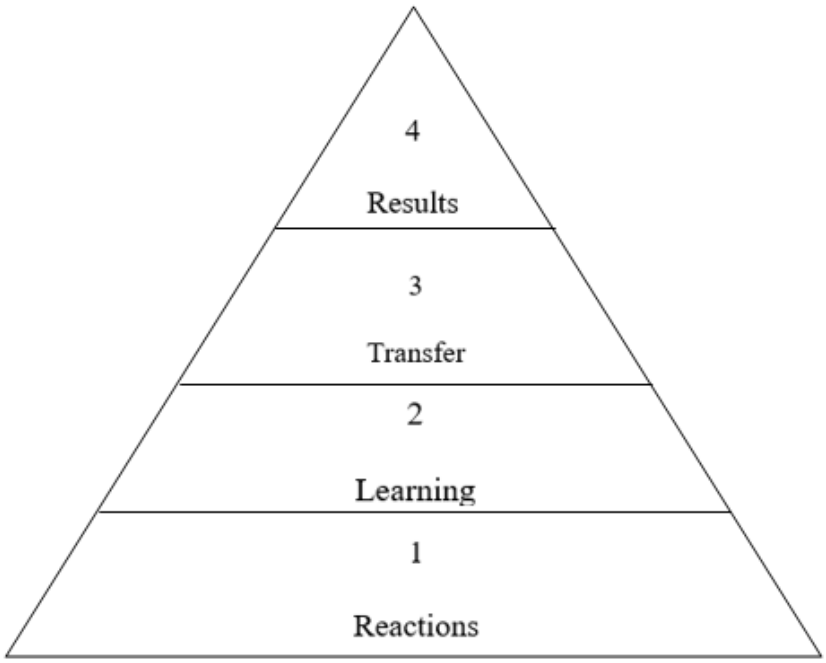
The Kirkpatrick Model is an intuitive, globally recognized, widely used framework for leadership development programs (Alliger & Janak, 1989; Ardent Learning, 2020; Lacerenza et al., 2017; Leskiw & Singh, 2007; Neuberger, 2016)**.** Although leadership development allows for a degree of flexibility, a framework such as Kirkpatrick's allows for a systematic evaluation of whether the program was effective or not. We will thus anchor our practical understanding of evaluations in Kirkpatrick's techniques, formulated as four steps:

1) reaction, 2) learning, 3) behavioral change, and *4) result*s, each representing different stages of the program.

These steps aim to stimulate the facilitators to analyze and assess each step in-depth. We emphasize that the future of SPV's training program depends on the ability to both evaluate and use the results of said evaluation. Accordingly, Kirkpatrick (1979) recommends evaluating programs constantly, especially after administering them. This should ensure that the program has been successful and indicate whether the desired results have been reached. If not, Kirkpatrick (1979) advises altering some of the steps in the program.

Figure 1

Kirkpatrick's framework for leadership development programs.



Reaction. The first step equates to the general approval of the program – which indicates subjectivity (Kirkpatrick, 1979). These *reactions* are measured right after the program, providing information on the immediate reactions. Exclusively adverse reactions and dissatisfaction might indicate less transfer of knowledge post-program. This suggests that the program needs further planning, revitalization, and change. As such, analyzing reactions would allow a consultant to determine whether the program was satisfactory and if they should tweak certain aspects of the program. Kirkpatrick advocates using surveys that may capture as much as possible to measure reactions. In doing this, he recommends consultants follow five principles:1) *to determine what to find out*, 2) *to use a written comment sheet covering the items determined in the first step*, 3) *to design the form so the questions may be tabulated and quantified*, 4) *to obtain the subjects' reactions by making the form anonymous*, and 5) *to encourage the subjects to write in additional comments not covered by earlier questions* (1979, p. 78).

Summarized, Kirkpatrick encourages surveys with questions that concentrate on the experiences the individuals will transfer towards the next step. The answers should be quantifiable and aligned with what a given organization wants from a given program.

Learning. Following reactions, Kirkpatrick (1979) recommends that consultants measure learning. As learning is not the focus of the current thesis, a somewhat limited definition of the term will be utilized: "*To what degree principles, facts, and techniques are understood and absorbed by the subjects*" (Kirkpatrick, 1979, p. 92). The definition does not diminish the significance of learning, though, as learning might stimulate behavioral change. Five principles are presented here as well: *1) the actual learning should be quantifiable,* 2) *one should use a pre-and post-approach so that the actual learning might indicate relevance to the program,* 3) *learning should be measured objectively*, 4) *there should both be a control- and an experimental group* and 5) *the results should be analyzed statistically*.

Behavioral Change. "*Behavioral change is the holy grail of leadership development*" (Discovery in Action, 2020). Leadership development programs aim to develop knowledge, skills, and abilities related to leadership. Hence, one may assume previous schematics on how to exercise leadership should be left behind or tweaked in favor of the behaviors taught at the program. In turn, this could indicate that learning might stimulate behavioral change.

Although rooted in social modeling, Sims and Manz (1982) present three distinct types of behavior change:

1) Establishing new behaviors, 2) changing the frequency of existing behaviors, and 3) providing behavioral cues.

Therefore, one could argue that the program's degree of success depends on how much participants have changed their behavior (O'Loughlin, 2013). Hence, behavioral change is the holy grail of leadership development.

Furthermore, even if a leader possesses the abilities to exert leadership, there is no guarantee that the leader will utilize those skills or permit behavioral change. As such, five essential requirements are presented for behavioral change to happen after learning:

1) *wanting to change*, 2) *recognizing one's weakness*, 3) *working in a permissive climate*, 4) *receiving help from someone interested and more skilled*, and 5) *having an opportunity to try out new ideas* (Kirkpatrick, 1979).

Consequently, tweaking existing behaviors might make a leader more susceptible to utilizing necessary skills. For example, a given training program may focus on developing communication skills so that the leaders will be more likely to use different communication tools and – techniques. However, as behavioral change is considered the holy grail, there is a tendency to end the evaluation at this level (Kirkpatrick, 1979; Lacerenza et al., 2017; Prochaska & DiClemente, 1982; Prochaska et al., 1992; Prochaska et al., 2015). Instead, Kirkpatrick advises the consultant to infer the next step.

Results. Rather than viewing behavioral change as the endgame, Kirkpatrick proposes the fourth step. A consultant may assess whether a given program or measure has benefited the organization by measuring results. Examples may be a reduction of turnover, increased motivation, or, as stated in the current program, maintaining drive and improving feedback skills. However, contingent factors may disturb the evaluation process of the results alone. Therefore, it is advised to repeatedly evaluate the program to gain a holistic perspective: capturing the respondents' *reactions* and measuring whether *learning* has occurred, thus stimulating *behavioral change*. Nevertheless, Kirkpatrick (1979) emphasizes that certain jobs allow for a fast and easy evaluation, e.g., the effectiveness of typewriters in typing a total amount of words. Hypothetically, a before- and after-test for the total amount of words per minute is possible. Behavioral change is, on the other hand, both multifaceted and complex. For instance, external factors outside the program may drive the causality of behavior change.

Therefore, not measuring results may present a challenge in stating whether behavioral change is a causal effect of the program. Therefore, a multifaceted evaluation of *reactions*, *learning*, and *behavior* may be necessary, including an evaluation of the long-term goals in the form of a return on investment.

In practice, Kirkpatrick's techniques would have a given consultant measure the participants' reactions to the program. Using anonymous surveys with questions that focus on the experiences of the program would make it easier to predict behavioral change. An example could be, "To which degree will you change your behavior due to the gathering?". These questions should be quantifiable, most commonly by a Likert scale, to allow for quantitative analyses.

Further evaluation would have assessed learning. A common learning method is *classroom performance* (Kirkpatrick, 1979). Typical activities in this method include role-play, discussions after the role play, and demonstrations, which were also used at SPV's gathering. However, this is merely the activity where people learn. Exams are a classic example of measuring learning. However, using 360-degree feedback pre- and post-program could tie learning to observed behavioral change. The given consultant could also supplement the evaluation of behavioral change by administering a self-survey to the participants. This survey could include questions such as "*I feel the hunger for improvement has lately increased among my employees and me*."

Lastly, the consultant should not end their evaluation based on the behavioral change alone. Instead, the consultant should assess the behavioral change's long-term results. For example, suppose the goal is to train the management in creative solutions. The consultant could then evaluate whether different tasks and issues have been solved using new solutions.

Bandura's Social Cognitive Theory

As Kirkpatrick addresses a holistic organizational perspective, understanding individual social processes may need to be understood. For this, one may use Bandura's Social Cognitive Theory ("SCT") (1969, 1986, 2001).

Out of SCT's six constructs, specifically, *self*-efficacy and *observational learning* are relevant. Research shows that these constructs may influence learning (Blume et al., 2010) and possibly advance the extent of behavioral change (O'Loughlin, 2013).

Self-efficacy. Self-efficacy is defined as "*estimates of one's ability to successfully execute the behavior required to produce desired outcomes*" (Dwyer, 2019, p. 637). As self-efficacy might influence willingness to challenge oneself, choice of behavior, the direction of effort, and persistence in meeting obstacles (Bandura, 1986; Gist, 1987; Lent et al., 1987; Schaub & Tokar, 2005), the process may be incremental to learning; even in leadership development programs. According to Blume et al. (2010), there might be moderately strong relationships between self-efficacy and transfer of learning to the work setting, further indicating that self-efficacy may influence behavioral change. As self-efficacy has gradually become popular in organizational research in the last decades (Dwyer, 2019), a definition that delimits itself to leadership has been proposed: "one's *self-perceived capability to perform the cognitive and behavioral functions necessary to regulate group process in relation to goal achievement*" (McCormick, 2001, p. 30).

McCormick's definition highlights leadership self-efficacy as a somewhat neutral term, indicating that it may positively and negatively affect learning. In negative contexts, the individual may learn the contents of the program. However, a poor level of self-efficacy may prevent the individual from truly utilizing what one has learned. On the other hand, in positive contexts, high levels of self-efficacy beforehand may lead to better performance and assessments post-training (Goldstein & Ford, 2002). Consequently, these findings indicate

that self-efficacy might influence behavioral change and hence influence leadership behaviors. This further supports a complementary relationship between performance and self-efficacy.

Individuals with higher levels of self-efficacy tend to be more likely to try out new skills and tasks while also showing confidence (Goldstein & Ford, 2002). On the other hand, low levels of self-efficacy may lead to hesitance when answering questions such as "*I feel I have the competence I need to motivate my employees to strive for improvement*." By using motivational techniques such as praising or constructive criticism, a consultant should try to increase a participant's level of self-efficacy. This could motivate the participants to try out, e.g., a new feedback method.

Observational Learning. Learning does not necessarily occur explicitly in a classroom setting. However, it may also take effect through social modeling, also referred to as observational learning (Bandura, 1969, 1977b) or imitation. When learning by observation, an individual can learn a specific behavior by observing external stimuli (Sims & Manz, 1982), thereby possibly undergoing behavioral change. According to Bandura (1986), four concepts are central to observational learning. *Attention* refers to the observed situation. *Retention* is how observed activities are remembered as abstract symbols, acting as guides (Bandura, 1986). The third concept is *Production*, which converts symbolic conceptions into an appropriate action for a situation. The last factor is *motivation*. As individuals will not necessarily perform everything they observe, Bandura (1986) argues that they recall and act upon what they have observed if there are incentives to do so.

Bandura (1986, pp. 47-48) proposed that observation has a significant role in influencing values, patterns of thought and behavior, and attitudes. This may lead to a greater ability to contextualize and use different behaviors for different situations. This type of

learning may also provide social cues to exercise what one has learned in the past, which may or may not be inhibited (Andersen & Cole, 1990).

In a given program, the participants could observe a role-play by a role model. They could learn the behavior and repeat what the role models have done. In this particular program, SPV emphasized this kind of modeling. The attending leaders were told to participate actively and model the behavior. The principle of social modeling might also relate to self-efficacy, as low levels might lead to less confidence in the modeled behavior (Goldstein & Ford, 2002). Therefore, by considering self-efficacy when conducting role-play, one could easier allow for the transfer of knowledge towards behavioral change.

Nevertheless, SCT should be supplemented with an understanding of the underlying processes of behavioral change. The Transtheoretical Model ("TTM") by Prochaska and DiClemente (1982) might help explain this complexity.

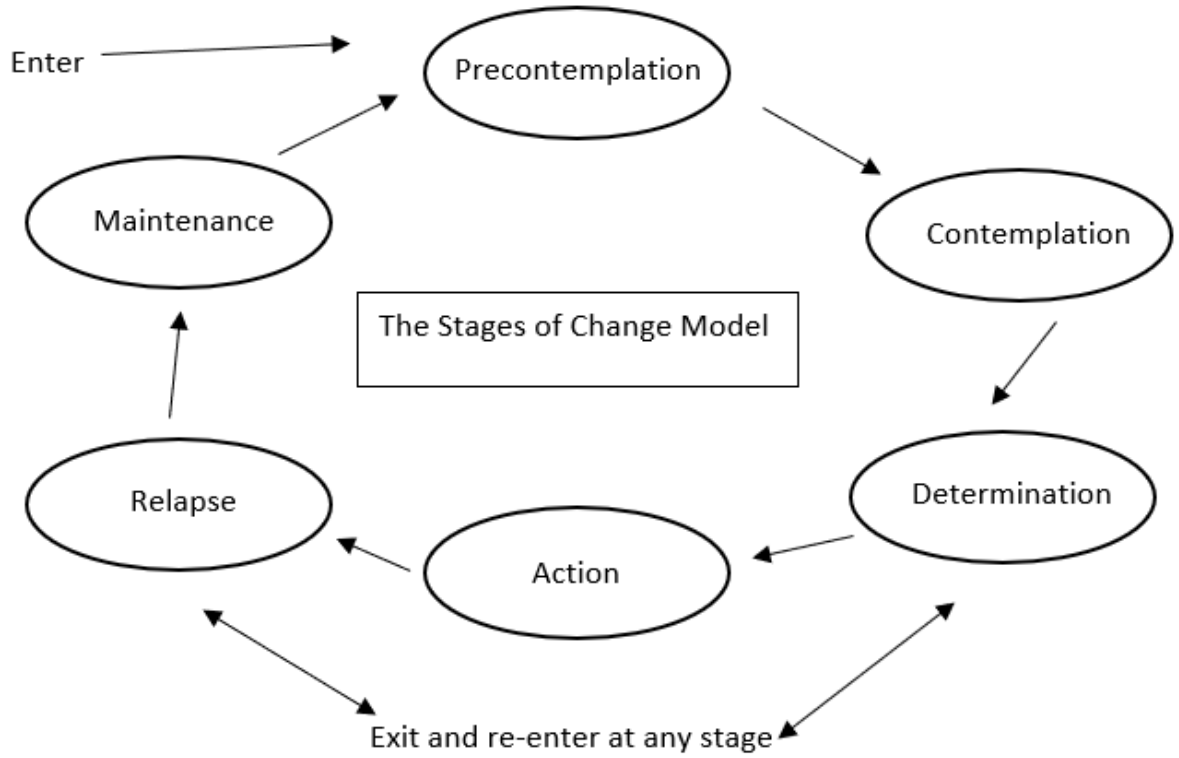
The Transtheoretical Model of Behavioral Change

TTM aims to understand decision-making with the intent to change. The theory was initially conceived to assess the experiences of smokers who quit by themselves and smokers who required further treatment. Aiming to explain why some people can and some cannot change their behavior. Results show that behavioral change may take place if one were ready to do so themselves (Prochaska & DiClemente, 1982; Prochaska et al., 1992)

As the model might clarify individuals' decision-making processes, it may be applied to organizational development. The theory incorporates motivational and cognitive aspects, social learning, and relapse prevention theories. Prochaska and DiClemente (1982) claim that long-term behavior cannot simply change; one must *want* long-term change, indicating a motivational aspect.

Figure 2

Prochaska and DiClemente's (1982) Transtheoretical Model of Change.



Six stages are identified where both overt and covert processes may occur and guide the intervention. Before entering a program, one may start with no intention to change, perhaps even unknowingly resisting it (Riggio, 2008). If a participant does not intend to change their behavior in the near future, one is categorized as a *precontemplator* (Prochaska et al., 1992). By *raising consciousness* over the problem, allowing *dramatic relief*, and conducting an *environmental reevaluation*, the consultant may motivate the participant to contemplate change, thus leading to a greater intention to change.

Moving towards *contemplation*, one might be aware that change is needed, even assessing the pros and cons of change, though with ambiguity towards changing their behavior (Prochaska & DiClemente, 1982; Prochaska et al., 1992). By using the process of *self-reevaluation*, the consultant may aid the participant in preparing to change.

After contemplation, the participant is *prepared* or *determined*, ready to act within a short period of time. With the belief that changing one's behavior may be productive, intention - and behavioral criteria are combined (Prochaska et al., 1992). With the consultant facilitating *self-reevaluation* and *self-liberation* from earlier behaviors, the participant would have an easier time accepting long-term behavioral change.

Specific *actions* may have been made within the last six months. Nevertheless, as actions are directly observable, they tend to equate to behavioral change (Prochange.com, 2020), thus ending the process. However, as behavioral change needs maintenance (O'Loughlin, 2013), the consultant should continue the process to have an actual behavioral change (Prochaska et al., 1992). By actively *self-liberating*, the participant would have an easier time maintaining the behavior.

By maintaining behavioral modifications, one would work towards preventing *relapse* and falling into old patterns. Consequently, one would be increasingly confident in maintaining changes (Prochaska & DiClemente, 1982; Prochaska et al., 1992; Prochaska et

al., 2015). Thus, one would apply change processes less frequently as a behavior already has been somewhat transferred. By addressing the participants' *self-efficacy*, the consultant could raise the participant's confidence, making it easier to maintain their new behavior. To prevent relapsing, the consultant could still guide the participants to substitute earlier behaviors with new ones, find supportive relations that guide the change, reward each other for their new behaviors, and recreate the environment to encourage healthy behavior.

Reaching *termination*, one would leave the circular model, ideally experiencing zero temptation to relapse and 100 % self-efficacy (Prochaska et al., 2015). Still, the consultant should consider that the process is demanding with both trials and errors. The program will be successful if the participants act as if their previous patterns never existed. Failing to maintain a specific behavior may be due to too high criteria (Prochaska et al., 2015), such as demanding behavioral change immediately after action. Therefore, consultants should implement sub-goals to make *termination* easier, which will make it easier for the participants to keep the behavioral change long-term.

Predictions and Hypotheses

By using Kirkpatrick's evaluation techniques, we will measure whether SPV's leadership development program is effective. Based on the presented literature, our understanding of Kirkpatrick's techniques will be supplemented by primarily two theories. As Kirkpatrick's definition of learning is somewhat limited, we will apply Bandura's SCT. Understanding SCT might shed light on how to understand learning and how it might occur implicitly by constructs such as self-efficacy and social modeling. Further, Kirkpatrick's techniques explain how behavioral change occurs from an organizational level. TTM will hence be used to explain the underlying behavioral change processes from an individual level.

As SPV's goal is for behavioral change to occur, we will measure the program's effectiveness as such. We will collect our data with surveys right before the program – Expectations, and in February – Experiences. This three-month gap is because behavioral

change takes time (Day et al., 2014; Prochaska & DiClemente, 1982). Furthermore, the amount of women in leadership positions within finance is growing. Gender will therefore be included as a variable to assess whether there are different perceived utilities between them.

As literature generally predicts behavioral change after a completed program, we predict that behavioral change will occur after the program. Furthermore, although the literature is inconclusive on whether mixed- or specific-gender approaches are most effective for a less represented gender, SPV believes their gender-neutral program to be effective. Hence, we predict that the program will have the same effect on both genders. As such, we present two hypotheses:

Hypothesis 1: *Behavioral change will occur at least three months after the management training program. Specifically, there will be a positive change in hunger and feedback skills, thereby making the program effective.*

Hypothesis 2: *The management training program will have equal effectiveness for both women and men.*

Method

Procedure

The data for this thesis was collected in collaboration with SPV through their semi-annual leadership development program. The program was located at SPV's headquarters in Bergen, totaling two gatherings. The first one was held on November 11th and 12th, and the second gathering was held on the 15th and 16th of November 2021. Day one of both gatherings lasted from 12:00 to 18:00 and ended with a social dinner. On day two, the training lasted from 9:00 to 14:00.

The participants were tasked with reading the literature by Ankersen (2016) and Heen & Stone (2014) before the program and gathering 360-degree feedback from their employees. On day one, the CEO gave an opening remark speech regarding hunger, their current results, and how they want to keep improving. In their first group assignment, the participants gathered group-wise. They discussed the results from the 360-degree feedback they had acquired from their co-workers. After a brief pause, the groups gathered and discussed in the plenum. Later in the program, the renowned consultant Suzanne Peterson had a workshop with the theme "Making others better through feedback" via Zoom. The workshop focused on improving skills through feedback.

On day two, the organizational developers presented the theme of the day. They then invited leaders from the gathering to present hunger in practice. After this, the participants received a group assignment where they conducted role-play scenarios. In the scenarios, the participants practiced how to give and receive constructive feedback as both leaders and co-workers. The role-play was then discussed within the groups and shared with the broader gathering. The day concluded with a group assignment on how to transfer the acquired knowledge into actual results in practice.

Data Collection

The T1 survey was administered digitally and consisted of preliminary demographic questions and 20 questions that assessed the participants' subjective level of *hunger & leadership* and *feedback*. The T2 survey consisted of the same preliminary demographic questions and 19 additional questions. The data was collected at two different points in time. The first collection, "T1", was sent out two weeks ahead of the program, on November 11th, to measure the expectations of the coming program. SPV conducted a second data collection to measure the immediate reactions to the program. This survey is not part of our data due to the replies not being anonymous. We were merely given the results from their analysis. The table showing the results is included in Appendix A. Lastly, "T2" was sent out in mid-February 2022, aiming to measure long-term behavioral change resulting from the program.

Participants

There were a total number of 86 attendees in the program. The total pool consisted of men and women aged from 20 to 51+, with years of leadership experience varying from under two to 11+ years. These participants were leaders belonging to the corporate management and managers in categories LU2 – two levels below the CEO, and LU3 – three levels below the CEO. SPV's "Aspiring leaders" initiative included a fourth level of participants in the program.

The participants differed in their respective departments and groupings. One such grouping is *Aspiring leaders*, consisting of regular employees admitted to the program and paired with a mentor to practice and hone their skills. Other departments include *Bulder*, *Business Markets* ("BM"), *Private Markets* ("PM"), and general *staff and support* under SPV centrally.

As we could not link respondent IDs from T1 to T2, our assessment was concerned with a group - rather than an individual perspective. We measured the participants at two time points. However, there may have been different respondents at different times: A respondent

ID at T1 may not have been the same person as the respondent ID in T2, e.g., ID # 3 at T1 was not necessarily the same person at T2.

Frequencies

Table 1

Baseline characteristics of respondents to the T1 - and T2 surveys.

	Respondents T1 (*n* = 64)		Respondents T2 (*n* = 45)	
Variables	n	%	n	%
Gender				
Male	41	64.1	29	64.4
Female	23	35.9	16	35.6
Age				
20-30	3	4.7	4	8.9
31-40	16	25.0	15	33.3
41-50	27	42.2	12	26.7
>51	18	28.1	14	31.1
Department				
BM	12	18.8	8	17.8
Bulder and others	8	12.5	3	6.7
Aspiring Leaders	3	4.7	6	13.3
Staff & Support	20	31.3	11	24.4
PM	21	32.8	17	37.8
Years of leadership experience				
<2	14	21.9	16	35.6
2-5	21	32.8	11	24.4
6-10	11	17.2	8	17.8
>11	18	28.1	10	22.2

T1. SPV had 86 participants in the overall program. Our anonymous T1 survey was answered by 64 participants, with a response rate of 74 %. Illustrated in Table 1, there were more men than women. The majority of the participants belonged to the age category *41-50*, while *PM* was the most represented department. Lastly, the participants also differed in their amount of leadership experience, with *2-3 years* of experience relating to the majority of the participants.

T2. Our anonymous T2 survey was answered by 45 participants, with a response rate of 52 %. Illustrated in Table 1, most of the participants were men. The majority of the participants belonged to the age category *31-40*, while *PM* was the most represented department. Lastly, the participants also differed in their amount of leadership experience, with *<2* years of experience relating to the majority of the participants.

Instruments

SPV administered the surveys in their digital, anonymous survey administration. The surveys measuring T1 and T2 were based on a survey by Bergheim et al. (2007) that evaluated female-oriented leadership development programs in the Norwegian Armed Forces. However, our questions were modified following the interests of the current thesis and the needs and interests of SPV. The surveys were administered in the employees' native language of Norwegian.

A five-point Likert scale was used in surveys T1 and T2 to capture the respondents' level of agreement or disagreement per question: 1 – "Strongly Disagree"; 2 – "Disagree"; 3 – "Neither agree nor disagree"; 4 – "Agree"; 5 – "Strongly Agree." All surveys are included in both Norwegian and English in Appendices B to G.

T1

This survey was split into three categories later computed into scales in SPSS. The scale *Expectations* was measured using five items, with questions such as "*I believe I am going to learn something at the gathering which I can use in my everyday life*." The scale *Hunger and leadership* was measured using seven items, such as "*I feel that my employees and I are hungrier for improvement than we were three months ago*." Lastly, *feedback* was measured using eight items, such as "*Sometimes I avoid tough feedback conversations. "*

Reactions. SPV administered a non-anonymous survey to the participants for their internal training evaluation. This survey consisted of three questions and aimed to capture the participants' immediate reactions. An example is "*To which degree will you change your behavior due to the gathering?*"

T2

This survey consisted of four categories and aimed to mirror the questions used in the T1 survey whenever possible. The survey assessed whether there had been an indication of behavioral change in the group. Otherwise, older questions were removed, and new ones were added.

The scale *Experiences* was measured using four questions, such as "*I have utilized what I have learned at the gathering in practice*." The scale *Hunger and leadership* was measured using four questions, such as "*I feel that my employees and I are hungrier for improvement than we were three months ago*." *Feedback* was measured using eight questions, such as "*Sometimes I avoid tough feedback conversations*."

Lastly, as Norway experienced new lockdowns in early December, employees were urged to work from home. Consequently, a fourth category, *home office*, was included to theorize whether this factor may or may not have influenced residual results in the T2 survey.

This category was measured using three questions such as "*The hunger has decreased while working from home*."

Design and Statistics

The current study utilized IBM SPSS Statistics (Version 27.0) for all conducted analyses. The data was provided as an excel-document and exported into SPSS. All analyses were measured from a group perspective.

Preliminary Data Screening and Data Preparation

We first screened the dataset and checked for errors by inspecting each variable's frequencies (see Table 1). We found the need to reverse items with negative wording for the measurements to go in the same direction. The item "*Sometimes I avoid tough feedback conversations"* was thus reversed.

Sum scores for the variables were further computed into their respective scales and adjusted by dividing sums by the number of items on the scale. This was done to ease the interpretation of the descriptive statistics (Pallant, 2020). A reliability test was performed to control for internal consistency of the scales used for *expectations/experience, hunger and leadership, feedback, and home office*.

Further, an exploratory analysis was conducted. This analysis was used to check whether the assumptions for the inferential statistics were met. By inferring the explore function in SPSS, we checked the normality rate by histograms for each scale and assessed potential outliers through the box plots. Tabachnick and Fidell (2014) highlight that a certain degree of skewness and kurtosis may occur in statistical analyses. The current study is not exempted from this, as all scales had a certain degree of non-normality.

Regarding outliers, some were found on different scales. However, these did not lead to an error in the dataset. The participants were also a natural part of the population. Therefore, we did not delete any of them.

Main analyses

Our main analyses were conducted as independent samples t-tests as we wanted to compare the means between two independent groups: timeframe (T1 and T2) and gender. We merged the files manually through Microsoft Excel and added the variable "Timeframe" with values 1 and 2. Our first main analysis was a simple independent samples t-test with *Hunger and Leadership*, *Feedback*, and *Sometimes I avoid tough feedback conversations* as scales. Some items were not mirrored from T1 to T2, so they were cut from the main analysis. The following items were removed from the main analysis: "*I feel I have the competence I need to motivate my employees to strive for improvement*," "*I encourage my employees to challenge the status quo,*" and "*I have had specific conversations regarding hunger and development with my employees the last month*." *Feedback* remained untouched. We used the variable *timeframe* as the grouping variable.

To assess whether the scale *Home Office* impacted the results, we ran a non-parametric correlation in the T2 dataset. Lastly, we ran an independent samples t-test to assess the perceived effectiveness for the genders. This was done in the merged dataset, first with each gender filtered away per analysis and then with timeframe filtered away per analysis.

Ethics

All surveys were administered by SPV using an anonymous digital survey solution that does not collect IP addresses. Thus, the replies could not be linked to the individuals. All questions were made in large answer groups to avoid the possibility of identification through the combination of answers. No personally identifiable information was collected. All respondents were informed that answering the survey was voluntary, and no personally identifiable information was collected.

SPV conducted its own survey-based evaluation of the participants' reactions ("RT") as a part of their own evaluation of the leadership training. This survey was not anonymous. Only the gender distribution, means, and standard deviations were shared, not the underlying

data. To prevent personal identification, it was not possible to establish links between the respondents in T1, RT, and T2.

Results

Preliminary Analyses

Reliability Analyses

SPSS' reliability function was used to calculate each of the scales' Chronbach's α. According to DeVellis and Thorpe (2021), a score of .7 or above indicates an acceptable internal consistency. All levels of Cronbach's α are included in the table in Appendix I.

T1. The question "*I will learn things at the leadership development program that I would not have learned by regular practice as a leader*" was removed from T1 as it had to be removed from T2. *Expectations* thus showed a Cronbach's α of .818, indicating internal consistency. *Hunger and leadership* indicated acceptable internal consistency with a Cronbach's α of .704.

Feedback initially showed a low Cronbach's α of .694. Following the reliability analysis, the item "*This past week, I have requested informal feedback from my co-workers regarding a specific area*" was removed from the original scale. " *Sometimes I avoid tough feedback conversations"* was also removed but kept as a sole item due to being a predictor for behavioral change. When removed, the relevant scale indicated acceptable internal consistency with Cronbach's α at .769.

T2. The first scale, *Experiences*, initially showed a Cronbach's α of .575, indicating little internal consistency. The reliability analysis indicated that Cronbach's α would rise to .749 if the item "*I could not have learned what I learned merely by merely practicing it*" was removed. On the other hand, *Hunger and Leadership* showed a Cronbach's α of .761, indicating acceptable internal consistency. The scale *Feedback* also indicated internal consistency. The same items were removed as in T1. The Cronbach's α was .708. Lastly, the

scale *home office* initially showed a Cronbach's α of .264, indicating a lack of internal consistency. Following the reliability analysis, the item "*I have been able to use what I have learned at the gathering when working from home*" was removed, resulting in a Cronbach's α of .750.

Main Analysis

Measuring the overall effectiveness of the program

Table 2

Items in T1 and T2 with means, SD, Cronbach's α, and differences between T2 and T1.

	T1			T2			Difference	
Variable	*M*	*SD*	α	*M*	*SD*	α	*MD*	*t*
Hunger and Leadership	3.80	.566	.704	3.74	.634	.761	.056	.487
Feedback	4.20	.505	.769	4.20	.455	.708	-.001	-.010
Avoid Difficult Feedback	2.98	1.266	-	3.33	1.158	-	-.349	-1.472

We ran an independent samples t-test to compare the scores from the 64 participants of T1 to the 45 participants of T2 (Table 2). For *Hunger and Leadership*, we saw a slight non-significant decrease in scores (t = .487, p = .627) from T1 to T2. The scale *Feedback* indicated no decrease in the mean score, though a slight decrease in standard deviation (t = -.010, p = .992). The last item we were interested in was "*Sometimes I avoid tough feedback conversations*." There was a non-significant increase in means and a decrease in the standard deviation score (t = -1.472, p = .144).

Home office as a possible influence. To assess whether the scale *Home Office* impacted the residual scales in T2, we conducted a Spearman correlation analysis of these variables. As our *N* is low with 86 total participants, these correlations were non-parametric.

Table 3

Non-parametric Spearman's correlation of Home Office and residual variables in T2

Variable	Experiences	Hunger/Leadership	Feedback	Avoid Difficult Feedback
Home Office	.102	.333*	.028	.191

Note: *= Correlation is significant at the .05 level (2-tailed).

Table 3 shows the correlations. The variables *Experience, Feedback,* and *Avoiding Difficult Feedback* all had very low correlations with *Home Office* and a lack of significance. On the other hand, the variable *Hunger and Leadership* had a low correlation, although significant at .333.

Measuring the perceived effectiveness for men versus women

Table 4

Means and standard deviations between men and women from T1 to T2

	Men						Women					
	T1 (n = 41)		T2 (n = 29)		Difference		T1 (n = 23)		T2 (n = 16)		Difference	
Variables	*M*	*SD*	*M*	*SD*	*MD*	*t*	*M*	*SD*	*M*	*SD*	*MD*	*t*
Hunger and	3.88	.507	3.81	.700	.068	.470	3.66	.647	3.63	.492	.038	.198
Feedback	4.25	.489	4.17	.476	.085	.728	4.09	.529	4.25	.426	-.156	-.977
Avoid Difficult	3.22	1.215	3.62	1.015	-.401	-1.455	2.57	1.273	2.81	1.223	-2.47	-.606
Feedback*												

Note: *= Difference is significant at the .05-level (2-tailed) between men and women before the program.

To assess the perceived effectiveness of the program for men versus women, we conducted yet another independent samples t-test. Table 4 shows the means and standard deviations for both genders in T1 and T2. There was a non-significant decrease in the mean score for men and an increase in standard deviation for *Hunger and Leadership* (t =.470, p = .640). Regarding *Feedback* (t = .728, p = .469), there was a non-significant decrease in the mean score and a decrease in standard deviation, indicating less variance in their replies. Lastly, regarding "*Sometimes I avoid tough feedback conversations*" (t = -1.455, p = .150), there was a non-significant increase in the mean score and a decrease in standard deviation.

For women, there was a non-significant decrease in the mean score and standard deviation of *Hunger and Leadership* (t = .198, p = .844). Regarding *Feedback* (t = -.977, p = .335) and "*Sometimes I avoid tough feedback conversations*" (t = -.606, p = .548), there was a non-significant increase in the mean scores and a decrease in standard deviation. However, the latter variable indicated a significant difference between men and women at T1 (p = .046) and T2 (p = .022), which indicates that women avoid giving difficult feedback to a greater degree than men.

Discussion

The current thesis aimed to measure the effectiveness of the leadership development program administered by SPV. Our intent to evaluate the program was twofold: to assess whether the organization needed to change its future programs and highlight the need for systematic evaluations of programs. The evaluation was conducted with quantitative analysis to measure the overall group improvement and the perceived effectiveness for men and women in the population. Hypothesis 1, which was that behavioral change would occur three months after the program, was not supported. The findings showed no significant increase in the sum scales of *Hunger and Leadership*, *Feedback*, and *Sometimes I avoid tough feedback conversations*. Table 2 also indicates an already-high baseline among the participants in T1.

Men and women had equal effectiveness in the program, supporting hypothesis 2. Table 3 indicates no overall significant difference between men and women. When filtering out timeframe, our results indicate a significant difference between men and women in *Sometimes I avoid tough feedback conversations* in both timeframes. However, the results indicate no significant difference in the perceived effectiveness for men and women.

Behavioral Change

Hypothesis 1: Behavioral change will occur at least three months after the management training program. Specifically, there will be a positive change in hunger and feedback skills, thereby making the program effective.

The results did not support H1 as the sum scores for the three variables did not show a significant increase after the program. Table 2 showed an already high baseline in the participants' mean scores in T1. This indicates that the participants reported a very high level of "hunger" and feedback skills before the program, which did not increase after the training. Table 2 shows a slight decrease in the mean score in T2, which could be explained by random variation caused by unknown factors. After completing the program, the respondents did not report better feedback skills than before. The participants reported that they had gotten better

at not avoiding tough feedback conversations after completing the program. However, our analysis showed no significant increase in said skills. In the T2 measurement, there was a significant correlation between the home-office variable and hunger & leadership. This could indicate that the participants who rated themselves high in hunger & leadership found it more difficult working from home (see table 3).

Previous research has shown that leadership development programs had mixed results (Avolio et al., 2009; Powell & Yalcin, 2010). While our results did not show an effect, this is not conclusive proof that the leadership program is ineffective. As mentioned in the investigated literature, behavioral change is the holy grail of leadership development (Discovery in Action, 2020). Even if a leader learns the skills and abilities after a program, there is no guarantee that the leader will utilize those skills. Out of Kirkpatrick´s(1979) four steps, *reaction, learning, behavioral change,* and *results*, behavioral change is the most challenging step.

There may be multiple reasons why the program did not show the expected results. Our results could be explained by other factors that might have influenced our process. The factors discussed further are the ceiling effect, the use of a needs assessment, and the home office as an effect on our results.

The ceiling effect as a possible explanation

SPV's goal was to induce behavioral change amongst the participants. However, the attendees were all mid-to high-level leaders in SPV's hierarchy. Most of them had already attended similar programs on behalf of SPV. Our results showed high mean scores in *Hunger & Leadership* and *Feedback* before the program (See Tables 2 and 4). These high results could indicate that there had been a ceiling effect from the start. A ceiling effect is a measurement limitation. This occurs when the highest possible score or close to the highest score on a measurement instrument is reached, further decreasing the accuracy of the intended measurements (Salkind, 2010).

One explanation for why this ceiling effect occurred in our results may have something to do with how the questions in the survey were formulated. The questions may have been too easy, such that the participants scored high mean scores. How the facilitators present and formulate the questions in a survey does influence the received data. In our case, we asked questions such as "I am good at giving feedback." Our questions may have been too simple for an educated mass or might not have captured the full extent of what they were supposed to measure. The phrasing of general questions such as the one mentioned may be an explanatory factor in why our results indicated no significant change. Future studies should ask more specific questions to capture what they need. One could ask more precise questions linked to specific leadership behavior. For example, "As a leader, I seek feedback to improve interactions with others."

Another explanation, following the above, might be that the participants already before the program were highly competent in hunger and feedback skills. This indicates that there was little room for significant improvement. As mentioned, the participants were either high- or middle-level leaders with several years of leadership experience. One could assume that their high level of experience has led to entrenched behaviors that are more difficult to

change. For example, a high-level leader who has used the same feedback methods over the last ten years might struggle to let go of those techniques even after a training program. At the same time, a low-level leader with less experience may be more willing to try out new skills and transfer them to behavior. With more experience, it may be more challenging to transfer the learning into behavior. This may be called a "transfer problem," which refers to not being able to generalize learned material to the job and the maintenance of trained skills over a period (Baldwin & Ford, 1988).

This assumption is supported by Lacerenza et al. (2017), who found that leaders across multiple levels were equally motivated to learn and likely to improve organizational - and subordinate outcomes. However, their results differed in Kirkpatrick´s step of transferring learning to behavioral change. That is, applying what they had learned from the development programs to the actual job was more difficult for high-experienced leaders than low-experienced leaders. The transfer effect was circa four times weaker for leaders high up in the hierarchy. In practice, this could mean that the leaders with a high level of experience would have a harder time implementing what they have learned and thus require more training in this area. Lacerenza et al. (2017) argue that these upper leaders experienced the ceiling effect only in transferring what they had learned into behavior and not the other levels in Kirkpatrick's framework. This may further indicate that upper-level leaders' high level of experience may have led to entrenched behaviors that are difficult to change on the job.

These findings do not mean that upper-level leaders do not have room for improvement but instead have a "transfer problem" (Baldwin & Ford, 1988). Based on the findings of Lacarenza et al. (2017), it may be advisable that SPV focuses on training and following up their leaders after the intervention to enhance the transfer of behavior if that is the case.

Maintaining rather than developing behaviors

Training on something one is already good at may seem pointless, but it can also be important to maintain strong skills if one wants to stay on top. In short, maintenance is the same as training on a specific behavior to prevent relapse. Just as top athletes train to maintain their strengths and skills to stay on top, so should leaders. Therefore, the question is whether SPV's understanding of leadership development should be to develop new behaviors and skills or to include maintenance in this understanding.

In the circular process of TTM, behavioral change is a long and arduous process for the individual. As mentioned in the theoretical framework, TTM views behavioral change as a six-step process - precontemplation, contemplation, preparation/determination, action, maintenance/relapse, and termination. The ceiling effect that we discussed might call for this maintenance or training rather than inducing new behaviors.

Maintaining has a single yet challenging goal: To avoid relapsing into previous behaviors. Since these leaders already had a high baseline, as indicated in Tables 2 and 4, SPV should focus on maintaining these behaviors as there is not much to improve. In fact, it became clear during SPV's gathering that they wanted to stay on top of their game. If this is the case, maintenance should be prioritized rather than development – especially when "developing" the same leaders each year.

However, maintenance is not without its troubles. Firstly, motivation and incentives will always be relevant factors in behavioral change. As observational learning may significantly influence values, behaviors, and attitudes (Bandura, 1986; Sims & Manz, 1982), we suggest that SPV facilitates a supportive, incentivizing climate where models can learn from each other. This could lead to a greater ability to contextualize what the participants have learned by themselves. This could also provide social cues on how to act, possibly hindering a decrease in the behaviors while making it easier to maintain them.

Secondly, it would be interesting to measure whether a leadership development program has successfully maintained skills that would otherwise have been reduced. This would mean that one group would have received the originally intended training, and the other would receive training in the maintenance of behavior. After that, one should compare the results of these two groups to assess which approach has the desired results. However, even though maintaining strong skills is relevant, investing in areas with more significant potential for change may be more reasonable. It may even yield a higher return on investment.

A new needs assessment

Building on what was discussed above, one may assume that the interventions in the program might have focused on the wrong areas for improvement. As such, we suggest that SPV conducts a new needs assessment. Developing a program from a needs assessment may yield a greater degree of learning and transfer, thereby reducing the chance of generic training programs (Lacerenza et al., 2017). This could also help SPV with focusing on the right areas.

Had SPV conducted a thorough needs assessment from the start, they could have discovered the ceiling effect and tweaked their program early on. This tendency to neglect is not a new one. Some organizations even view them as a waste of resources or believe they will not reveal any new information (Goldstein & Ford, 2002). Martin (2009, as cited in O'Loughlin, 2013) even highlights that needs assessments may identify the participants' current learning level and performance gap. The practical implications of SPV conducting thorough needs assessments in the future may be that i) it will be easier to allocate resources for future programs, ii) it could be easier to identify which areas the leaders need training on, and iii) it could be easier to determine the leaders' current level of learning, making it easier to hit the spot.

As needs assessments may pinpoint the participants' current level of learning, they may also help the facilitators. As discussed in the section about the ceiling effect, our

questions might have been too simple. If SPV had conducted a needs assessment beforehand, it would have been easier for us to pinpoint the participants' current level of learning and thus provided clues on how to formulate our questions.

Home office as a possible explanation

Table 3 shows a significant correlation between the experience of *home office* and *hunger & leadership*. The correlation tells us that those who scored high on hunger & leadership also scored high on home-office. This alone points to the fact that working from home may have taken its toll on some participants or that they experienced it as an inconvenience. The result implies that people who are hungrier and ready to lead might have had a more challenging time. In fact, leading virtually and from home is more demanding than in person. Leaders are primarily trained in face-to-face leadership, which requires different skills than virtual leadership. This was also the case for SPV's program. The idea of virtual teams and – leadership is a concept that has existed for a long time, though this trend has accelerated as a result of the covid-pandemic (Hulbach & Refsnes, 2021).

Apart from the stress associated with a looming pandemic, a lockdown makes it harder to work proactively with development. Most leadership development programs develop face-to-face skills that require a degree of socialization. Working from home requires broader technical support and new ways to follow up with the employees. In this sense, socialization methods such as social modeling will also be more challenging. The employees will no longer be able to fully socialize with their peers or leaders as intended. For example, cues from body language will be harder to interpret. In this sense, it would be harder to model the behavior of leaders who score high in, e.g., hunger & leadership, as a computer screen mediates the socialization.

Lockdowns have not only challenges for the individuals but also for the organization. The available time and capacity to finish projects might drastically decrease and be put on

hold. For example, following up on the leaders' change might be harder as new challenges present themselves in lockdowns. Therefore, this lack of a follow-up might not motivate the participants to further practice their skills, which is already hard enough when leading virtually.

The effect of virtual teams and working from home is still not thoroughly known. Multiple variables may affect the performance of virtual teams, indicating that virtual team interactions fluctuate from day to day (Hulbach & Refsnes, 2021). SPV's program did not factor in the possibility of mandatory home offices as something that would return. There was a general belief in Norway that no further lockdowns would commence. The Norwegian workforce had already been mandated to work from home for over two years, which takes its toll. When Norwegian society opened again in September 2021, no one expected another lockdown three months later. It is reasonable to assume that they felt it was hard to be a good leader when working from home, or that drive and innovation, categorized as hunger, had decreased.

Gender

H2: The management training program will have equal effectiveness for both women and men.

The results supported H2, which was formulated as a null hypothesis. There is no indication that the program's effectiveness was significantly different when looking at gender in our data. As a group, the participants are alike. There is nothing significant in our data that indicates that women have worse or better effects of the program than men. This implies that SPV should continue to conduct mixed-gender leadership development programs.

As shown in Table 4, there was no significant rise or fall in the participants' hunger and feedback skills. Although there were some discrepancies between men and women, they were not significant. The largest discrepancy, which was significant, was between men and

women in T1 and T2 regarding how they handle tough feedback conversations. This significant difference was already present before the program and still existed in February. However, there was no overall difference in the perceived efficiency of the program, as this significant difference was not a result of the program. We thus find support for our hypothesis. Our results further support other studies that show equal effectiveness between men and women in mixed-gender leadership development programs (Burke & Day, 1986; Ely et al., 2011).

Both women and men had a high baseline level of *hunger & leadership* and *feedback skills* with no significant differences before or after the program. These results may be explained through a selection effect. In practice, this would mean that the leaders who have been selected share specific competencies and skills they want in a leader at SPV. This implies that SPV focuses on competence and attributes that can be developed when choosing leaders. An implication of this may be that they select men and women who are more similar to each other than what men and women may be in the population. Hence, the differences based on gender becomes less clear. This could further support their gender-neutral focus. However, another implication of this is that the gender differences may be greater in less selected samples.

Even though men and women experienced equal effectiveness, our data showed a significant discrepancy between men and women in how they handle tough feedback conversations. According to our results, women avoid more tough feedback conversations than men. This discrepancy did not change after the program. The significant discrepancy could be explained by differences in self-ratings and self-confidence between men and women. Men tend to rate themselves significantly more effectively than women do in senior leader positions (Paustian-Underdahl et al., 2014). Furthermore, men tend to overestimate or rate themselves accurately, whereas women tend to either underestimate or rate themselves

accurately (Beyer, 1990). To control for this, one could use 360-degree feedback surveys across multiple timeframes to get more objective measures for male and female leaders. This could make it easier to avoid such sources of error.

Moreover, research shows consistent gender differences in self-confidence levels, with women generally having less confidence than men (Instone et al., 1983). Hence, it is possible that female leaders may avoid more tough feedback conversations because of their lower confidence. To investigate this, one could have conducted in-depth interviews with the female leaders on the topic of feedback conversations to better understand why women tend to avoid tough feedback conversations to a greater degree than men.

In summary, the results suggest SPV´s leadership development program has equal effectiveness for both men and women, further supporting their gender-neutral focus as one of their equality measures. As there were differences in how they handle tough feedback conversations before and after the program, we suggest that SPV conduct 360-degree feedback and in-depth interviews to better measure and understand why these differences exist.

Limitations

Time

The current study is not without its limitations. Our results indicated that there was no significant increase in behavioral change. As stated in the theoretical framework, Prochaska et al. (2015) claim that behavioral change may occur after at least six months. It is, therefore, important that there is sufficient time between measurements. The current evaluation took place three months after the program, which might not have been enough time to conduct a thorough evaluation. However, some leadership development programs may show a larger effect at the start, only for this effect to fall with time. This is in stark contrast to our initial assumption but might be due to various outside factors (Lacarenza et al., 2017). However, a key factor for this potential fall of effect is whether the program only has one activity or if the

training is systematically tailored and lasts over time. Nevertheless, having more time could present opportunities such as more than two measurements.

Shamir (2011) points out that most evaluations collect data from two time points at best but that multiple measures would benefit many studies. In practice, more than two measurements would have allowed us to assess a possible effect over time. For example, we would be able to measure the effect not only after three months but also after six or eight months. Had this been possible, the study could also have been truly longitudinal. This means that we would have been able to measure the same participants over a long period, e.g., a year. Although this study is somewhat longitudinal, a more extended measurement period could have helped us identify long-term behavioral patterns. On the other hand, evaluating the same people multiple times might decrease scores over time. This might be due to a lack of motivation or fatigue as the evaluation drags on.

A lack of time plus the organization's needs also hindered us from i) using a qualitative section at the end of the survey, ii) conducting qualitative interviews next to the quantitative analyses, and iii) going in-depth in Kirkpatrick's steps. We believe that the two former points would have let us understand the nuances not covered by the questions in the survey. The behaviors that we measured are hard to measure objectively, especially when the data is derived from self-report surveys. Future practical studies and consultants should thus address time as an important factor when conducting an evaluation.

Methodological limitations

First and foremost, our sample size was representative of the population at SPV. The population consisted of 86 people, with a high response rate of 74 % in T1 and 52 % in T2. Respectively, the sample consisted of 64 and 45 individuals. Generally speaking; however, the population is small. This means that we also have a small sample size in general terms, which might present an issue. Statistical significance is very dependent on the *n* when dealing with small sample sizes. This means we would require much larger mean differences to get significant values. The largest implication of the low *n* is that our small sample size leads to a lack of statistical power. This would mean that there might have been an effect that our analysis could not find. However, it should be noted that our small level of mean differences reduces the likelihood of our power being the issue.

A secondary challenge with a small sample size could be that outliers would have more influence in our analysis. This would mean that respondents who have rated themselves very high or very low would have influenced the mean scores and standard deviations to a greater degree than normalized samples. Although sample – and population sizes depend on the study, these methodological limitations should be accounted for.

Furthermore, Kirkpatrick recommends the use of control groups when evaluating learning. The use of control groups would have allowed us to assess two things: i) whether learning has or has not taken place in a group, and ii) to investigate whether behavioral change is dependent on learning or not in that particular setting. Due to available time, resources, and the number of participants, this was not possible.

Additionally, we cannot establish respondent IDs at T1 and T2, making it more challenging to evaluate individual behavioral change. We would have combined this with multiple measurements to evaluate both individual – and group change if we could. This would have allowed us to assess group trends and individual learning curves.

Lastly, had we been able to administer a 360-degree feedback survey, we believe our results might have had a stronger impact and ultimately benefited the organizations. In this case, we would have used the 360-degree surveys to ask the coworkers about specific areas such as "The manager has spoken to me about hunger the last couple of weeks." We would have combined these results with the self-reporting surveys. These surveys would have used general questions such as "The hunger has increased lately."

Contributions and Implications

Although we have our limitations, our thesis also offers some contributions. Firstly, most articles regarding leadership development and Kirkpatrick's four steps are primarily theoretical. A quick search for leadership development on Google Scholar shows many articles, though most are literature reviews, meta-analyses, and other theoretical contributions. Therefore, it was challenging for us to find data on specific, practical examples of leadership development programs in scientific journals. This was also the case for Kirkpatrick's model. Although we did find some empirical evidence of how the model functioned, we could not find practical examples of how each of the steps had been evaluated. This includes the University of Bergen's "Bergen Open Research Archive," where we only managed to find one practical evaluation from 20 years ago.

The few practical evaluations and reports we found were mostly limited to the Norwegian Armed Forces ("Forsvaret"). However, generalizing these results to the broader Norwegian workforce might present some challenges. First and foremost, Forsvaret is a military organization rather than a civilian one like SPV. Their selection process emphasizes physical, as well as mental, and intellectual attributes. Furthermore, the culture and understanding of leadership is different in Forsvaret than in civilian organization:

"*A military leader must be prepared to put forward their own life on behalf of the state and political goals, take others' lives, and give assignments that can put subordinates' and*

others' lives in danger … The core of military leadership is linked to planning and conducting military operations" (Forsvaret, 2020, pp. 6, 8).

This paraphrasing indicates that their understanding of leadership is operational. As such, military organizations have different needs than civilian organizations. On the other hand, our study was conducted in a Norwegian civilian organization and will therefore be easier to generalize. Our study is realistic. The more practical and realistic the study is, the less purely experimental it is. It, therefore, becomes more valid. However, the issue is that third variables could influence the results, such as with our study.

Furthermore, there is a lack of empirical quantitative studies of gender differences in leadership development in Norwegian contexts. Many studies are qualitative or have mixed methods, emphasizing qualitative measurements. Much of the available quantitative data are in descriptive reports that do not present theory, such as the data provided by CORE in our introduction. However, the current thesis provides quantitative empirical data on this theme.

Our thesis also has direct implications for SPV. The organization aimed to evaluate its leadership development program to assess i) its general effectiveness and ii) its utility for men and women. The current thesis draws results directly from their leadership population and has given SPV an insight into how the respondents experienced the program. Although our results may have been affected by outside factors, there is a tradeoff with this study being realistic, which makes it more generalizable. We have also offered explanations of these and what SPV should aim to accomplish in future programs. Our results have further indicated that there is little difference in the perceived effectiveness of the program between men and women. Although there was a significant difference between men and women in avoiding difficult feedback in both T1 and T2, there was no perceived difference in the program's efficiency, as seen in Table 2. This might, in turn, bolster the idea of gender equity in SPV.

Future research

Our results show little change in behavior from T1 to T2. This might be due to factors such as the ceiling effect, wrong focus area, a lack of needs assessments, or lack of time. Future research should consider these factors and our limitations when conducting similar analyses.

As mentioned, time is important when evaluating leadership development programs. Although longitudinal studies have their drawbacks, they have their benefits as well. If we had more time, we would do more than two measurements to capture the effect over a longer period. We recommend that future researchers do this as behavioral change does not happen overnight.

We further recommend that future researchers send out a survey two weeks beforehand to measure the participants' baseline levels, followed by a reactions survey immediately after the program, which should include a qualitative section. Furthermore, we recommend that new surveys be sent out three, six, and, e.g., nine months later. These surveys should be 360-degree surveys to measure specific areas relating to behavior and self-reporting surveys that measure general areas relating to learning.

We would also have followed more gatherings over time and measured this development. This might have allowed us to establish one of the conditions as a control group. By establishing at least one condition as a control group, it would be easier to measure whether learning occurs. In practice, this can offer two possibilities: i) to predict behavioral change, or ii) to establish one group as a "maintenance" group. The latter can help future researchers to identify whether maintenance would benefit the organization. However, this was not possible for us due to our low *N*. The number of participants and sample size depend on factors beyond the consultant's control. However, following more gatherings over time could solve this problem.

Furthermore, more time would have allowed us to conduct a thorough needs assessment beforehand. A qualitative section in the pre-survey could also provide valuable information that would benefit this needs assessment. We believe that this insight might have caught a potential ceiling effect early on and allowed SPV to focus on, e.g., maintenance of behavior. Future researchers should thus have this in mind.

Besides the points stated above, a lack of time and mandate hindered us from seeing the program's long-term effects. Kirkpatrick's last step, *results*, is a form of a long-term effect and would shed light on a return on investment. Examples of such long-term organizational changes could be decreased turnover, higher revenue, increased employee feedback scores, or other results that the organization strives for. Multiple measures post-program, as mentioned above, could provide some insight into these results.

We further recommend that future research utilizes a mixed-methods approach. This means quantitative data collection and analyses and a qualitative section in the survey. In this qualitative section, the participant could discuss themes not covered by the surveys. This will also make it easier to tweak future surveys or could benefit a needs assessment, as stated above. These surveys should not be entirely anonymous but rather confidential. By this, we mean that only the researchers will know the participants' IDs. In practice, this would allow the researchers to link the respondent IDs in each of the timeframes and allow them to assess individual changes in addition to group change. We further believe that qualitative interviews of the minority gender could paint a more nuanced picture of their position in the organization. In practice, this could yield results that provide a holistic perspective on how they feel about the program, their peers in general, and whether the program should utilize a gender-specific approach or continue with a mixed-gender approach.

Conclusion

This study has evaluated SPV's semi-annual leadership development program. We based our evaluation on Kirkpatrick's techniques: measuring *reaction, learning, behavioral change,* and *results*. We further based our further understanding of behavioral change on Social Cognitive Theory and the Transtheoretical Model of Behavioral Change. The latter two theories offer an individual explanation rather than an organizational one, as Kirkpatrick's techniques do.

We had two points of data collection: A pre-program survey sent out in November and a post-program survey sent out in February. These surveys consisted of demographic questions, questions relating to their expectations or experiences, and SPV's focal points: hunger and leadership, and feedback. These surveys aimed to provide data that could indicate a behavioral change three months after the program.

Our analyses derived from two research questions: i) whether the program was overall effective in changing the participants' behaviors, and ii) whether the program had equal effectiveness for men and women. To test the hypotheses, we conducted various t-tests to measure the means and standard deviations of the respondents and see whether there was a significant difference between the groups. Our results in Table 2 indicated no significant increase or decrease among the participants before and after the program. As we operated with a small sample, we did not have enough statistical power to conclude that the program was ineffective. Instead, we argue that other factors might have been at hand.

For example, we noticed a high baseline among the participants in the pre-program measurement. This indicates a ceiling effect, which begs to question of whether it would be realistic to improve their behaviors. This would, in turn, call for the maintenance of behaviors rather than developing new behaviors. In TTM, behavioral change is understood as a circular process. In this process, maintenance is the process of training and practicing these behaviors

so that they stick. We thus recommend that SPV aims to maintain existing behaviors by training the participants rather than inducing new ones. This would, in turn, call for new needs assessments and a change of strategy. We also noticed a significant correlation between home office and hunger & leadership. We believe that the participants were unable to practice their skills face to face during the lockdown in early December and that this might have affected our main analysis. These factors are ones that SPV should focus on in the future.

Furthermore, our results supported our second hypothesis. Table 4 indicates that there was no significant difference in the effectiveness for men and women. We believe that this is due to the selection effect, meaning that SPV hires specific types of people based on intrinsic factors. That is, external differences become less prominent as people are relatively alike in their skillsets or personality types. However, we found a significant difference between them in how they handle tough feedback conversations. However, this significant difference was already present before the program and thus did not affect our results.

Our study has both its limitations and contributions. Time may be the most important limitation in our study, next to the population size. However, these should not be taken as a defeat but rather as a way to improve future research. We further contribute with research on practical evaluations. Most evaluations are theoretical, while practical evaluations are delimited to the Norwegian Armed Forces. However, results from military organizations cannot be transferred to civilian ones as they have vastly different cultures, selection methods, and fundamental understanding of the term leadership.

References

Alliger, G. M., & Janak, E. A. (1989). Kirkpatrick's levels of training criteria: Thirty years later. *Personnel psychology, 42*(2), 331-342. https://doi.org/10.1111/j.1744-6570.1989.tb00661.x

Amagoh, F. (2009). Leadership development and leadership effectiveness. *Management decision, 47*(6), 989-999. https://doi.org/10.1108/00251740910966695

Andersen, S. M., & Cole, S. W. (1990). " Do I know you?": The Role of Significant Others in General Social Perception. *Journal of Personality and Social Psychology, 59*(3), 384-389. https://doi.org/10.1037/0022-3514.59.3.384

Ankersen, R. (2016). *HUNGER in PARADISE*. Turbulenz Publishing

Ardent Learning. (2020, February 19th). *What is the Kirkpatrick Model? Learn the 4 Levels of Evaluation*. Ardentlearning.com. Retrieved October 18th from https://www.ardentlearning.com/blog/what-is-the-kirkpatrick-model#:~:text=The%20Kirkpatrick%20Model%20is%20a,learning%2C%20behavior%2C%20and%20results.

Arthur Jr, W., Bennett Jr, W., Edens, P. S., & Bell, S. T. (2003). Effectiveness of training in organizations: A meta-analysis of design and evaluation features. *Journal of Applied Psychology*, *88*(2), 234. https://doi.org/10.1037/0021-9010.88.2.234

Avolio, B. J., Reichard, R. J., Hannah, S. T., Walumbwa, F. O., & Chan, A. (2009). A meta-analytic review of leadership impact research: Experimental and quasi-experimental studies. *The Leadership Quarterly, 20*(5), 764-784. https://doi.org/10.1016/j.leaqua.2009.06.006

Backhaus, K., & Tikoo, S. (2004). Conceptualizing and researching employer branding. *Career development international, 9*(5), 501-517. https://doi.org/10.1108/13620430410550754

Baldwin, T. T., & Ford, J. K. (1988). Transfer of training: A review and directions for future research. *Personnel psychology, 41*(1), 63-105. https://doi.org/10.1111/j.1744-6570.1988.tb00632.x

Bandura, A. (1969). *Principles of Behavior Modification*. Holt, Rinehart & Winston.

Bandura, A. (1977a). Self-efficacy: toward a unifying theory of behavioral change. *Psychological Review, 84*(2), 191-215.

Bandura, A. (1977b). *Social Learning Theory*. Prentice Hall.

Bandura, A. (1986). *Social Foundations of Thought and Action: A Social Cognitive Theory*. Prentice Hall.

Bandura, A. (2001). Social Cognitive Theory: An Agentic Perspective. *Annual review of psychology, 52*(1), 1-26. https://doi.org/10.1146/annurev.psych.52.1.1

Beyer, S. (1990). Gender differences in the accuracy of self-evaluations of performance. *Journal of Personality and Social Psychology, 59*(5), 960-970. https://doi.org/10.1037/0022-3514.59.5.960

Blume, B. D., Ford, J. K., Baldwin, T. T., & Huang, J. L. (2010). Transfer of training: A meta-analytic review. *Journal of management, 36*(4), 1065-1105. https://doi.org/10.1177/0149206309352880

Bouty, I. (2000). Interpersonal and interaction influences on informal resource exchanges between R&D researchers across organizational boundaries. *Academy of Management journal, 43*(1), 50-65. https://doi.org/10.5465/1556385

Burke, M. J., & Day, R. R. (1986). A cumulative study of the effectiveness of managerial training. *Journal of Applied Psychology, 71*(2), 232-245. https://doi.org/10.1037/0021-9010.71.2.232

Cardy, R. L., & Lengnick-Hall, M. L. (2011, 2011/06/01). Will They Stay or Will They Go? Exploring a Customer-Oriented Approach To Employee Retention. *Journal of Business and Psychology, 26*(2), 213-217. https://doi.org/10.1007/s10869-011-9223-8

Carlyle, T. (1841). *On Heroes, Hero-Worship, and the Heroic in History* James Fraser. https://books.google.dk/books?id=HJwQAAAAYAAJ&hl

Collins, D. B., & Holton III, E. F. (2004). The effectiveness of managerial leadership development programs: A meta-analysis of studies from 1982 to 2001. *Human Resource Development Quarterly, 15*(2), 217-248. https://doi.org/10.1002/hrdq.1099

CORE - Centre for Research on Gender Equality. (2021, May 25th). *Norwegian Gender Balance Scorecard*. Institute for Social Research. Retrieved September 27th from https://www.samfunnsforskning.no/core/publikasjoner/core-indikator-status/core-topplederbarometer/core-topplederbarometer-2020/flak/core-topplederbarometer-2020.pdf

Crawford, J. A., & Kelder, J.-A. (2019). Do we measure leadership effectively? Articulating and evaluating scale development psychometrics for best practice. *The Leadership Quarterly, 30*(1), 133-144. https://doi.org/10.1016/j.leaqua.2018.07.001

Cummings, T. G., Worley, C. G., & Donovan, P. (2015a). Chapter 5: Diagnosing. In T. G. Cummings, C. G. Worley, & P. Donovan (Eds.), *Organization Development and Change* (10e ed., pp. 89-122). Thomson/South-Western.

Cummings, T. G., Worley, C. G., & Donovan, P. (2015b). Chapter 16: Talent Management. In T. G. Cummings, C. G. Worley, & P. Donovan (Eds.), *Organization Development and Change* (10e ed., pp. 474-493). Thomson/South-Western

Dalakoura, A. (2010). Differentiating leader and leadership development: A collective framework for leadership development. *Journal of Management Development, 29*(5), 432-441. https://doi.org/10.1108/02621711011039204

Day, D. V. (2000). Leadership development: A review in context. *The Leadership Quarterly, 11*(4), 581-613. https://doi.org/10.1016/S1048-9843(00)00061-8

Day, D. V., Fleenor, J. W., Atwater, L. E., Sturm, R. E., & McKee, R. A. (2014, 2014/02/01/). Advances in leader and leadership development: A review of 25years of research and theory. *The Leadership Quarterly, 25*(1), 63-82. https://doi.org/10.1016/j.leaqua.2013.11.004

DeVellis, R. F., & Thorpe, C. T. (2021). *Scale development: Theory and applications*. Sage publications.

Discovery in Action. (2020). *Behavioural change is the holy grail of leadership development*. Discovery in Action. Retrieved March 15th from https://discoveryinaction.com.au/behavioural-change-is-the-holy-grail-of-leadership-development/

Dwyer, L. P. (2019). Leadership self-efficacy: review and leader development implications. *Journal of Management Development, 38*(8), 637-650. https://doi.org/10.1108/JMD-03-2019-0073

Eagly, A. H., & Johnson, B. T. (1990). Gender and leadership style: A meta-analysis. *Psychological Bulletin, 108*(2), 233-256. https://doi.org/10.1037/0033-2909.108.2.233

Ely, R. J., Ibarra, H., & Kolb, D. M. (2011). Taking gender into account: Theory and design for women's leadership development programs. *Academy of Management Learning & Education, 10*(3), 474-493. https://doi.org/10.5465/amle.2010.0046

Forsvaret. (2020). *Forsvarets grunnsyn på ledelse*. https://www.forsvaret.no/soldater-og-ansatte/regelverk/Forsvarets-grunnsyn-ledelse.pdf/_/attachment/inline/74262ea3-bbbd-468d-8464-66136d4c7391:2579a67ab56d8ef777618b0da381a722513075b2/Forsvarets%20grunnsyn%20pa%CC%8A%20ledelse.pdf

Gist, M. E. (1987). Self-efficacy: implications for organizational behavior and human resource management. *Academy of Management Review, 12*(3), 472-485. https://doi.org/10.2307/258514

Goldstein, I. L., & Ford, J. K. (2002). *Training in Organizations*. Wadsworth Centage Learning.

Halvorsen, M., & Johansen, S. I. (2013). Kjønn og ledelse: hvilken betydning har lederstil og kjønn for medarbeidernes oppfattelse av en effektiv leder? http://hdl.handle.net/11250/165401

Heen, S., & Stone, D. (2014). Find the coaching in criticism. *Harv Bus Rev*, 108-111. http://www.pdngkc.com/wp-content/uploads/2014/01/Find-the-coaching-in-Criticism-HBR-2014.pdf

Herman, R. E. (2005). HR managers as employee-retention specialists. *Employment Relations Today, 32*(2), 1-7. https://doi.org/10.1002/ert.20058

Hopkins, M. M., O'Neil, D. A., Passarelli, A., & Bilimoria, D. (2008). Women's leadership development strategic practices for women and organizations. *Consulting Psychology Journal: Practice and Research, 60*(4), 348-365. https://doi.org/10.1037/a0014093

Hrivnak, G. A., Reichard, R. J., & Riggio, R., E. . (2009). A Framework for Leadership Development In S. J. Armstrong & C. Fukami (Eds.), *The Sage handbook of management learning, education and development* (Vol. 1, pp. 456-475). SAGE.

Kirkpatrick, D. L. (1979). Techniques for evaluating training programs. *Training and development journal*.

Kroon, B., & Freese, C. (2013). Can HR practices retain flexworkers with their agency? *International Journal of Manpower, 34*(8), 899-917. https://doi.org/10.1108/IJM-07-2013-0169

Lacerenza, C. N., Reyes, D. L., Marlow, S. L., Joseph, D. L., & Salas, E. (2017). Leadership training design, delivery, and implementation: A meta-analysis. *Journal of Applied Psychology, 102,* 1686-1718. https://doi.org/https://doi.org/10.1037/apl0000241

Larsen, H. H. (1988). *Ledelsesutvikling*. Universitetsforlaget AS.

Lent, R. W., Brown, S. D., & Larkin, K. C. (1987). Comparison of three theoretically derived variables in predicting career and academic behavior: Self-efficacy, interest congruence, and consequence thinking. *Journal of Counseling Psychology, 34*(3), 293-298. https://doi.org/10.1037/0022-0167.34.3.293

Leskiw, S. L., & Singh, P. (2007). Leadership development: Learning from best practices. *Leadership & Organization Development Journal, 28*(5), 444-464. https://doi.org/10.1108/01437730710761742

Lind, H. R. (2007). Lederutvikling gjennom 50 år. En empirisk studie av Solstrandprogrammets formidling av kunnskap om organisasjon og ledelse i perioden 1953-1998. Universitetet i Tromsø. https://hdl.handle.net/10037/1083

McCormick, M. J. (2001). Self-efficacy and leadership effectiveness: applying social cognitive theory to leadership. *Journal of Leadership Studies, 8*(1), 22-33. https://doi.org/10.1177/107179190100800102

Midtun, E. K., & Jordahl, A. (2002). Lederutvikling - virker det? En evalueringsstudie av Solstrandprogrammet In *The Solstrand Program* (Vol. 4, pp. 16-23). AFF.

Nawaz, Z., & Khan, I. (2016). Leadership theories and styles: A literature review. *Leadership, 16*(1), 1-7. https://www.researchgate.net/profile/Allah-Nawaz-2/publication/293885908_Leadership_Theories_and_Styles_A_Literature_Review/links/56bcd3ad08ae9ca20a4cdea2/Leadership-Theories-and-Styles-A-Literature-Review.pdf

Neuberger, O. (2016). *Personalentwicklung*. De Gruyter Oldenbourg.

Northouse, P. G. (2019). Introduction. In P. G. Northouse (Ed.), *Leadership: Theory and Practice* (8 ed., pp. 1-18). SAGE.

O'Loughlin, K. (2013). Leadership Development Programs and Participant Behavioral Change. *Seminar Research Paper Series*. http://digitalcommons.uri.edu/lrc_paper_series/29

Pallant, J. (2020). *SPSS Survival Manual* (7 ed.). Routledge

Paustian-Underdahl, S. C., Walker, L. S., & Woehr, D. J. (2014). Gender and perceptions of leadership effectiveness: A meta-analysis of contextual moderators. *Journal of Applied Psychology, 99*(6), 1129-1145. https://doi.org/10.1037/a0036751

Powell, K. S., & Yalcin, S. (2010). Managerial training effectiveness: A meta-analysis 1952-2002. *Personnel Review, 39*(2), 227-241. https://doi.org/10.1108/00483481011017435

Prochange.com. (2020, February 3rd). *Transtheoretical Model (or Stages of Change) - Health Behavior Change*. Retrieved November 18th from https://www.prochange.com/transtheoretical-model-of-behavior-change

Prochaska, J. O., & DiClemente, C. C. (1982). Transtheoretical therapy: Toward a more integrative model of change. *Psychotherapy, 19*(3), 276-288. https://doi.org/10.1037/h0088437

Prochaska, J. O., DiClemente, C. C., & Norcross, J. C. (1992). In search of how people change: Applications to addictive behaviors. *American Psychologist, 47*(9), 1102-1114. https://doi.org/10.1037/0003-066X.47.9.1102

Prochaska, J. O., Redding, C. A., & Evers, K. E. (2015). The Transtheoretical Model and Stages of Change. In K. Glanz, B. K. Rimer, & K. Viswanath (Eds.), *Health Behavior: Theory, Research and Practice* (5 ed., pp. 125-148). Jossey-Bass.

Riggio, R. E. (2008). Leadership development: The current state and future expectations. *Consulting Psychology Journal: Practice and Research, 60*(4), 383-392. https://doi.org/10.1037/1065-9293.60.4.383

Ruderman, M. N., & Ohlott, P. J. (2005). Leading roles: What coaches of women need to know. *Leadership in action: A publication of the Center for Creative Leadership and Jossey-Bass, 25*(3), 3-9. https://doi.org/

Salkind, N. J. (2010, 2022/04/11). Encyclopedia of Research Design. https://doi.org/10.4135/9781412961288

Schaub, M., & Tokar, D. M. (2005, 2005/04/01/). The role of personality and learning experiences in social cognitive career theory. *Journal of Vocational Behavior, 66*(2), 304-325. https://doi.org/10.1016/j.jvb.2004.09.005

Shamir, B. (2011, 2011/04/01/). Leadership takes time: Some implications of (not) taking time seriously in leadership research. *The Leadership Quarterly, 22*(2), 307-315. https://doi.org/10.1016/j.leaqua.2011.02.006

Sims, H. P., & Manz, C. C. (1982). Social Learning Theory. *Journal of Organizational Behavior Management, 3*(4), 55-63. https://doi.org/10.1300/J075v03n04_06

Skogstad, A. (2015). Effektiv og ivaretakende ledelse: Viktige perspektiver og modeller. In S. Einarsen & A. Skogstad (Eds.), *Ledelse på godt og vondt: Effektivitet og Trivsel* (2 ed., pp. 15-45). Fagbokforlaget.

Sparebanken Vest. (2020). *Årsrapport* https://www.spv.no/om-oss/investor-relations/rapporter

Stogdill, R. M. (1948). Personal factors associated with leadership: A survey of the literature. *The Journal of psychology, 25*(1), 35-71. https://doi.org/10.1080/00223980.1948.9917362

Sturges, J. (1999). What it means to succeed: Personal conceptions of career success held by male and female managers at different ages. *British journal of management, 10*(3), 239-252.

Suvatne, S. S. (2020, October 28th). *Bruker millioner på amerikansk coaching*. Dagbladet. Retrieved October 5th from https://www.dagbladet.no/nyheter/bruker-millioner-pa-amerikansk-coaching/72996756

Tabachnick, B. G., & Fidell, L. S. (2014). *Using multivariate statistics: Pearson new international edition*. Pearson.

Tronsmo, P. (1996). Ledelse og lederutvikling i staten. In I. T. Colbjørnsen (Ed.), *Tanke, følelse, handling. Perspektiv på lederutviking AFFs Årbok 1996*. Cappelen Akademisk Forlag.

Tsai, W., & Ghoshal, S. (1998). Social capital and value creation: The role of intrafirm networks. *Academy of Management journal, 41*(4), 464-476. https://doi.org/10.2307/257085

Van Velsor, E., McCauley, C. D., & Ruderman, M. N. (2010). *The center for creative leadership handbook of leadership development* (Vol. 122). John Wiley & Sons.

Vardiman, P. D., Houghton, J. D., & Jinkerson, D. L. (2006). Environmental leadership development: Toward a contextual model of leader selection and effectiveness. *Leadership & Organization Development Journal, 27*(2), 93-105. https://doi.org/10.1108/01437730610646606

Appendix

Appendix A - Respondents, percentages, means and standard deviations of the RT-survey

Variables	*n*	%	M	SD
Gender				
Male	34	62		
Female	21	38		
Items				
How inspiring was the gathering?			4.70	.50
To which degree were you able to practice relevant skills=			4.62	.53
To which degree will you change your behavior due to the gathering?			4.49	.64

Note: Response rate out of 86 participants was 65 %.

Appendix B - T1-survey in Norwegian

Gender

	Kjønn	**Mann**	**Kvinne**	**Annet**

Age

	Alder	**20-30 år**	**31-40 år**	**40-50 år**	**50-60 år**	**61 år eller eldre**

Department

	Avdeling	**PM**	**BM**	**Lederspirene**	**Stab og støtteavdelinger (SDD, OBS, RM, Øk og fin)**	**Bulder og andre**

Leadership experience

	Hvor lenge har du vært leder?	**Under 2 år**	**2-5 år**	**6-10 år**	**11 år eller mer**

Expectations

	Forventninger	**Helt uenig**	**Uenig**	**Hverken enig/uenig**	**Enig**	**Helg enig**
1	Jeg tror jeg vil lære noe på samlingen som jeg kan bruke i min hverdag.	1	2	3	4	5
2	Jeg er motivert for å delta i lederutviklingsprogrammet.	1	2	3	4	5
3	Jeg lærer ting på lederutviklingsprogrammet som jeg ikke lærer av vanlig praksis som leder.	1	2	3	4	5
4	Jeg tror denne ledersamlingen vil gjøre meg til en bedre leder.	1	2	3	4	5
5	Jeg tror denne ledersamlingen vil gi meg bedre lederferdigheter.	1	2	3	4	5

Hunger and Leadership

	Sult og ledelse	**Helt uenig**	**Uenig**	**Hverken enig/uenig**	**Enig**	**Helg enig**
6	Jeg føler at jeg og mine ansatte er mer sulten på forbedring enn vi var for 3 måneder siden.	1	2	3	4	5
7	Jeg føler meg trygg på at jeg har den kompetansen jeg trenger for å motivere mine ansatte til å jage forbedringer.	1	2	3	4	5
8	Jeg motiverer mine medarbeidere til å utvikle og forbedre seg.	1	2	3	4	5
9	Jeg oppmuntrer mine medarbeidere til å utfordre status quo.	1	2	3	4	5
10	Jeg har hatt konkrete samtaler om sult og utvikling med mine medarbeidere siste måned.	1	2	3	4	5
11	Jeg/min gruppe har forbedret prestasjonene våre signifkant siste måned.	1	2	3	4	5
12	Jeg føler sulten på forbedring hos meg og mine ansatte har økt i det siste.	1	2	3	4	5

Feedback

	Tilbakemeldinger	**Helt uenig**	**Uenig**	**Hverken enig/uenig**	**Enig**	**Helg enig**
13	Jeg føler meg kompetent til å gi og få ærlig tilbakemelding.	1	2	3	4	5
14	Jeg er alltid åpen for konstruktive tilbakemeldinger fra mine medarbeidere.	1	2	3	4	5
15	Den siste uken har jeg etterspurt uformell tilbakemelding fra mine medarbeidere på et konkret område.	1	2	3	4	5
16	Jeg føler jeg har tilstrekkelig kompetanse og verktøy for å ta krevende tilbakemeldingssamtaler.	1	2	3	4	5
17	Jeg jobber aktivt med egen utvikling basert på de konstruktive tilbakemeldingene jeg mottar.	1	2	3	4	5
18	Jeg er god på å ta tilbakemeldinger.	1	2	3	4	5
19	Jeg unngår noen ganger vanskelige tilbakemeldingssamtaler. *	1	2	3	4	5
20	Jeg er god på å gi tilbakemeldinger.	1	2	3	4	5

Note: *=This item was reversed in the analysis

Appendix C - Reactions-survey in Norwegian

	Reaksjoner	**Ikke i det hele tatt**	**Ikke veldig**	**Hverken eller**	**Litt**	**Veldig**
1	Hvor inspirerende var samlingen?	1	2	3	4	5
2	I hvor stor grad fikk du trent på relevante ferdigheter?	1	2	3	4	5
3	I hvor stor grad vil du endre atferd som følge av samlingen?	1	2	3	4	5

Appendix D - T2-survey in Norwegian

Gender

	Kjønn	**Mann**	**Kvinne**	**Annet**

Age

	Alder	**20-30 år**	**31-40 år**	**40-50 år**	**50-60 år**	**61 år eller eldre**

Department

	Avdeling	**PM**	**BM**	**Lederspirene**	**Stab og støtteavdelinger (SDD, OBS, RM, Øk og fin)**	**Bulder og andre**

Leadership experience

	Hvor lenge har du vært leder?	**Under 2 år**	**2-5 år**	**6-10 år**	**11 år eller mer**

Experiences

	Erfaringer	**Helt uenig**	**Uenig**	**Hverken enig/uenig**	**Enig**	**Helg enig**
1	Det jeg lærte og erfaringene jeg gjorde har kommet organisasjonen til gode	1	2	3	4	5
2	Det jeg lærte på samlingen har jeg fått anvendt i praksis	1	2	3	4	5
3	Det jeg lærte kunne jeg ikke ha lært kun i vanlig praksis som leder	1	2	3	4	5
4	Samlingen var en positiv erfaring	1	2	3	4	5

Hunger and Leadership

	Sult og ledelse	**Helt uenig**	**Uenig**	**Hverken enig/uenig**	**Enig**	**Helg enig**
5	Jeg føler at jeg og mine ansatte er mer sulten på forbedring enn vi var for 3 måneder siden.	1	2	3	4	5
6	Jeg motiverer mine medarbeidere i større grad til å utvikle og forbedre seg	1	2	3	4	5
7	Jeg/min gruppe har forbedret prestasjonene våre signifkant siste måned	1	2	3	4	5
8	Jeg føler sulten på forbedring hos meg og mine ansatte har økt i det siste	1	2	3	4	5

Feedback

	Tilbakemeldinger	**Helt uenig**	**Uenig**	**Hverken enig/uenig**	**Enig**	**Helg enig**
9	Jeg føler meg mer kompetent til å gi og få ærlig tilbakemelding	1	2	3	4	5
10	Jeg er mer åpen for konstruktive tilbakemeldinger fra mine medarbeidere	1	2	3	4	5
11	Den siste uken har jeg etterspurt uformell tilbakemelding fra mine medarbeidere på et konkret område.	1	2	3	4	5
12	Jeg føler jeg har tilstrekkelig kompetanse og verktøy for å ta krevende tilbakemeldingssamtaler.	1	2	3	4	5
13	Jeg jobber aktivt med egen utvikling basert på de konstruktive tilbakemeldingene jeg mottar.	1	2	3	4	5
14	Jeg har blitt bedre på å ta imot tilbakemeldinger	1	2	3	4	5
15	Jeg unngår noen ganger vanskelige tilbakemeldings-samtaler.	1	2	3	4	5
16	Jeg har blitt bedre på å gi tilbakemeldinger.	1	2	3	4	5

Home office

	Hjemmekontor	**Helt uenig**	**Uenig**	**Hverken enig/uenig**	**Enig**	**Helg enig**
17	Det jeg lærte på samlingen har jeg fått bruk for på hjemmekontor	1	2	3	4	5
18	Jeg synes det er vanskelig å være en god leder på hjemmekontor	1	2	3	4	5
19	Sulten har falt mens vi har jobbet på hjemmekontor	1	2	3	4	5

Appendix E - T1-survey in English

Gender

	Gender	Male	Female	Other

Age

	Age	20-30	31-40	40-50	50-60	61 or older

Department

	Department	PM	BM	Aspiring Leaders	Staff & Support (SDD, OBS, RM, Øk & fin)	Bulder and others

Leadership experience

	How long have you been a leader?	Under 2 years	2-5 years	6-10 years	11 years or more

Expectations

	Expectations	Fully disagree	Disagree	Neither agree, nor disagree	Agree	Fully agree
1	I believe I am going to learn something at the gathering which I can use in my everyday life	1	2	3	4	5
2	I am motivated to participante in the leadership development program	1	2	3	4	5
3	I will learn things at the leadership development program that I would not have learned by regular practice as a leader.	1	2	3	4	5
4	I believe this gathering will make me a better leader.	1	2	3	4	5
5	Jeg tror denne ledersamlingen vil gi meg bedre lederferdigheter.	1	2	3	4	5

Note: *=This item was reversed in the analysis

Hunger and leadership

	Hunger and leadership	**Fully disagree**	**Disagree**	**Neither agree, nor disagree**	**Agree**	**Fully agree**
6	I feel that my employees and I are hungrier for improvement than we were three months ago.	1	2	3	4	5
7	I feel I have the competence I need to motivate my employees to strive for improvement.	1	2	3	4	5
8	I motivate my employees to develop and better themselves.	1	2	3	4	5
9	I encourage my employees to challenge the status quo.	1	2	3	4	5
10	I have had specific conversations regarding hunger and development with my employees the last month.	1	2	3	4	5
11	My group/I has/have improven our performance significantly the last month.	1	2	3	4	5
12	I feel the hunger for improvement has lately increased among my employees and I.	1	2	3	4	5

Feedback

	Feedback	**Fully disagree**	**Disagree**	**Neither agree, nor disagree**	**Agree**	**Fully agree**
13	I feel competent enough to give and receive honest feedback.	1	2	3	4	5
14	I am always open for constructive feedback from my employees.	1	2	3	4	5
15	I have requested informal feedback from my employees regarding a specific area the last week.	1	2	3	4	5
16	I feel I have adequate competence and tools to hold demanding feedback conversations.	1	2	3	4	5
17	I work actively with my own development based on the constructive feedback I receive.	1	2	3	4	5
18	I am good at receiving feedback.	1	2	3	4	5
19	Sometimes I avoid tough feedback conversations.*	1	2	3	4	5
20	I am good at giving feedback.	1	2	3	4	5

Appendix F - Reactions-survey in English

	Reactions	**Not at all**	**Not so much**	**Neither**	**A little**	**Very much**
1	How inspiring was the gathering?	1	2	3	4	5
2	To what degree did you get to exercise relevant skills?	1	2	3	4	5
3	To which degree will you change your behavior due to the gathering?	1	2	3	4	5

Appendix G - T2-survey in English

Experiences

	Experiences	Fully disagree	Disagree	Neither agree, nor disagree	Agree	Fully agree
1	What I have learned, and the experiences I have had, has benefited the organization.	1	2	3	4	5
2	I have utilized what I have learned at the gathering, in practice.	1	2	3	4	5
3	I could not have learned what I learned merely by practice.	1	2	3	4	5
4	The gathering was a positive experience.	1	2	3	4	5

Hunger and Leadership

	Hunger and leadership	**Fully disagree**	**Disagree**	**Neither agree, nor disagree**	**Agree**	**Fully agree**
5	I feel that my employees and I are hungrier for improvement than we were three months ago.	1	2	3	4	5
6	I motivate my employees to develop and better themselves more than before.	1	2	3	4	5
7	My group/I has/have improven our performance significantly the last month.	1	2	3	4	5
8	I feel the hunger for improvement has lately increased among my employees and I.	1	2	3	4	5

Feedback

	Feedback	Fully disagree	Disagree	Neither agree, nor disagree	Agree	Fully agree
9	I feel more competent to give and receive honest feedback.	1	2	3	4	5
10	I am more open for constructive feedback from my employees.	1	2	3	4	5
11	I have requested informal feedback from my employees regarding a specific area the last week.	1	2	3	4	5
12	I feel I have adequate competence and tools to hold demanding feedback conversations.	1	2	3	4	5
13	I work actively with my own development based on the constructive feedback I receive.	1	2	3	4	5
14	I have become better at receiving feedback.	1	2	3	4	5
15	Sometimes I avoid tough feedback conversations.*	1	2	3	4	5
16	I have become better at giving feedback.	1	2	3	4	5

Note: *=This item was reversed in the analysis

Home office

	Home office	**Fully disagree**	**Disagree**	**Neither agree, nor disagree**	**Agree**	**Fully agree**
17	I have been able to use what I have learned at the gathering when working from home.	1	2	3	4	5
18	I feel it is hard to be a good leader when working from home.	1	2	3	4	5
19	The hunger has decreased whilst working from home.	1	2	3	4	5

Appendix H – SPV's leadership development program handout

Agenda

Day 1		Day 2	
Time	**Theme**	**Time**	**Theme**
12:00	How do we look when we are hungry for development?	09:00	NBS Leadership.
13:45	Pause.	09:20	Hunger in practice, examples from leaders and aspiring leaders.
14:00	Group work: rundown before the gathering.	09:50	Pause.
15:00	Pause.	10:00	Feedback training.
15:10	Discussion in plenum – reviewing the group work.	12:00	Lunch.
15:40	Pause with lunch.	12:45	Prelude to group work.
16:00	Making others better through feedback: Suzanne Peterson through Zoom, group work.	13:00	Group work: How do we transfer knowledge to results in practice?
17:30	Pause.	13:45	End.
17:40	A bubbly sum-up with Prosecco.		
18:10	Joint trip to eat dinner.		

Appendix I – Means, Standard Deviations, Cronbach's alpha, and *n* with items that were excluded in the main analyses

T1

Variable	Mean	SD	α	*n*
Expectations	4.72	.416	.826	64
Hunger and Leadership	3.95	.492	.704	64
Feedback	4.20	.505	.769	64
Avoid Difficult Feedback	2.98	1.266	-	64

T2

Variable	Mean	SD	α	*n*
Experiences	4.53	.485	.749	45
Hunger and Leadership	3.74	.634	.761	45
Feedback	4.20	.455	.708	45
Avoid Difficult Feedback	3.33	1.148	-	45
Home Office	3.31	.967	.750	45