



BI Norwegian Business School - campus Oslo

GRA 19703

Master Thesis

Thesis Master of Science

Leader Behavior in Performance Appraisal Conversations:
The Fostering of Psychological Safety in Subordinates

Navn: Ellinor Therese Jacobsen, Ida Birgitte
Theisen Norbye

Start: 15.01.2020 09.00

Finish: 01.09.2020 12.00

Leader Behavior in Performance Appraisal Conversations: The Fostering of Psychological Safety in Subordinates

Hand-in date:
01.07.2020

Campus:
BI Oslo

Supervisor:
Ide Katrine Birkeland

Program:
Master of Science in Leadership and Organizational Psychology

Content

CONTENT I

ACKNOWLEDGEMENTS..... III

ABSTRACTIV

INTRODUCTION..... 1

THEORETICAL FRAMEWORK 3

PERFORMANCE APPRAISALS 3

TRUST 4

MODEL I AND MODEL II BEHAVIOR 5

Single-Loop and Double-Loop Learning..... 6

Listening as Model II Behavior 7

PSYCHOLOGICAL SAFETY..... 8

METHODOLOGY..... 10

RESEARCH DESIGN..... 10

DIRECTIVE QUALITATIVE CONTENT ANALYSIS 11

TEMPLATE CODING 11

PARTICIPANTS AND ORGANIZATIONAL CONTEXT 12

DATA COLLECTION 13

ETHICAL CONSIDERATIONS..... 13

DATA ANALYSIS 14

CODEBOOK..... 14

SURVEY MEASURES 15

Common Items 15

Leader Items 16

Subordinate Items 16

ANALYTICAL PROCEDURES..... 16

FINDINGS 18

OVERALL FINDINGS FROM SURVEY DATA 18

OVERALL CODING FINDINGS FROM THE PERFORMANCE APPRAISALS 19

Model I and Model II Behaviors in Leaders..... 20

Leaders’ Trust-Building Behavior..... 21

Active-Empathetic Listening Behaviors in the Performance Appraisal..... 21

The Interaction Between Model II and Psychological Safety..... 22

DISCREPANCIES IN OBSERVED BEHAVIOR AND SURVEY RATINGS 23

OVERALL FINDINGS FROM WITHIN THE PERFORMANCE APPRAISALS..... 24

Leaders’ Lack of Open Inquiry..... 24

Leaders’ Model I Collaborative Planning..... 26

Active-Empathetic Listening as Model II behavior 26

<i>Leaders Ensuring Psychological Safety</i>	28
<i>Counterproductive Laughter</i>	30
<i>Giving Subordinates Feedback</i>	30
DISCUSSION	32
WHAT MAKES THE PA A POTENTIALLY UNCOMFORTABLE ORDEAL FOR BOTH THE LEADER AND THE SUBORDINATE?	32
<i>Insufficient Use of The Performance Appraisal Framework</i>	32
<i>Lack of Relevant Inquiry</i>	32
<i>The ‘Humming-Trap’</i>	33
<i>Poor Collaborative Planning</i>	34
<i>Inadequate Feedback</i>	34
<i>Off-Topic and Laughter</i>	35
<i>Is there Really a Shared Reality?</i>	35
<i>Potential Overconfidence in Leaders</i>	36
HOW DOES LEADERS FOSTER PSYCHOLOGICAL SAFETY WHEN CONDUCTING PERFORMANCE APPRAISALS?	36
GENERAL DISCUSSION	38
THEORETICAL AND PRACTICAL IMPLICATIONS	40
LIMITATIONS AND FUTURE RESEARCH	41
CONCLUSION	43
REFERENCES	44
APPENDIX	53

Acknowledgements

First and foremost, we want to express our deepest gratitude to our supervisor, Associate Professor Ide Katrine Birkeland, for her time, extensive feedback, and support throughout this process. You have been such a positive source of inspiration and motivation. Further, we want to thank you for introducing us to such an interesting field of research. It has been inspiring and something we will bring with us into both our work life and private life going forward.

Secondly, we would like to thank the research group that we were lucky to be part of, for great discussions and brainstorming.

We are also extremely grateful towards the organization who let us in, and granted us access to their performance appraisal conversations, regardless of our ‘unconventional’ methodology. Especially, we want to thank our liaison for their continuous promotion of the project and recruitment of participants, not to mention the great collaboration we have had.

Further, we want to express gratitude towards friends and family who have supported us through the rough patches and cheered with us during the highlights. Ida’s roommates deserve a special thanks for letting us occupy their living room for two full months during the Covid-19 pandemic.

Last but not least, we would like to thank each other. This project has been extremely rewarding and fun, and this is due to our great teamwork and continuous support of each other. We went into this project as great friends and come out even closer.

Oslo, 25th of June 2020



Ellinor Therese Jacobsen



Ida Birgitte Theisen Norbye

Abstract

Purpose: The purpose of this study is to shed light on the leader-subordinate interaction as it unfolds during the performance appraisal conversation. More specifically, the aim is to look into how Model II behavior in the leaders may contribute to psychological safety in the subordinates. By doing this, we answer the call of more qualitative studies in this field, which is needed in order to bridge the gap between scholars and practitioners.

Research Methodology: The study was conducted by recording the performance appraisals of five leader-subordinate dyads in an organization, and by handing out a survey to the parties in the dyad afterwards. The performance appraisals were transcribed and coded based on a codebook partially developed by Meyer and colleagues (2019). The coded findings were then compared to answers from the survey. Additionally, patterns found within the performance appraisal conversation were highlighted.

Findings: Our findings suggest that leaders managed to maintain the relationship with the subordinate, but especially due to a lack of inquiry, the subordinate's voice remained limited. The result is an ineffective PA, suppressed psychological safety, and limited learning opportunities for both parties. Further, our findings propose that the overuse of active-empathetic listening behavior such as humming can be counterproductive.

Implications: Our overall findings suggest keeping the traditional performance appraisal, but with some alterations and leadership development in order to make it more effective. Further, our findings arguably reflect the importance of the leader's behavioral nuances in relation to psychological safety in the performance appraisal. This notion directs future research into also looking at authentic behavior, and not only subordinates' perceptions in surveys, when studying the performance appraisal, not.

Introduction

In today's world of business, it is unlikely to find an organization without any form of formal performance appraisal management. This is probably because it is found to be beneficial for several important subordinate attitudes (Brown et al., 2010). However, it can yield negative outcomes as well (Kuvaas, 2006; Sumelius et al., 2014). A quick Google search reveals a number of sources discussing how both leaders and subordinates are reluctant to the performance appraisal conversation (PA) (e.g. Cappelli & Tavis, 2016; Grint, 2003; Sias, 2014). Reasons for reluctance include use of numerical ratings, poor communication, and the PAs distance from reality (Grint, 2003; Roberts & Prgitzer, 2007). One suggestion to why the PA may be perceived negatively is that the leader's understanding of the social dynamics remains limited (Miller & Gordon, 2014; Meineke et al., 2017). For most organizations, the PA between leaders and subordinates is considered an important part of the appraisal system (Kuvaas, 2006), thus there has been conducted a substantial amount of research to be able to design good practices thereafter. However, the complex context of the PA has made it difficult to establish practices that generate consistent positive outcomes across subordinates (Brown et al., 2019; Miller & Gordon, 2014).

The PA may be the only arena where subordinates get to sit down with their leader and communicate on a deeper level (Roberts, 2002). Moreover, several leadership behaviors have been found to contribute positively to the sense of psychological safety and trust in the leader-subordinate relationship (Edmondson, 1999; Kim et al., 2018; Levy & Williams, 2004). Furthermore, the level of trust affects the overall effectiveness of the PA. This suggests that the leader's abilities to build trust will play an important role in whether the PA is perceived successful by the subordinate or not. Nevertheless, it still remains unclear how this is actually formed (Fenniman, 2010).

In a recent literature review by Brown and colleagues (2019) on PAs, it was found that most previous research has been conducted using quantitative methods, thus making a call for more qualitative methods to better capture the multidimensional nature of PAs in organizations. Further, Miller and Gordon (2014) encourages studying the interaction in the PA and help HRM-practitioners in developing effective conversational practices thereafter. We draw on these notions and conduct a qualitative study with an emphasis on an aspect of the PA that is not much touched upon: how leader behavior can promote psychological

safety, which may stimulate for organizational learning (Edmondson, 1999). To do this, we will shed light on the leader-subordinate interaction as it unfolds during the PA. Our interest is to discover how these interactions affect the dynamic between leader and subordinate during the PA. Further, PA perceptions have been found to have a significant impact on the perceived overall effectiveness of the appraisal system (Kuvaas, 2008), therefore it is interesting to uncover *how* employees were led to these perceptions.

This study aims at contributing to bridging the gap between scholars and HR-practitioners. Further, researchers highlight the need for studies using ‘real people’ and real PAs, instead of situational interviews and simply ticking boxes (Brown et al., 2019; Le Fevre & Robinson, 2015). To do this, we further incorporate Argyris & Schön’s (1974) research on professional effectiveness. They proposed two behavioral modes that may be exhibited by the leader, Model I and Model II. Model II reasoning is governed by values that may lead to openness, trust and transparency. As being open and honest as a leader is key to ensure organizational learning, and with a notion of subordinate psychological safety promoting organizational learning, we propose that Model II behaviors are important to both ensure psychological safety, and further promote organizational learning. Therefore, we propose an investigation of the following question:

How can the leader’s actions and inactions explain or relate to the level of perceived psychological safety for subordinates during and after performance appraisal conversation?

RQ1: What makes the PA a potentially uncomfortable ordeal for both the leader and the subordinate?

RQ2: How does leaders foster psychological safety when conducting performance appraisals?

Theoretical Framework

Performance Appraisals

Performance appraisal conversations may be defined as recurrent conversations between the leader and the subordinate, where the focus is on subordinate performance and development (Asmuß, 2008). Annual performance appraisals (PAs) are critical to the effectiveness of both subordinates and the organization (Meinecke et al., 2017). If the PA is effective, it enables subordinates to speak up, state their own opinions, and plan future actions. Internal communication has been highlighted as necessary for increased organizational performance; thus, poor organizational communication might limit organizational performance (Asmuß, 2013). Recently, some organizations have chosen to move away from the traditional PA conversations being an annual evaluation exercise towards providing real-time feedback to the subordinates (Cappelli & Tavis, 2016). However, research done by Roberts (2002) has concluded that only the conversation part of the PA can give insights on subordinate voice. Therefore, instead of arguing for a removal of the traditional PA entirely, the current study strives to generate more knowledge on how to do it more effectively, thus enhance psychological safety and trust, that further promotes learning as perspectives are exchanged.

The majority of previous research is based on questionnaires, and not on authentic PA conversations that offer insights into what actually happens during the PA. Asmuß (2013) argues that PA conversations should be studied as dialogical interactions between equal partners, as this provides researchers with a social interaction ‘lens’ that offers the opportunity to look at different concepts compared to what can be drawn from questionnaires. For instance, one might look at real-time communication patterns as they unfold during the conversation. This pattern is, however, likely to be influenced by the degree of structure, topics, and current dyadic power distance between the leader and the subordinate (Asmuß, 2013).

Nishii and Wright (2007) argue that it is the subordinate perceptions, and not the specific practices, that subsequently influence outcomes of the PA. The PA conversation is a timely process and one of the few direct and tangible encounters subordinates have with HRM practices. Thus, the feelings that arise from PAs are interesting to look at, as they are arguably determinant to perceptions of other

features of the HRM-system (Bowen & Ostroff, 2004). Additionally, previous research has demonstrated a positive correlation between PA perceptions and important factors such as organizational commitment, trust in management, and individual performance (Kuvaas, 2006; Kuvaas, 2008). Kuvaas' studies further propose that the PA may provide negative outcomes, specifically for subordinates low on intrinsic motivation, who are in even more need of a positive encounter with the PA. However, it is important to note that Kuvaas' studies do not consider leader behavior, only subordinate perceptions. Brown and colleagues (2010) have suggested four indicators of a positive PA experience: trust, communication, clarity, and fairness. In addition, the experience is dependent on how subordinates are treated by their leader during the PA. In the current study, trust and communication will be emphasized as these refer to the extent of information that flows in the dyad, and the belief that the leader is competent and will act in the subordinate's best interest (Brown et al. 2010).

Trust

Trust is both a key ingredient in problem-solving as well as an important element in the relationship between the leader and the subordinate (Kim et al., 2018). Further, trust is essential in interpersonal communication and in order to reach a shared reality of the state of affairs (Hardin & Conley, 2000). Shared reality is "the product of the motivated process of experiencing with others, a commonality of the inner states about the world" (Echterhoff et al, 2009, p. 496). When such shared realities are non-existent, they may impair the relationship dynamics and mutual reciprocity in the leader-subordinate relationship (Hardin & Conley, 2000).

In an organizational setting, Tan and Lim (2009) have differentiated between trust in coworkers and trust in the organization. They have defined *trust in coworkers* as "the willingness of a person to be vulnerable to the actions of fellow coworkers whose behavior and actions the person cannot control" (p. 46). Further, *trust in organizations*, has been defined as "employee willingness to be vulnerable to the actions of the organization, whose behavior and actions he or she cannot control" (Tan & Lim, 2009, p. 45). In the current study, trust will mainly be seen in the light of the last definition, as the leaders represent the action of the organization in relation to the subordinate.

Additionally, Dirks and Ferrin (2002) categorize trust into two dimensions; affective and cognitive. *Affective trust* is based on care and emotional bonds

between the one trusting and the one being trusted (Kim et al., 2018). There is a special relationship where concern about welfare can be demonstrated (Dirks & Ferrin, 2002). *Cognitive trust* relates to beliefs the one trusting has about the ability, reliability, fairness, and integrity of the one being trusted (Dirks & Ferrin, 2002; Kim et al., 2018). Both forms of trust have an important impact on outcomes of subordinates but have different importance under differing conditions (Dirks & Ferrin, 2002; Kim et al., 2018). In the PA, both kinds of trust can be seen as important; affective trust may play a role as the subordinate may interpret issues discussed in the PA differently based on the emotional bond between the leader and the subordinate. Further, cognitive trust is believed to play a significant role as the subordinate must trust the leader's competence to value their opinions.

“One critical determinant of trust is the leader’s ability to deal with difficult issues in a respectful way” (Robinson, 2009, p. 1).

It is hard to develop a culture of trust and respect in the workplace if dealing with challenging issues are put off or done incompletely. Robinson (2009) states that a critical requirement for the development of trust is the ability to engage in behaviors involving: description of problematic situations, listening to the views of others, detecting and challenging own and others' assumptions, welcoming alternative views, ability to both give and receive negative feedback, and to deal constructively with conflict. These skills are related to what Chris Argyris (Argyris, 1976, 1993, 2010; Argyris & Schön, 1974) referred to as Model II behavior and will be detailed in the following section.

Model I and Model II Behavior

Chris Argyris (Argyris, 1976, 1993, 2010; Argyris & Schön, 1974) introduced the concepts of Model I and Model II as models of reasoning, where Model I refers to defensive reasoning, while Model II refers to productive reasoning. The purpose of Model I is to defend and protect the individual against change that feels disruptive (Argyris, 2010). This will lead to a mind-set where the individual becomes defensive in explaining their own actions, often when facing potentially embarrassing or threatening situations. Model II can be used to prevent the counterproductive effects of Model I. However, to develop the productive

reasoning mindset of Model II, it is crucial to be able to produce Model II governing values. These values can further lead to openness, trust and transparency.

A central issue is that individuals tend to support the Model II reasoning, but the behaviors they enact are still based on Model I (Argyris, 2010). Argyris and Schön's (1974) work referred to this potential discrepancy as espoused theory and theory-in-use. Individuals are seldomly aware of their theory-in-use; they may believe that they act in a Model II fashion, yet their actual behavior is often characterized by Model I.

To enact Model II and productive reasoning, and thus build trust, the challenge is to create a dialogue where participants can clear up uncertainties and assumptions, as well as reconcile differences. This is especially important in the PA, since conversations about performance quality and improvement can be difficult, as they may trigger discomfort and defensiveness, thus pose a threat to the relationship between leader and subordinate (Robinson, 2009).

Single-Loop and Double-Loop Learning

Model I and Model II behaviors promote fundamentally different learning modes; single-loop and double-loop (Argyris, 1976, 1993). Single-loop learning is sufficient in achieving current policies and objectives, therefore maintaining status quo, as it simply corrects errors by changing routine behavior. However, in the dynamic environment that most organizations are facing today, single-loop learning will simply postpone problems rather than solve them. Argyris (1976) suggests that children learn Model I during their upbringing, with influences from e.g. parents, school, and social settings in general. With most people being unaware of their Model I behavior, they go on enacting it despite its disadvantages (Argyris & Schön, 1974). In contrast, Model II and double-loop learning occurs when errors are corrected by examining the underlying assumptions, values, and policies of the organization, thus increasing growth, learning and effectiveness. By doing so, organizational members may become more effective, strengthen the quality of relationships, and continuously improve the organization through organizational learning. Organizational learning is the outcome of the process of assimilating and translating information into knowledge, applying this knowledge and then revising the information and reshape knowledge after feedback from real-life situations (Gill, 2000b). This means that subordinates learn, not primarily from training or

workshops, but from their everyday work life. Thus, to ensure that organizational learning occurs, communication and dialogue is important (Thames, 2015).

Model II thinking focuses on how people are searching to improve the quality of their own thinking and decision-making (Robinson, 2009). Contrarily, in Model I thinking, people assume their views are valid and try to impose these on others, often in a nice and polite manner. The difference between these two relates to learning each other's points of view and being open to both these views being valid. Further, there are three behavioral dimensions to Model II that may determine the effectiveness and collaborativeness of conversations: advocacy, inquiry, and collaborative planning (Meyer et al., 2019).

The most central tenet to conversations where a leader exhibits Model II behavior is the ability to tackle difficult matters while maintaining trust-building with subordinates. Nevertheless, solving problems and building trust is rarely seen by leaders as things that may occur simultaneously as leaders tend to see trust as something that must be built *before* they can engage in difficult conversations (Robinson, 2009). However, the Model II framework suggests that problem-solving and trust building should, in fact, co-occur; in this framework the leader and the subordinate are perceived as collaborators, they are less defensive, and more open to learning (Argyris & Schön, 1974). Further, Argyris & Schön (1974) predicts that in a “behavioral world of Model II” (p. 91), individuals will have a stronger sense of ‘psychological success’. The current definition of ‘psychological safety’ was not specified at the time when Argyris’ and Schön’s book was published, as it was coined by Edmondson in 1999 (Fenniman, 2010). We therefore suggest that what Argyris and Schön (1974) meant by ‘psychological success’ can be related to today’s understanding of psychological safety and will elaborate on this shortly.

Listening as Model II Behavior

An aspect of learning others’ point of view is to simply listen (Romero, 2001). The construct of listening has been hard to define and involve several dimensions (Castro et al., 2016). However, after reviewing several different aspects, Castro and colleagues (2016, p. 763) have chosen to define listening as “a behavior that manifests the presence of attention, comprehension and good intention towards the speaker”. Rogers (1951) stated that listeners need to create an atmosphere of safety for the speaker, which has later been referred to as psychological safety (Castro et al., 2016). Itzhakov and colleagues (2017) has found that high quality

listening elicits awareness and tolerance of inconsistencies in terms of objective attitude, and therefore the speaker's attitude is changeable by merely being provided with high-quality listening. Empathy has been closely associated with listening (Drollinger et al., 2006), and as empathy entails an understanding of the others point of view, it also requires an open exploration of it, for example through Model II behavior. Drollinger (2006) further introduces the concept of active-empathetic listening. As listening creates an atmosphere of psychological safety, and Model II behavior elicits a strong sense of psychological success, active-empathetic listening can be seen as Model II behavior.

Psychological Safety

The term psychological safety is defined as “people’s perceptions of the consequences of taking interpersonal risks in a particular context such as a workplace” (Edmondson & Lei, 2014, p. 23). It entails taken-for-granted beliefs of how others will respond when one shows and employs one’s self (Castro et al., 2016; Edmondson, 2004). Psychological safety decreases concern about being seen as incompetent when asking for help from others in superior positions (Carmeli et al., 2009). Further, it lessens concerns about feedback, as there is little fear of criticism or humiliation. If one experiences psychological safety, actions will be taken without fear of negative consequences to aspects such as status, career, or one’s self-image (Castro et al., 2016). Experiencing psychological safety at work is important for learning behaviors such as information seeking, paying attention to feedback, asking for help, and speaking up to test assumptions (Tynan, 2005). In contrast, a lack of psychological safety is associated with lowered levels of performance, learning, well-being, innovation, and adaptiveness both outside and within the corporate setting (Roberto, 2002).

Leadership behavior has been found to be a strong predictor of team-level psychological safety (Tynan, 2005), and most of the research in the field has focused on team performance (Edmondson, 1999; Tynan, 2005). In terms of psychological safety on an individual level, the research has mainly focused on areas such as organizational commitment, job engagement, creativity, and learning from mistakes (Edmondson & Lei, 2014). Drawing on that, Tynan (2005) has introduced the concept of psychological safety as a dyadic construct between the leader and subordinate. She distinguishes between self-psychological safety and others-psychological safety in dyadic leader-subordinate relationships. This

concept is tightly coupled with team-level psychological safety, yet it is aimed at measuring the subordinate's perception of psychological safety in the leader-subordinate dyad and is therefore applicable to the PA conversation context. Tynan (2005) defines self-psychological safety as the degree to which an individual feels emotionally safe with another, for example how much the individual feels trusted or respected by the other. Further, other-psychological safety is defined as how safe the individual perceives the other to be in the relationship. For instance, if the subordinate is low in other-psychological safety, they will be more likely to tiptoe around the leader's feelings.

Psychological safety is further linked to subordinates speaking up and 'voice' (Detert & Burris, 2007; Edmondson & Lei, 2014). Detert and Burris (2007) defines 'voice' as a provision of information with the intention of improving function within the organization or making a perceived authority act to improve the organization's well-being, even though this information may challenge or upset the status quo. Challenging the status quo and offering ideas for improvement is a vital force for helping organizational learning (Edmondson & Lei, 2014). Detert and Burris (2007, p. 871) explained psychological safety in relation to voice as "the belief that engaging in risky behaviors like voice will not lead to personal harm".

Methodology

The purpose of this study is to gain insights into what can potentially be uncomfortable about the PA and how leaders foster psychological safety during the PA. Because our research questions aim at gaining an in-depth understanding of the *how* in the PA, we have chosen a qualitative strategy of analysis in order to acknowledge the richness of our data.

Research Design

To answer our research question, we used triangulation and collected data from three different sources (the leader survey, the subordinate survey and the PA interaction), thus providing a strong design which helps to minimize common method bias due to separated sources of variables (Podsakoff et al., 2003). Further, this supports Guest's (2011) suggestion that multisource data in particular is wanted in HRM research, as it is naive to assume that *one* perspective alone (e.g. leader) can provide valid information about a phenomenon, such as the PA interview.

The study has an exploratory design, which is facilitated through directed content analysis with an aim of gaining new ideas, insights, and understanding. Qualitative research strategy usually emphasizes words rather than numbers when collecting and analyzing data, and the main focus is to understand the social world through examining the interpretation of participants experiencing that world (Bell et al., 2019). Our research approach is interpretivist in its epistemology, entailing that it is primarily concerned with understanding human behavior rather than explaining it. Interpretivism is underpinned by a social constructionist ontology, suggesting that categories, e.g. culture, are socially constructed entities coming to life by the actions and understandings of humans. Lastly, our approach is fundamentally naturalistic as a great concern is to study the participants in their natural habitat with as little researcher interventions as possible (Bell et al., 2019).

The benefits of qualitative research involve being able to collect deep and rich data in a natural setting for the participants (Bell et al., 2019). Further, there is more room for attention to, and interpretation of, details that could otherwise seem trivial. However, some criticism of qualitative research has to be noted, especially issues related to subjectivity and replicability. As qualitative research has less structure and there is no standard procedure, there is almost impossible to conduct a true replication of previous studies. Traditionally, qualitative methods have been associated with an inductive approach where the theory emerges from the data

(Hyde, 2000). This study, however, despite its qualitative method, exhibits a more deductive approach where we let the current theory color our data, yet with inductive elements where new things are allowed to emerge from the data.

Directive Qualitative Content Analysis

The data was examined through directive qualitative content analysis (DQCA), where the goal is to “validate or extend conceptually a theoretical framework or theory” (Hsieh & Shannon, 2005, p. 1281). With the use of existing theory and prior research, DQCA exhibits a deductive approach, as it helps focus the research questions and determine the coding scheme and relationship between codes (Hsieh & Shannon, 2005). Further, the predetermined codes are used to search for themes and meanings in the data. The objective is to structure and transform large amounts of unstructured data, thus facilitating a more systematic analysis. The main strength of DQCA is that it allows for existing theory to be supported and extended, however, as existing theory sets the direction of the research, it also poses a strong bias that is important to be aware of. Lewinski and colleagues (2019) proposed using content analysis by incorporating aspects of conversation and discourse analysis when analyzing synchronous conversations. Insights of the latter methods provide the possibility of examining *how* behaviors are used in a conversational sequence with an aim of in-depth analysis rather than generalization (Dick, 2004; Lewinski et al., 2019). Further, the analysis was partly facilitated through the computer-assisted qualitative data analysis software, ATLAS.ti. It allowed us to code, retrieve and work efficiently with the large amounts of unstructured data.

Template Coding

Coding involves the data being broken down into components where the different parts are given labels (Bell et al., 2019). With the answer to our research questions being hidden in the data, a coding template was designed using a priori codes and categories drawn from the theoretical background and questions of interest of the current study. It is important that in the early stages, the coding is as inclusive as possible to ensure that all the relevant data is counted for (Gill, 2000a). Using a coding template, as opposed to open coding where the codes emerge from the data, allows the data to speak *through* the researcher rather than *at* them (Blair, 2015). This notion highlights the importance of reflexivity: the researchers’ ability to constantly scrutinize and question their particular form of subjectivity (Blair,

2015; Peshkin, 1988). In this study, reflexivity was exhibited by going through the coded transcripts in several passes, addressing all divergences made by the researchers.

Participants and Organizational Context

The data of the current study was collected in a large Norwegian retail enterprise. The total dataset includes performance appraisal conversations and surveys from five different leaders and their respective subordinates (Table 1), resulting in five complete dyads. In order to get participants to the project, an insider liaison was responsible for the informing and recruiting of participants. The liaison did this by first establishing contact with a group of leaders, and then having the leaders seek consent from their respective subordinates. Snowball sampling was found to be the best sampling method in this case, as the organization was reluctant to let the researchers establish contact with possible participants directly. Participants chose to participate because of their genuine interest in the project and research topic. A frequent comment among the participants was that “they had nothing to hide nor to be ashamed of”. The liaison was in charge of all contact made with participants and making a list of scheduled conversations. In that way, we managed to collect data without collecting the name or contact information of the participants, as anonymity was a great concern for all involved parties.

Table 1

Performance appraisal overview

Interview	Participants	Dyadic Tenure	Length
PA1	Leader A ♂, 49 Subordinate a ♀, 38	6 months	1 h 5 min
PA2	Leader B ♀, 44 Subordinate b ♂, 34	2 months	1 h 20 min
PA3	Leader C ♀, 44 Subordinate c ♂, 42	8 months	1 h 26 min
PA4	Leader D ♂, 43 Subordinate d ♂, 28	2 years	1 h 21 min
PA5	Leader E ♂, 46 Subordinate e ♀, 41	1.5 years	57 min
Total: 5	5 leaders 5 subordinates		6 h 9 min (average: 82.8 min)

The PAs in the organization takes place annually in the spring. Nevertheless, most leaders conduct a follow-up conversation during the fall. Leaders in the organizations are recommended to follow an organization-specific conversational guide. However, this is done to a varying extent by the leaders. These conversations are often supplemented by shorter and informal ‘check-ins’, which are also done to a varying extent, ranging from never to weekly (Liaison, personal communication, December 03, 2019).

Data Collection

As the aim of the study is to get insight to what can potentially be uncomfortable in the PA and how leaders can foster psychological safety during the PA, the data collection was twofold. The PA data was recorded by the participants themselves, using Dictaphones that were handed out before the PA. By minimizing research interventions, the goal was to maintain the interaction as genuine as possible. Along with the Dictaphone, a paper-survey was distributed to both the leader and the subordinate which was completed separately as soon as they could after finishing the PA. This survey was used to investigate the differences and similarities between what was heard in the PA recording and what was reported in the individual surveys when the parties were not affected by each other's presence. The recordings were then transcribed and coded.

Ethical Considerations

In order to access real PAs, a confidentiality agreement with the organization where the PAs took place had to be signed. This was in order to maintain the security of sensitive information surfacing during the PAs. Immediately after picking up the Dictaphone, the audio files were uploaded into a secure cloud-service and deleted from the Dictaphone. As the surveys were in paper-format, all answers were logged and saved in a different secure cloud-service. In order to know which surveys belonged to which PA, each dyad was marked with its own unique code (e.g. 86WJQS6). All audio files were then transcribed, and all information that could lead to recognition of the participants or the company involved was changed to ensure complete anonymity of the dataset, as in accordance with ethical guidelines provided Norwegian Center of Research Data (NSD) and the American Psychological Association (APA). Subsequent to the finalized transcription, the audio-file was deleted from the cloud. The data collection method and data conservation in this study is approved by NSD.

Data Analysis

Codebook

For our analysis, the qualitative data was transcribed and coded in consensus using a codebook: after the initial coding, all 2881 codes were revised to ensure coding agreement (see Appendix A for codebook). Further, we identified and discussed patterns and segments that were of particular interest to the research questions. In order to discover patterns in the data, a codebook based on Model I and Model II behavior (or utterances) as well as the observable (audible) dimension of the traits measured in the surveys was utilized. The final codebook was a fusion of Model I and Model II codes developed by Meyer and colleagues (2019) and codes designed specifically to the context, scope, and research questions of the study. The codebook developed by Meyer and colleagues (2019) includes indicators of Model II behaviors such as advocacy, inquiry and collaborative planning. Indicators of Model I is the ‘negative counterpart’ of the Model II behaviors. Other indicators such as active-empathetic listening (AEL), based on work by Drollinger and colleagues (2006), and shared reality (SR) based on Echterhoff’s (2012) work, were included as a result of the proposed theoretical framework in relation to Model II-values.

In order to measure active-empathetic listening, we looked for utterances that could provide cues of the listening quality. Signs of listening may be manifested in different ways, for example by paraphrasing or giving confirmation such as “yes, I agree with that” or “mhm”. Shared reality was operationalized to the observed quality of agreement in between the parties. Codes were also made to distinguish what type of performance evaluation or feedback that was given (positive, neutral, and negative).

The codes developed for subordinates were specifically designed to detect utterances suggesting a state of trust and psychological safety. The codes for trust (T) and self/other-psychological safety (S-PS and O-PS) were developed based on the operationalization of Pugh and colleagues’ (2003) trust-in-supervisor scale and Tynan’s (2005) dyadic psychological safety scale. Based on the definitions used in the current study, a manifestation of subordinate trust may be valuing the leader’s opinion or asking the leader for advice. Likewise, subordinates high in self-psychological safety may be more prone to take interpersonal risk with their leader, such as opening up about difficult experiences at work or admitting weaknesses. As

for other-psychological safety, the subordinate may be less reluctant to express concern towards the leader or the organization. The main concern was to code utterances that could manifest any type of change in behavior as a result of an utterance made by the leader.

In the coding process, the aim was to code as fine-grained as possible. Thus, all coded units could consist of anything from silence to full sentences. To avoid forcing codes upon units showing different types of expressions or leaving relevant units uncoded, filler-codes such as laughter and interruptions were made (Lehmann-Willenbrock & Allen, 2018). These were applied to both leader and subordinate and were analyzed in relation to codes depicting Model I, Model II, and psychological safety.

Survey Measures

After the PA, both the leader and subordinate were asked to fill out a survey, thus offering the opportunity to discover congruences or discrepancies between observed and reported data. The survey was given out with a purpose of gaining a better understanding and support of our qualitative data. In order to reduce common method variance, the items for the constructs were intermixed (Podsakoff et al., 2003). The items in the survey followed a Likert scale ranging from 1 to 7 where '1' equals strongly disagree and '7' equals strongly agree. All measurements and items were based on already existing and verified scales (Drollinger et al., 2006; Dysvik & Kuvaas, 2011; Pugh et al., 2003; Schmalbach et al., 2019; Tynan, 2005), however, minor contextual adjustments and translation from English to Norwegian were done.

Common Items

The measurement of shared reality is based on definitions of shared reality theory (Echterhoff, 2012), with 4 items developed by Schmalbach and colleagues (2019). An example of the original items is 'I think that X and I are on the same wavelength with regards to Y'. The items were adapted to the current study, thereby resulting in items such as 'My leader and I see things in the same way' and 'My subordinate and I see things in the same way' as it was measured dyadically in the leader-subordinate relationship. Shared reality was measured as it is related to trust, and to get an indication of whether the parties left the PA with overall shared perception. In order to measure listening, items from Drollinger and colleagues'

(2006) scale on active-empathetic listening were utilized. This scale consists of 11 items measuring the three dimensions of active-empathetic listening: sensing, processing, and responding. Examples of items are ‘I am sensitive to what others are saying’ and ‘I assure others that I will remember what they say’. In the current study, our main interest is the leader’s ability to listen, therefore, the leader was self-reporting their own listening skills, and the subordinate was reporting on how they perceived their leader’s listening skills (same items with different angling). PA perceptions were measured using Kuvaas’ (2006) PA perceptions scale, consisting of subordinate items such as ‘I am satisfied with the way I receive feedback’. The leaders were also asked to report their PA-perceptions, hence Kuvaas’ (2006) items were flipped, resulting in items such as ‘I feel that I am good at giving feedback’.

Leader Items

In addition to common items, leaders were asked to report the effectiveness of the subordinate. This was done by utilizing Dysvik and Kuvaas’ (2011) employee effectiveness scale, with items such as ‘My subordinate intentionally expends a great deal of effort in carrying out their job’. This was done in order to gain insight to how the leader evaluated the subordinate and to discover whether it was congruent with feedback and praise given in the actual PA.

Subordinate Items

Subordinates were measured on psychological safety and trust. The measure of self-other psychological safety is based upon Edmondson’s (1999) team-level measure of psychological safety. Tynan (2005) broke the construct down in two interdependent measures to facilitate measurement at the individual level, with items such as ‘My leader had the best intentions towards me’ (self-psychological safety) and ‘At some level, I felt I had to tiptoe around my leader’s feelings’ (other-psychological safety). Trust was measured using Pugh and colleagues’ (2003) scale on supervisor trust, with items such as ‘I believe that my leader’s motives are good’.

Analytical Procedures

To answer our first research question exploring how the leader’s conversational behaviors may lead to both parties being uncomfortable in the PA-setting, we looked at the prevalence of the different codes in the transcribed PAs.

Thereafter, we explored segments which had a high amount of Model I codes. Further, we looked into specific segments which could exemplify why subordinates or leaders may feel uncomfortable for different reasons. Examples can be to not inquire into sensitive topics that were brought up, or continuing surface level conversation instead of discussing subordinate performance.

To answer our second research question concerning how leaders foster psychological safety in the PA, we looked at which leader behaviors led to expressed psychological safety by the subordinate. Model I and Model II behaviors surrounding psychological safety expressions were further interpreted in context to see patterns. Further, we looked for patterns of non-action made by the leader, e.g. not inquiring at all into the other's beliefs when the subordinate initially displayed psychological safety by showing vulnerability or concern.

To support and understand findings in the PA conversations related to the research questions we compared these to the survey data. Firstly, we looked at how the performance evaluation by leaders differed between the codes in the PA and the answers from the survey, thus offering insights to the leader's espoused theory versus theory-in-use. Thereafter, we looked for patterns in the survey measures for both leaders and subordinates and the coded transcripts in relation to active-empathetic listening, psychological safety, trust and shared reality, and compared these.

Findings

In this chapter, the findings from both surveys and the coding of the PAs will be presented and discussed. Firstly, the survey findings will be presented. Thereafter, the overall findings and large patterns from the PAs will be elaborated in-depth with excerpts. One key finding was that regardless of highly positive survey results, there were several ineffective aspects of the PA, which did not foster psychological safety, possibly causing uncomfortableness in for both the leader and the subordinate. Furthermore, our findings suggest that Model II behaviors such as active-empathetic listening plays a larger role in the interaction when psychological safety is low. It is important to note, as previously stated by Meyer and colleagues (2019) that the coded behaviors are not proof of leader or subordinate values, but they provide a way of linking values and observable actions.

Overall Findings from Survey Data

There was a positive overall tendency in the survey data for both leaders and subordinates. By summarizing the data (see Table 2), we found that all subordinates reported high levels of both self- and other-psychological safety (S-PS and O-PS). However, there was a slight tendency of subordinates having higher degrees of self-psychological safety. All subordinates reported very high levels of trust (T) in the leader, with little variation between the different subordinates. All statements were either strongly agreed or agreed to, with a single exception of one subordinate only rating 'kind of agree' on two out of seven trust-related statements. Additionally, there was a close correspondence between the leader's and subordinate's internal shared reality (SR), yet the involved leaders reported a slightly higher level of shared reality. Three out of five leaders rated their active-empathetic listening (AEL) skills higher compared to the ratings given by their respective subordinate, while the two remaining leaders gave lower ratings than their subordinates did. Furthermore, on statements regarding the subordinate evaluation (SE), all subordinates were rated between 6 and 7, meaning that the leaders 'agreed' to 'strongly agreed'. Lastly, the survey data indicate positive performance appraisal perceptions, where both leaders (PAR-L) and subordinates (PAR-S) rated highly.

Table 2*Overall survey results*

Leader questionnaire		Mean	Subordinate questionnaire		Mean
PA1	AEL	4.75	PA1	AEL	5.88
	PAR-L	N/A		PAR-S	6.8
	SR	N/A		SR	6.33
	SE	6		O-PS	6.75
			S-PS	7	
			T	6.5	
PA2	AEL	6.13	PA2	AEL	5.87
	PAR-L	6.5		PAR-S	6
	SR	7		SR	6.33
	SE	7		O-PS	6
			S-PS	6.8	
			T	6.25	
PA3	AEL	5.5	PA3	AEL	5.1
	PAR-L	5		PAR-S	6.8
	SR	6.33		SR	6.33
	SE	6.8		O-PS	6.25
			S-PS	7	
			T	6.87	
PA4	AEL	6.75	PA4	AEL	5.12
	PAR-L	6.5		PAR-S	6.4
	SR	7		SR	6.66
	SE	6.9		O-PS	5.25
			S-PS	6.4	
			T	6.62	
PA5	AEL	4.75	PA5	AEL	5.6
	PAR-L	6		PAR-S	6.6
	SR	6.33		SR	5.66
	SE	6.2		O-PS	5.25
			S-PS	6	
			T	6.25	

Overall Coding Findings from the Performance Appraisals

In the following section, the main findings from the PAs will be presented. This is related to the different occurrences and the combinations of coded utterances made by both leaders and subordinates. The findings will be presented thematically.

Model I and Model II Behaviors in Leaders

Our overall findings indicate that leaders conducted more Model II than Model I behavior (see Table 3). Especially in the dimension ‘Advocacy’ where the leader states own beliefs, gives reasons for these beliefs and provides reasons for agreement or disagreement with subordinate’s opinion, the leaders showed a substantial larger amount of Model II behaviors. Further, the leaders showed great amounts of active-empathetic listening behaviors such as offering praise and support, giving confirmative cues of listening, expressing empathy and understanding and assuring that what the subordinate said will be remembered.

Table 3

Prevalence of different codes related to Model I and Model II behaviors in leaders

Dimension	Indicator	Overall	
		Model I	Model II
Advocacy	1.1 Discloses own belief	80	119
	1.2 Provides grounds for own problem belief	89	122
	1.3 Critiques/comments other's point of view	4	156
Inquiry	2.1 Inquires into other's beliefs	24	68
	2.2 Checks for understanding	12	29
	2.3 Explores other's reaction to own beliefs	42	24
Collaborative planning	3.1 Establish common ground	26	28
	3.2 Collaboratively plans for next step	45	12
	3.3 Fosters shared responsibility	2	6
Active empathetic listening	4.1 Positive active and empathetic listening	2	159
	4.2 Neutral active and empathetic listening	1	541
	4.3 Expressing empathy	4	32
	4.4 Remembering	10	8

However, some Model I behaviors stand out as more prevalent than their Model II counterpart, such as exploring the other’s reaction to own belief under the dimension ‘Inquiry’, where the leader showed 57% more Model I behaviors than Model II behaviors. Further, in the dimension of ‘Collaborative planning’, our findings suggest a general tendency of leaders conducting more Model I behaviors, especially when planning for next steps in a process.

Leaders' Trust-Building Behavior

In the PA, trust was measured by looking for statements indicating cognitive trust shown by subordinates. Cognitive trust, as stated, relates to the subordinate belief about ability, reliability, fairness and integrity of the leader (Dirks & Ferrin, 2002; Kim et al., 2018). We found that the subordinates showed some indications of cognitive trust through valuing the leader's opinion and integrity (see Table 4). However, we were unable to measure affective trust in the PAs, which relates to the emotional bonds between leader and subordinate (Kim et al., 2018). It seems as affective trust cannot be clearly measured through coding conversations, as it relies on a feeling that is not as easily expressed as cognitive trust may be.

Table 4

Prevalence of different codes indicating trust, psychological safety, and perceived shared reality in subordinates

Dimension	Indicator	Overall	
		No	Yes
Self-psychological safety	6.1 Willingness to take interpersonal risk	56	204
Other-psychological safety	6.2 Not afraid of hurting the leaders' feelings	27	136
Trust	7.1 Value leader's opinion and integrity	2	37
Shared reality	8.1 Agreeing with what the leader is saying	21	246

Active-Empathetic Listening Behaviors in the Performance Appraisal

As leaders were asked to rate their own level of active-empathetic listening, it allowed us to compare the ratings to active-empathetic listening behaviors observed in the PA. All the leaders exhibited different levels of listening behaviors, from merely answering "Mhm" to expressing empathy and encouragement during the PA. Table 5 shows that there is little correspondence between self-ratings and observed listening behavior, thus indicating discrepancies between leaders' espoused theory and theory-in-use. For instance, one of the leaders with the lowest self-rating (PA1) had the highest number of active-empathetic listening codes, even when neutral active-empathetic listening codes were removed. This removal was done due to the sheer amount of neutral active-empathetic listening codes, leading to clutter in the data, and will be elaborated shortly. On the contrary, the other leader with the lowest self-rating (PA5) had the lowest number of active-empathetic listening codes when neutral codes were removed.

Table 5*Active-empathetic listening (AEL) survey data and code prevalence*

	Leader AEL survey data mean	Subordinate AEL survey data mean	AEL codes	AEL codes (removing neutral AEL)
PA1	4.75	5.88	212	56
PA2	6.13	5.87	170	42
PA3	5.50	5.10	136	56
PA4	6.75	5.12	104	34
PA5	4.75	5.60	117	10

The Interaction Between Model II and Psychological Safety

Our analysis suggests that during the PA conversations both Model II behavior and expressed psychological safety fluctuated during the conversation, something in which was not captured by the positive survey results. Nonetheless, we found that 74.3% of all expressions of self-psychological safety, and 95.4% of other-psychological safety came following leaders exhibiting Model II behaviors. However, after finding out that 18.8% of all coded statements was coded as neutral active-empathetic listening such as “Yes” and “Mhm”, and that there was no pattern to what kind of statement that would precede or follow that code, we decided to remove all neutral active-empathetic listening codes which preceded expressions of psychological safety. When removing the code of neutral active-empathetic listening, we found that only 41.7% of the expressions of self-psychological safety, and 54.6% of other-psychological safety expressions followed a Model II statement. Combining this notion to the survey data results, indicating that subordinates were slightly lower on other-psychological safety, one may interpret this as Model II behaviors such as active-empathetic listening playing a larger role in the interaction when psychological safety is low.

However, we also found that Model II statements from leaders led to the subordinate responding by undermining the severity of an issue 68.2% of the time (‘non’-self-psychological safety). Further, 89.3% of codes indicating the subordinate tiptoeing around their leader’s feelings (‘non’-other-psychological safety), did also follow a Model II statement. Nonetheless, when removing the code for neutral active-empathetic listening, we found that only 15.9% of statements of undermining, and 28.6% of statements of inability to critique, followed leaders’ Model II statements. These numbers suggest a slight pattern between psychological safety and Model II among the participants, as there are small differences between

the number of psychological safety statements and ‘non’-psychological safety statements following a Model II utterance by leaders.

To further investigate the relationship between psychological safety and Model II in each of the conversation, we decided to look into the occurrence of Model II indicators in relation to Model I indicators, and indications of psychological safety in relation to ‘non’-psychological safety (Table 6).

Table 6

Model II and psychological safety ratios

	Model II ratio	Psychological safety ratio
PA1	3.2	6.9
PA2	1.6	2.3
PA3	1.6	8.2
PA4	1.6	7.1
PA5	0.8	2.1

Note. Ratio was calculated by dividing Model II utterances (excluding neutral active-empathetic listening) by Model I utterances, and psychological safety expressions by ‘non’-psychological safety expressions in each conversation.

Our sample suggests associations between Model II and self-other psychological safety, especially when examining the two extremes of our dataset, PA1 and PA5 (see Table 6). The psychological safety ratios of PA3 and PA4 may be seen in light of the fact that we did not distinguish the degree of vulnerability in the coded psychological safety utterances. Moreover, leaders engaged in ‘pseudo’ Model II behaviors, for example pseudo inquiry, which could be starting a sentence with genuine Model II inquiry into subordinates' belief but ending with a loaded Model I question. This may give the subordinate the impression that they are in a safe and open space, although the loaded question is still present. This will be explored further in the second part of our analysis.

Discrepancies in Observed Behavior and Survey Ratings

As shown in the active-empathetic listening section, there was little correspondence between self-ratings and the observed listening behavior, thus indicating that leaders were not completely aware of their listening behavior. In our sample, three out of five leaders had given themselves high ratings, yet this was not manifested in active-empathetic listening behavior. Subordinate ratings have been found to be more accurate, thus possibly indicating an overconfidence in some of

the leaders, which is to be discussed later (Sinnema et al. 2015). On the contrary, one of the leaders with the lowest self-rating (PA1) had the highest number of active-empathetic listening codes, even when neutral active-empathetic listening codes were removed. In this PA, the subordinate had given their leader a high rating, corresponding with the high number of Model II behavior. To support these findings, we further draw on Sinnema and colleagues (2015) findings that leaders who underestimate themselves are considered more effective and trustworthy. Further, the subordinate in PA1 had a lower psychological safety ratio than PA3 and PA4, yet their leaders had a lower Model II Ratio. Regardless of our small sample, it is interesting to speculate whether lower confidence in the leader combined with a high degree of Model II makes the subordinate feel uncomfortable.

Across all five dyads, there were significant discrepancies regarding the performance evaluation given in the survey and the actual feedback given in the interview. As mentioned, all subordinates were rated highly, but there was little manifestation of these ratings in the statements given by the leader in the PA.

Lastly, when comparing the leaders' and the subordinates' perception of shared reality to the number of agreements, inquiry, and performance evaluation utterances encountered in the transcripts, we see that the number of coded utterances is remarkably low. A question that arise is whether the parties can reach a shared reality, when there is little exploring and sharing of perspectives. We will return to this in the discussion.

Overall Findings from within the Performance Appraisals

When exhibiting a more in-depth analysis to the transcripts and the codes, several important notions about the interaction between the leader and the subordinate were revealed. In the following part, qualitative findings from the analysis will be presented. The headings represent each of the main findings from the interactions as they unfolded in the PA.

Leaders' Lack of Open Inquiry

As earlier stated, leaders showed 57% more Model I behaviors than Model II behaviors when exploring other's reaction to own belief under the dimension 'Inquiry', and this was mainly done through loaded questions. These questions often encouraged the subordinate to change their belief about the discussed matters. An example of this is:

Excerpt 1

*Right, but, but it may be that someone had thought that “ok, yes, but, but, do you, do you” *laugh*, well, that if you had thought it through, then you would not have asked that question, you know? (PA1)*

In excerpt 1, the leader asks a question, but in a way that strongly encourages the subordinate to agree with the leader’s view of the matter. The leader does not inquire into what the subordinate thinks of what was said in an open way. Instead, by ending with “you know?”, the leader seeks agreement and the subordinate may feel they have to alter their view. Overall, there were moderately few codes for the ‘Inquiry’ dimension, with one of the PAs having merely one coded Model II inquiry utterance.

Off-topic inquiry. We saw patterns of leaders inquiring more on parts that were off-topic than when subordinates displayed a need for inquiry into topics involving psychological safety, e.g. expressing needs or being honest about aspects lacking in their work life. An example is shown in excerpt 2 where the subordinate expresses that they question their own ability and would like more help. The subordinate is showing both self-psychological safety by admitting to own weakness and other-psychological safety by asking the leader for help. Nonetheless, they are met with the leader asking about formalities instead of inquiring into why the subordinate questions themselves or showing any signs of registering the subordinate’s vulnerability.

Excerpt 2

*Subordinate: And then it’s the HR related... Yes, it is taken, yes, and it is the first step in terms of employer brand. We’ve taken in student projects and an internship, I think that is very positive, eh ... I question my own ability *laugh* to follow up, I would like a bit more help on that. Eeeh ... It will be fine ... Especially when it is demanding thematic, and it’s also therefore I ... I felt that it was good that I got a chance when we had it at [gathering]*

Leader: Yes. Ehm ... internship, who is that?

Subordinate: They will be here in April

Leader: From where?

Subordinate: They are from [university] (PA2)

Leaders' Model I Collaborative Planning

Further, in the dimension 'Collaborative planning' we found that in general the leaders conducted more Model I behaviors, especially when planning for next steps in a process. In these encounters, the leaders often pushed their own solution and further repeated and elaborated their own solutions. An example is:

Excerpt 3

We'll start there, because, because then, for then I think that, and what to do, what can I say, courses, and to build competence, whether it is writing advertisements or if it is other parts of the process, that would be very good, right? (PA1)

In this example the leader talks in a factual way by stating that they both should start a process at a certain point. Further, the leader states their own opinion, but instead of asking the subordinate their opinion on the matter, or whether they agree, the leader ends with a loaded question, and thereby pushing their own belief onto the subordinate.

Active-Empathetic Listening as Model II behavior

By examining the interactions in the PA, we found that only three out of the four codes concerning active-empathetic listening may be seen as Model II behavior: positive active-empathetic listening, expressing empathy, and remembering. Nonetheless, our analysis suggests that the indicator neutral active-empathetic listening ("mhm" or "yes") may be counterproductive. On the one hand, it can be seen as a form of inquiry as the leader confirms that they hear what is being said, while also giving room for the subordinate to express their opinions and feelings. On the other hand, we saw that leaders used this way of active-empathetic listening instead of explicit inquiry or showing support to the subordinate. The use of neutral active-empathetic listening instead of asking questions and giving empathetic statements were seen in all PAs. In excerpt 4, the leader and subordinate are talking about how it is to both give and receive constructive criticism. The subordinate is showing both self- and other-psychological safety by expressing a wish to be told when they are not performing. Instead of confirming this need or inquiring into why the subordinate wants this, the leader only shows neutral active-

empathetic listening, which seems to lead the subordinate away from the topic. The issue merely fades away.

Excerpt 4

Subordinate: Eh, but I want it, because I do not wish to get into a situation where I do **not** perform and does not know it

Leader: Mhm

Subordinate: So that is important to me. That there is room for errors there and that in a way, it is acknowledged that it is just to ...

Leader: Yeah yeah

Subordinate: I would like to know that

Leader: Yes

Subordinate: Sooner rather than later!

Leader: Yes

Subordinate: If there is anything

Leader: Yes

Subordinate: Eh, but I ...

Leader: Yes, that's good, well, it is just (PA1)

Our analysis further suggests that purely showing neutral active-empathetic listening may be a symptom of single-loop learning. This is sufficient for correcting, but not changing the organizational routine (Argyris, 1976, 1993). This will only postpone and not solve problems, as clearly shown in excerpt 5. When the subordinate expresses psychological safety by stating their wishes and needs, the leader simply responds with “Mhm” and “Yes” instead of inquiring into why or how these wishes may be met. The leader ends the interaction by expressing vaguely that this is an issue that they have to get back to, instead of discussing it there and then, which would be expected in a PA. This is also an example of Model I collaborative planning as there is no inquiry to the others' thoughts regarding solutions.

Excerpt 5

Subordinate: So that is one of the things that I would have appreciated... is to be... get a little more... Get challenged a bit... Both on what I do, who I am and on how I contribute, and ...

Leader: Mhm.

Subordinate: Get challenged a bit on, yes, both work environment wise...It is not a commitment, but that one is sort of... As things are now, one becomes very individual in the way things are done... I just gotta fix my stuff, like, sometimes it will be like that... Anyways, that you always have that hand you are talking about, that safe hand [talks about oneself in third person] "Oh, I can always come back here, and maybe someone will challenge me a bit" ...

Leader: Yes ...

Subordinate: So that is good.

*Leader: Let's... Yes, it will come later too, but we absolutely have to... *sighs*
I think we have to set up... hmm... some time... eeehm... regularly (PA2)*

Leaders Ensuring Psychological Safety

Several interesting patterns emerged in the PAs in relation to how the different leaders ensured psychological safety in the dyad. The subordinates scored highly on psychological safety in the surveys. However, we found few examples of leaders inquiring and exploring situations where the subordinates expressed vulnerability or a concern.

As previously discussed, leaders exhibited several 'pseudo' Model II behaviors such as pseudo inquiry in the PA conversations. An example can be seen in excerpt 6 where the leader starts by inquiring into the subordinate's belief, but before the subordinate gets a chance to answer, the leader follows up with a loaded question, steering the conversation in a certain direction. Therefore, this seems like an inquiry into the subordinate's belief, but in all it is Model I behavior.

Excerpt 6

(...) That position you got, that was actually a position for both you and [colleagues name]. Does the collaboration between you two work? Or are there any confusions concerning who has responsibility for whom, or does that work fine, do you think? (PA3)

The same tendency was present on the subordinates' side with regards to psychological safety, e.g. by starting with a psychological safety statement, yet quickly turning to brushing off the severity of the issue. Excerpt 7 shows a subordinate starting out by explaining why they did not give a top score on a self-rating in a spreadsheet, stating that they would have more development if they were faced with more complex issues. However, they quickly shift and instead focus on how the organization can be elevated, therefore focusing on everyone's need instead

of expressing their own need of improvement. In such cases, the leader usually did nothing to explore the ‘hint’ that was given, instead they simply went on to the next topic.

Excerpt 7

(...) I can say why I put 8 and not 10 there... It is simply because... If I were to meet more complex issues, of course I would have developed myself more... eeh... but eeh... it is... I have... I think that by elevating the organization as it is now, that is not a minus either... (PA3)

Nonetheless, some leaders were better at inquiring and supporting than others. Especially the leader in PA1, who scored themselves low in the survey concerning active-empathetic listening, but showed the most active-empathetic listening behaviors, was good at both supporting and inquiring into the subordinate’s view. In excerpt 8, the leader from PA1 demonstrates Model II behavior through both inquiry and active-empathetic listening, while the subordinate shows self-psychological safety by admitting shortcomings.

Excerpt 8

Leader: Do you think you have a good overview of, what can I say, [organization] as, as an organization?

Subordinate: I still believe that, I, I... I should have read more about the structure and how [organization] look internally...

Leader: Yes

Subordinate: Eh... Because it is... it has been... it is complex

Leader: Yeah yeah

Subordinate: And in the beginning it was very confusing... Now, it is getting better and better. I still believe that I can do more own research to find [organization] and unions, and all the things that are specific for, eh... Yes, it is complex, and it won’t stick immediately....

Leader: No no, I really understand you, it is... eh, I think that it is complex myself. There are many, many sides of the organization which are different. Employee groups and, what can I say, in that way a lot of different complexities, work structures, hm, Yeah... (PA1)

Counterproductive Laughter

We found that both leaders and subordinates tended to laugh in situations which could be interpreted as uncomfortable, e.g. situations with performance feedback or display of vulnerability. Leaders tended to laugh during or after providing feedback, while subordinates tended to laugh when talking about difficulties. When there was a situation of laughing, it seemed that it would prevent the parties to further understand the other's perspective, thus possibly preventing learning. Combined with the leaders' lack of inquiry, this could at times cause subordinates closing off and starting to talk about a different topic. In excerpt 9, a challenging issue regarding the subordinate's work-life balance is addressed. However, instead of the leader inquiring into why it is hard for the subordinate to say no to tasks, they instead tell the subordinate what they should do, and joins the laughter when the subordinate tries to 'cover' their uncomfortableness. The interaction soon steers off-topic.

Excerpt 9

Leader: Yes. And I believe maybe that here under time management, you should have a point concerning saying "no" more often.

*Subordinate: Yes *laughs**

*Leader: *laughs* Well, delimit what you are working on, with either saying "no" to yourself in terms of quality or saying "no" to new tasks. You have to judge that yourself, and we can have a running dialogue on it. But that something has to... something's gotta give ...*

Subordinate: Yes

Leader: Either you have to deliver with less quality than you usually do, or you have to deliver on fewer things.

*Subordinate: Yes. Yes, I will ... I just think it's hard to say ... It's just the word "no" *laughs**

*Leader: Yes *laughs**

[both laughing - then going off-topic] (PA2)

Giving Subordinates Feedback

Lastly, we found that leaders seemed to be struggling with giving performance related feedback. As mentioned, all leaders rated their subordinates highly in the survey. However, these scores were not reflected in the PA conversations. Most of the performance evaluation utterances were neutral, with leaders simply stating the observed such as "I see you do this...", very often based

upon a numerical evaluation form filled out by both parties. A richer feedback was not given unless there was a big discrepancy between the leader's and the subordinate's rating.

Discussion

The aim of this study is to uncover how the leaders' actions and inactions can explain or relate to uncomfortableness in both parties, and the level of perceived psychological safety in subordinates. Our analysis has revealed several important notions that helps us understand this with regards to Model I and II behavior, and psychological safety. In the initial part of the study, the research problem was broken down in two more specific research questions; in this chapter, these are to be discussed, along with theoretical and practical implications, limitations, and future research.

What Makes the PA a Potentially Uncomfortable Ordeal for both the Leader and the Subordinate?

In relation to our first research question, our analysis indicated several reasons to why the PA may be uncomfortable for both parties. Based on the findings from both the leader and subordinate survey, and the interactions in the PA conversation, we have identified several keynotes which could explain the potential uncomfortableness. Each of these will be presented in their own section.

Insufficient Use of The Performance Appraisal Framework

One of the overall apprehensions from the analysis was that leaders did not seem to fully know how to utilize the time at hand nor the 'framework' of the PA. This supports Meineke and colleagues' (2017) notion concerning the fact that the PA is negatively perceived by both parties due to limited understanding of the social dynamic in the PA. By strictly following a written guideline, the leaders will not be able to unravel the different issues and concerns subordinates may have. Further, and in line with Roberto (2002), it seems that numerical ratings are not sufficient in learning where there is good performance, and where there still is developmental potential. How can subordinates experience improvement when simply given a number on their current performance? How can leaders provide proper feedback if they are merely looking for congruence in their own and the subordinate's rating?

Lack of Relevant Inquiry

One of the most prominent findings uncovered is the leaders' lack of inquiry, or in other words, the lack of seeking information from the other. To better understand the implications of low occurrence of Model II inquiry, we draw on

findings from Schein (2013) suggesting inhibitory defensive routines, both within themselves and within the subordinates. In the current study, inhibitory defensive routines were manifested through e.g. loaded questions and counterproductive laughter. Schein's findings may help explain both the absence of Model II inquiry, and why some Model II inquiries turned into Model I, hence the leader defending own view. In terms of defensive routines in subordinates, it may help us understand why a significant amount of psychological safety utterances quickly turned to subordinate undermining the severity of the issue addressed (non-psychological safety). To support this, Leppänen and colleagues (2018) found that only inquiry can generate positive emotional expressions, which cannot be generated through merely advocacy.

Further drawing on Schein (2013), another possible explanation to why there was a lack of relevant inquiry may be the presence of a circular defensive routine. If the leader starts off with little inquiry, it may lead to the subordinate sharing less and less of their experience, which again leads to the leader inquiring even less, thus making both parties uncomfortable in the situation. This effect could be further strengthened by overconfidence in leaders, as their lack of self-awareness may inhibit them in questioning own behavior and belief (Moore & Healy, 2008). This may further lead them to not realize that their limited inquiry is the root of the problem, creating a negative loop. Overconfidence is to be discussed later in this chapter.

The 'Humming-Trap'

Neutral active-empathetic listening is essential part of the conversation since it gives the speaker room to be open, as the listener gives cues that they are paying attention. However, our findings indicated that an excessive use of utterances such as "Yes" and "Mhm" could be counterproductive and lead to the subordinate retracting their vulnerable or critical statement. We argue that it must be combined with other Model II behaviors in order to generate a positive outcome for subordinates, which further generate a positive outcome for leaders as it may create a positive loop. Considering what we found, there may be a presence of an inverted U-curve of the leader's use of neutral active-empathetic listening in relation to the subordinate's psychological safety. It is helpful up to a certain point, but beyond this it becomes counterproductive. This is supported by Mattias Heldner's finding that humming is essential, but if it turns absent, becomes

excessive, or comes in the wrong place, it may evoke insecurity in the other (Ashraf, 2014).

Poor Collaborative Planning

As of collaborative planning, leaders exhibited more Model I behavior. Especially the indicator ‘collaboratively plans for next step’ manifested more Model I behavior, thus considered as non-collaborative and typically entailing the leader pushing their own solution. This finding is further stressed by the notion of lack of inquiry: leaders outline their own view, does not inquire into the others’ view, and finish by pushing their own solution without checking for agreement. Earl and colleagues (2009) describe the situation of jumping to solutions that are not founded in evidence as ‘activity traps’, which are a waste of time and learning potential. It is important to remember the fact that all this usually happens in a very nice manner, hence not necessarily provoking any immediate reactions in the subordinate, thus, possibly explaining the positive survey results. However, when the leader is pushing their own solution without inquiring into the subordinate’s view, the subordinate may feel they have to alter their view. This may lead to an uncomfortable situation where the subordinate does not have the courage to be open and honest: they are no longer equal parties in the conversation. Taking a more holistic stance, this may further be a source of organizational inertia, where issues are postponed instead of solved.

Inadequate Feedback

As most of the performance evaluation utterances were neutral and given in relation to a numerical evaluation from a form, we conclude that the feedback given in the PAs was inadequate. We question the lack of inquiry in relation to giving feedback, and further how the leaders could give adequate feedback if they are mainly looking for congruence in ratings from a numerical form. What is the point of getting feedback if it will only be positive or negative when there is a big discrepancy? Further, how should the subordinates interpret neutral feedback without any indication of the leader’s standpoint in the matter? This lack of rich feedback could lead to lowered levels of organizational learning, as this relates to revising and reshaping knowledge the subordinate has assimilated (Gill, 2000b). If the subordinate does not get any indications in the PA of what they do right or wrong, they would not have any knowledge to revisit or reshape.

Off-Topic and Laughter

In a significant number of encounters across the PAs, laughter tended to surface in situations of performance feedback from leaders and vulnerable moments in subordinates, two ‘places’ in which there is learning potential for both parties. This could even be seen when giving positive feedback, sometimes even ebbing out in a sequence of jokes. One may interpret this as the leader finding it uncomfortable to provide praise, thus resulting in defensive routines (Schein, 2013).

Subordinates going off-topic may also be interpreted as a defensive routine as a result of the leader not inquiring into their belief. If they feel safer to talk about topics that do not concern themselves etc. As we also found that leaders were better at inquiring when off-topic, the subordinate may feel that what they initially were saying was not important and therefore stay off-topic. For both parties, staying off-topic will prevent them from both learning and problem-solving, and their perceived value and the organizational value of the PA may fade (Bowen & Ostroff, 2004).

Is there Really a Shared Reality?

As shown in the analysis, all leader-subordinate dyads were congruent in their reporting on shared reality. It is reasonable to assume that Model II behaviors contribute to the obtainment of a shared reality between parties, due to inquiry followed by sharing of beliefs. It is a part of human nature to strive for common grounds (Echterhoff, 2012), and with little evidence of Model II behavior, especially a lack of inquiry, the high level of shared reality may be understood as a product of the unconscious process of simply avoiding topics of possible disagreement. Several studies (e.g. Glauser, 1984; Higgins, 1992; Schmalbach et al., 2019) have found that individuals often tend to delay, distort, or simply not communicate information critical to organizational learning and performance. With this notion combined with little Model II behavior, it is interesting to speculate what may come at play when the both parties genuinely believe that they are on the same wavelength, yet they still have left to explore each other’s ‘realities’. In relation to RQ1, the lack of exploration of others’ realities may be important as it may cause misunderstandings or disagreement, in which the two parties are not aware of. This may be a source of conflict and uncomfortableness outside the PA.

All the subordinates in the current study were middle leaders and in positions of great autonomy, thus there may not be much of a power imbalance

between the parties. On the contrary, if our sample were to include shop-floor subordinates and their leaders, it is reasonable to assume that there may be more of a power imbalance between the parties. Thus, when such power imbalance is at play, one may see more of the subordinate adopting the view of the leader (Glauser, 1984), not only in the PA, but in general. There may be less Model II behavior, less psychological safety, combined with more agreement in between the parties. Arguably, this may further cause single-loop learning.

Potential Overconfidence in Leaders

To understand the discrepancies in the observed and reported data, we draw on research on overconfidence, or the excessive faith in yourself knowing the truth (Moore & Healy, 2008). Bazerman and Moore (2012) suggest cognitive dissonance as one of the psychological processes causing overconfidence. The survey data indicate that the subordinates reported very high levels of trust, regardless of leader exhibiting a lot of Model I behavior. A possible explanation may be that high levels of trust may further cause cognitive dissonance in the leader due to clashing roles. Further, Alvesson and Gjerde (2020) suggest that middle managers (as the leaders in the current study were) may find themselves in a ‘sandwiched’ position between top management and subordinates. On the one hand, they want to meet the subordinate’s expectations, and on the other hand, they want to meet the expectations of the organization. This potential cognitive dissonance may further cause overconfidence and Model I behavior, making the subordinate feel uncomfortable. The leader may have too much confidence in what they know. Overconfidence may also help us understand why the participating leaders had no reluctance towards the project.

On the contrary, one of the leaders had rated themselves lower than the others (PA1), yet they had the highest Model II ratio. A question that arise from this is whether it is possible for the leader to become too open and interested. Further, we question whether low self-rating indicates low confidence, and if this makes the subordinate tiptoe more around the leaders’ feelings.

How does Leaders Foster Psychological Safety when Conducting Performance Appraisals?

When discussing our first research question, we have highlighted different aspects of why the PA may be uncomfortable. In relation to our second research question, we will discuss how leaders may face this uncomfortableness through

fostering psychological safety. Based on our theory and analysis, we have identified several actions leaders may engage in to foster psychological safety in the PA setting. Each of these will be discussed in the following section. As stated initially in this study, psychological safety is related to perceptions of consequences of taking an interpersonal risk, in contexts such as the workplace (Edmonson & Lei, 2014). For subordinates to do this, leaders must ensure a safe and open space, free of judgement. This may, however, be easier said than done. The following section will provide a discussion of actions that leaders may take to promote this safe space.

The lack of inquiry has been highlighted as one potential cause to why subordinates may feel uncomfortable during the PA. Inquiry is arguably the most important aspect of Model II behaviors, as it explores the other's view and promotes subordinates' voice (Detert & Burris, 2007; Leppänen et al., 2018; Schein, 2013). Nonetheless, it is critical to note that in order for the inquiry to positively affect psychological safety, it must be open, and free from loaded questions that steers the conversation in a certain direction. By that, the subordinate may still feel like an equal part of the conversation, and speak their mind without reluctance, further increasing the learning potential of the PA.

As previously mentioned, we question whether there might be a misconception to whether neutral active-empathetic listening is enough to give the other the impression that one is engaged in the conversation. We further argue that the use of only humming and "Yes" can be interpreted as the other not paying attention, or a desire to move the conversation forward, as seen in many interactions across the PAs. Alone, these utterances may be enough for maintaining status quo, however, the other Model II behaviors, especially inquiry, is necessary for challenging it (Argyris, 1976; Leppänen et al., 2018). With an excessive use of humming from leaders, the conversation may be at risk of becoming more of a monologue, thus limiting the turn-taking behavior between the parties. This may be detrimental as conversational turn-taking has been identified as one important aspect of psychological safety in teams (Duhigg, 2016). The current study gives indications that the same may be at play in leader-subordinate interaction.

Moreover, trust-building is important to further promote psychological safety in the workplace, as the subordinate is willing to be vulnerable without concern of criticism or humiliation (Carmeli et al., 2009; Tan & Lim, 2009). Firstly, leaders need to genuinely inquire more into subordinates' views and assumptions, both by asking them their opinion, and exploring their reaction to own

statements. Further, leaders need to improve giving feedback, both positive and negative, which will further increase perceived trust (Robinson, 2009). The subordinates do not learn much from leaders' neutral expressions, which further may weaken organizational learning (Gill, 2000b). Increased inquiry and checking for agreement can improve the feedback leaders give, as they are more aware of an eventual shared reality. This may reduce the pattern we saw in the PAs, where leaders only gave richer feedback when there were significant discrepancies between ratings. If the leader and the subordinate are on the same page, the leaders may give feedback not only when there are big discrepancies, but also in relation to the everyday behaviors of subordinates.

Additionally, leaders should be more aware of their collaborative planning behaviors. When next steps are planned with the use of loaded questions and the pushing of own beliefs, the parties may get caught in an activity trap, thus wasting both time and learning potential (Earl et al., 2009). In addition, it may lead to the subordinate feeling that they are not equal in the conversation, as the leader essentially is stating that "We will do that, right?" instead of asking for the subordinate's view on possible next steps or whether they think it is a good plan. This may hinder the subordinate in using voice, which further lessens their experience of psychological safety. The leader should show dedication, and ask explicitly what the subordinate thinks, even when they may think they are on the same page and with a shared reality.

General Discussion

In the conversations, the overall tendency was success in avoiding conflict, thus maintaining the leader-subordinate relationship. However, the lack of inquiry may have come at the cost of exchanging information critical to learning and challenging status quo. If the world of business was stable and not requiring organizations to constantly adapt, this would have been sufficient. However, today's world of business is changing at a fast pace, hence highlighting the need of rapid information exchange. The PA offers a unique opportunity for exchanging deeper information, solving problems, and building the leader-subordinate relationship (Brown et al., 2010; Kuvaas, 2006; Roberts, 2002). Nonetheless, the current study offers insights to how small behavioral nuances (of action and inaction) may be detrimental to the effectiveness of the PA. All participants reported positive perceptions, however, this may not necessarily correlate to the

actual effectiveness of the PA. By effectiveness, we refer to the leader's ability to evoke psychological safety in the subordinate, thus making the PA a safe space where information can be mutually shared.

Several ineffective behaviors were found in an overall positive sample. When finding this many unfavorable actions among 'content' leaders and subordinates, one can only imagine the scenario with dyads experiencing lower levels of psychological safety and trust. We restate Kuvaas' (2006) suggestion that the traditional framework of the PA may have negative effects on subordinates low in intrinsic motivation which actually are in need of benefitting from the PA. In line with other researchers (Brown et al., 2019; Miller & Gordon, 2014; LeFevre & Robinson, 2015), we argue that a stronger focus on the *how* of communication, as the current study suggest, the PA could benefit a broader range of subordinates.

Furthermore, the longest tenure among the subordinates were two years, with one of the subordinates only being in the organization for two months. This is a relatively short time of employment, and therefore we question whether this could explain the positive responses. It may be that the subordinates still feel that the job is new and exciting, and they may not have experienced major conflicts yet (Owens et al., 2016). Further, it may be that they want to impress their new leader. Moreover, as the longest dyadic tenure was two years, it may affect the leaders' attitudes towards their subordinates, as they are not familiar with each other, and therefore may be reluctant to push boundaries or to dig too deep. With this in mind, it is also relevant to think of the younger generation of workforce, where a more rapid change of workplace is predicted (Gallup, 2016). This notion, combined with the increasing pace of change in the business environment, further highlight the need of effective communication, double-loop learning and the fostering of psychological safety. If turnover becomes inevitable, organizations will at least benefit from maintaining important information within.

From what we have found, it is fair to wonder if having the traditional PA is worth both the time and effort. An instant solution would be to remove the PA as a whole, as several organizations already have done (Cappelli & Tavis, 2016). However, we argue that if organizations are aware of the behavioral nuances as identified in this study, the PA can be beneficial in the long run. With leadership behavior as a strong predictor of team-level psychological safety (Tynan, 2005), the PA may pose positive synergies beyond the leader-subordinate dyad. This may further stimulate for organizational learning (Edmondson, 1999). Moreover, by

keeping the traditional PA, rather than replacing it with real-time feedback, the leader and subordinate may have a better foundation for solving problems. It may take some time to get to the bottom of an issue and resolve it, especially more fundamental ones (Roberts, 2002). We propose how this can be done while avoiding the leader's behavior causing an activity trap (Earl et al., 2009).

Furthermore, Nishii and Wright (2007) argue that it is the subordinate perceptions and not the specific practices that influence outcomes of the PA. These PA perceptions are important to aspects such as organizational commitment, trust in management, and individual performance (Kuvaas, 2006). However, survey measures are not very precise when measuring perceptions, as the subordinates are rating their agreement of statements, and not stating how they actually feel themselves (Arnulf et al., 2018). Nonetheless, as these perceptions of leaders are expressed also in the PA conversation, combined with the notion that it takes time to open up about issues, and that it is the only 'deep' encounter subordinates have with HR practices, this should be enough to keep the traditional PA, but with some alterations. Further, this may also affect the subordinates' perception of the HRM system as a whole (Bowen & Ostroff, 2004).

Theoretical and Practical Implications

Argyris and Schön's framework as presented in the current study, has mainly been applied to problem-solving conversation in an educational setting with school leaders (Le Fevre & Robinson, 2015; Robinson et al., 2020). To our best knowledge, the current study is the first to apply Argyris and Schön's (1974) theory to experienced leaders in a business setting. Our results are congruent with previous findings (Meyer, 2019; Robinson et al., 2020), thus making the framework both a good tool for leadership practice and development, and for developing guidelines and practices for the PA. As opposed to similar studies (Asmuß, 2013; Meinecke et al., 2017; Robinson et al., 2020), we incorporate the dimension of subordinates' psychological safety as a possible outcome of leader Model II behavior, thus proposing an expansion of Argyris and Schön's framework and a more practical contribution to the *how* of psychological safety.

Most research conducted with regards to PA has been done using quantitative methods looking into employee perceptions (Brown et al., 2019). With a focus on leader behavior and the authentic leader-subordinate interaction, we provide insight to *how* the PA may be exploited in a way that may generate positive

outcomes for leaders, subordinates, and the organization as a whole. Further, earlier research has had an emphasis on the subordinate perceptions (Kuvaas, 2006; Nishii & Wright, 2007); our findings suggest that perceptions may not be enough when determining the value of the PA, it should be combined with the effective practices as suggested in this study. As ineffective practices are often done in a very nice manner, single-loop learning way still be at play, even though the subordinate has positive perceptions.

Limitations and Future Research

All participants were very positive towards the project and chose to participate due to that, thus representing a major bias in our study. The majority potential participants found the thought of having their PA recorded uncomfortable, hence limiting the number of complete dyads. In addition, the data-collection was cut short due to Covid-19. In contrast to our homogenous sample, it would be interesting to get a peek into the PAs of leaders not as confident in their role as both a supporter and critic. It is reasonable to think that low trust in leader, low psychological safety, and low leader confidence may have been valid reasons for people to decline our invitation to participate. However, this study may be the first step in the right direction in reaching these people. We encourage future research to take the next step, trying to make the participation of these kinds of studies nonthreatening, and instead seen as a learning opportunity.

Further, all of the subordinates in the study were middle managers. We believe that if we would gain access to subordinates at the shop-floor, we would have gotten more diverse results, due to the potential power imbalance (Asmuß, 2013). Furthermore, with most of the participants being relatively new to the organization, it may also be interesting to further uncover how being new could affect the ability to take interpersonal risk. Thus, future research should strive to investigate a more diverse group of participants, in terms of aspects such as age, dyadic tenure, employee performance, trust, and motivation. Further, one of the questions that remain unanswered due to our small sample is whether there is such thing as too much Model II behavior, and if low confidence in leaders may have a negative impact on psychological safety in subordinates.

Due to the small number of participants we question the generalizability of this study. On the one hand, with triangulation, authentic PAs, and the exploration of processes that are mainly unconscious, this study provides better grounds with

regards to generalizability. On the other hand, our intention is to provide an in-depth understanding of leader and subordinate behavior in the PA, and not to create a generalizable theory of those behaviors.

Regarding the codebook, there was only one code for each of the two psychological safety dimensions. This led to several different utterances showing psychological safety being given the same code. For instance, some subordinates used shorter time to express a big uncertainty, while others needed a longer dialogue where they expressed psychological safety gradually. The large utterance would be equally weighted as a psychological safety expression as the individual small utterances would. Future research should work on extending the operationalization.

In line with the preferences of the participants, the survey was distributed in paper format. Yet anonymity was a major concern, it may be that the participants were primed to answer the survey was answered in a socially desirable way, thus partly explaining the very positive survey results and not necessarily reflecting what happened in the PA. Additionally, we restate Arnulf and colleagues (2018) findings concluding survey scores filter out attitude strength. Based on that, we encourage open-ended survey questions in the measurement of perceptions.

Lastly, with shared reality representing only a small fragment of this study, we argue that it may be interesting to look deeper into how shared reality may be formed, and what implications it may have when two parties genuinely believe they share reality, when they actually do not. As much as we identified a substantial lack of inquiry, and if these findings are true in a wider setting, we believe this may be the case in many organizations.

Conclusion

The leaders managed to maintain the relationship with their subordinates but did not manage to strengthen it. They were polite and advocated own views, but failed in inquiring into the subordinates' beliefs, and therefore hindered the exchange of perspectives. Further, they showed interest by showing active-empathetic listening, but with an excessive amount of humming, they missed out on many opportunities for learning and growth. However, both leaders and subordinates left the PA feeling satisfied.

Our investigation aimed at answering how leaders' action and inactions relate to level of psychological safety in subordinates during the annual performance appraisal interview. With the narrative above, we stress the need of the current study. During conversations, we tend to be unaware of the nuances of how we say things, and how that affects the consequences, especially in a bigger picture. In the current study we have linked the leader's exhibition of Model II behavior to psychological safety and how this interaction may be further linked to learning, both individual and organizational, thus making the performance appraisal more effective. On the contrary, Model I behavior may only be sufficient in the short run and to maintain status quo.

Our findings provide understanding to why several organizations move away from the traditional performance appraisal; the leaders are not aware of their behavioral nuances, nor do they know how to properly utilize the time and framework. It is important that leaders are aware of the potential of a performance appraisal, and we have highlighted several aspects of improvement in leader behavior. In sum, our recommendation is for organizations to keep the traditional performance appraisal, but to be aware of the communicational aspect.

References

- Argyris, C. (1976). Single-Loop and Double-Loop Models in Research on Decision Making. *Administrative Science Quarterly*, 21(3), 363-375.
<https://doi.org/10.2307/2391848>
- Argyris, C. (1993). Education for leading-learning. *Organizational dynamics*, 21(3), 5-18.
- Argyris, C. (2010). *Organizational Traps*. Oxford University Press.
- Argyris, C., & Schön, D. (1974). *Theory in practice: Increasing professional effectiveness* (Jossey-Bass higher and adult education series).
Jossey-Bass.
- Arnulf, J. K., Larsen, K. R., Martinsen, O. L., & Egeland, T. (2018). The failing measurement of attitudes: How semantic determinants of individual survey responses come to replace measures of attitude strength. *Behavior Research Methods*, 50(6), 2345-2365. <https://doi-org.ezproxy.library.bi.no/10.3758/s13428-017-0999-y>
- Ashraf, A. F. (2014, April 23). Så viktig er “mhm” og “jaha” for samtalen. *Aftenposten*.
<https://www.aftenposten.no/kultur/i/L410p/saa-viktig-er-mhm-og-jaha-f-or-samtalen>
- Asmuß, B. (2008). Performance appraisal interviews: Preference organization in assessment sequences. *Journal of Business Communication*, 45, 408–429. <https://doi.org/10.1177/0021943608319382>
- Asmuß, B. (2013). The emergence of symmetries and asymmetries in performance appraisal interviews: An interactional perspective. *Economic and Industrial Democracy*, 34, 553–570.
<https://doi.org/10.1177/0143831X13489045>

-
- Bazerman, M., & Moore, D. A. (2012). *Judgment in Managerial Decision Making*, (8th ed.). Wiley.
- Bell, E., Bryman, A., & Harley, B. (2019). *Business research methods* (Fifth ed.). Oxford University Press.
- Blair, E. (2015). A reflexive exploration of two qualitative data coding techniques. *Journal of Methods and Measurement in the Social Sciences*, 6(1), 1-14. <https://doi.org/10.2458/v6i1.18772>
- Bowen, D., & Ostroff, C. (2004). Understanding HRM-firm performance linkages: The role of the “strength” of the HRM system. *Academy of Management Review*, 29, 204–221.
<https://doi.org/10.5465/amr.2004.12736076>
- Brown, M., Hyatt, D. and Benson, J. (2010). Consequences of the performance appraisal experience. *Personnel Review*, 39(3), 375-396.
<https://doi.org/10.1108/00483481011030557>
- Brown, T. C., O’Kane, P., Mazumdar, B., & McCracken, M. (2019). Performance management: A scoping review of the literature and an agenda for future research. *Human Resource Development Review*, 18(1), 47–82. <https://doi.org/10.1177/1534484318798533>
- Cappelli, P. & Tavis, A. (2016, October). *The Performance Management Revolution*. HBR.
<https://hbr.org/2016/10/the-performance-management-revolution>
- Carmeli, A., Brueller, D., & Dutton, J. E. (2009). Learning behaviours in the workplace: The role of high-quality interpersonal relationships and psychological safety. *Systems Research and Behavioral Science: The Official Journal of the International Federation for Systems Research*, 26(1), 81-98. <https://doi.org/10.1002/sres.932>
-

-
- Castro, D. R., Kluger, A. N., & Itzchakov, G. (2016). Does avoidance-attachment style attenuate the benefits of being listened to?. *European Journal of Social Psychology, 46*(6), 762-775.
<https://doi.org/10.1002/ejsp.2185>
- Detert, J. R., & Burris, E. R. (2007). Leadership behavior and employee voice: Is the door really open?. *Academy of Management Journal, 50*(4), 869-884. <https://doi.org/10.5465/amj.2007.26279183>
- Dick, P. (2004). Discourse analysis. In C. Cassell & G. Symon (Eds.), *Essential guide to qualitative methods in organizational research* (pp. 203-213). Sage Publications.
<http://dx.doi.org/10.4135/9781446280119.n17>
- Dirks, K., & Ferrin, D. (2002). Trust in leadership: Meta-analytic findings and implications for research and practice. *Journal of Applied Psychology, 87*(4), 611-628. <https://doi.org/10.1037/0021-9010.87.4.611>
- Duhigg, C. (2016). What Google learned from its quest to build the perfect team. *The New York Times Magazine, 26*, 2016.
- Dysvik, A., & Kuvaas, B. (2011). Intrinsic motivation as a moderator on the relationship between perceived job autonomy and work performance. *European Journal of Work and Organizational Psychology, 20*(3), 367-387. <https://doi.org/10.1080/13594321003590630>
- Earl, L. M., Katz, S., & Ben Jaafar, S. (2009). *Building and connecting learning communities: The power of networks for school improvement*. Thousand Oaks, CA: Corwin Press.

-
- Echterhoff, G. (2012). Shared-reality theory. In Van Lange, P. A., Kruglanski, A. W., & Higgins, E. T. (Eds.) *Handbook of theories of social psychology: volume 2* (Vol. 2, pp. 180-199). Sage Publications.
<http://dx.doi.org/10.4135/9781446249222.n35>
- Echterhoff, G., Higgins, E. T., & Levine, J. M. (2009). Shared reality: Experiencing commonality with others' inner states about the world. *Perspectives on Psychological Science*, 4(5), 496-521.
<https://doi.org/10.1111/j.1745-6924.2009.01161.x>
- Edmondson, A. (1999). Psychological Safety and Learning Behavior in Work Teams. *Administrative Science Quarterly*, 44(2), 350-383.
<https://doi.org/10.2307/2666999>
- Edmondson A. C. (2004). Psychological safety, trust, and learning in organizations: a group-level lens. In R.M. Kramer, & K.S. Cook (Eds.) *Trust and distrust in organizations: Dilemmas and approaches* (pp. 239-273). Russell Sage Foundation.
- Edmondson, A. C., & Lei, Z. (2014). Psychological safety: The history, renaissance, and future of an interpersonal construct. *Annu. Rev. Organ. Psychol. Organ. Behav.*, 1(1), 23-43.
<https://doi.org/10.1146/annurev-orgpsych-031413-091305>
- Fenniman, A. (2010). Understanding each other at work: An examination of the effects of perceived empathetic listening on psychological safety in the supervisor– subordinate relationship. (Doctoral Dissertation), George Washington University. (UMI Number: 3389636)
- Gallup. (2016). *How Millennials want to work and live*.
<https://www.gallup.com/workplace/238073/millennials-work-live.aspx>
-

-
- Gill, R. (2000b). Discourse analysis. In M. W. Bauer & G. Gaskell (Eds.), *Qualitative researching with text, image and sound, 1*, (pp.172-190). Sage Publications.
- Gill, S. (2000a). *The managers pocket guide to organizational learning* (Managers Pocket Guide Series). HRD Press.
- Glauser, M. J. (1984). Upward Information Flow in Organizations: Review and Conceptual Analysis. *Human Relations*, 37(8), 613–643.
<https://doi.org/10.1177/001872678403700804>
- Grint, K. (2007). What's Wrong with Performance Appraisals? A Critique and A Suggestion. *Human Resource Management Journal*, 3(3), 61-77.
<https://doi-org.ezproxy.library.bi.no/10.1111/j.1748-8583.1993.tb0031.x>
- Guest, D. E. (2011). Human resource management and performance: still searching for some answers. *Human resource management journal*, 21(1), 3-13. <https://doi.org/10.1111/j.1748-8583.2010.00164.x>
- Hardin, C. D. & Conley, D. C. (2000). A relational approach to Cognition: Shared experience and relationship affirmation in social cognition. In G. Moskowitz, (Ed.). *Cognitive Social Psychology* (pp. 3-18). Psychology Press.
- Higgins, E. T. (1992). Achieving 'shared reality' in the communication game: A social action that creates meaning. *Journal of Language and Social Psychology*, 11(3), 107-131. <https://doi.org/10.1177/0261927X92113001>
- Hsieh, H., & Shannon, S. (2005). Three Approaches to Qualitative Content Analysis. *Qualitative Health Research*, 15(9), 1277-1288.
<https://doi.org/10.1177/1049732305276687>
-

-
- Hyde, K. F. (2000). Recognising deductive processes in qualitative research. *Qualitative Market Research*, 3(2), 82-89.
<https://doi.org/10.1108/13522750010322089>
- Itzchakov, G., Kluger, A. N., & Castro, D. R. (2017). I am aware of my inconsistencies but can tolerate them: The effect of high quality listening on speakers' attitude ambivalence. *Personality and Social Psychology Bulletin*, 43(1), 105-120.
<https://doi.org/10.1177/0146167216675339>
- Kim, T., Wang, J., & Chen, J. (2018). Mutual trust between leader and subordinate and employee outcomes. *Journal of Business Ethics*, 149(4), 945-958. <https://doi.org/10.1007/s10551-016-3093-y>
- Kuvaas, B. (2006). Performance appraisal satisfaction and employee outcomes: Mediating and moderating roles of work motivation. *The International Journal of Human Resource Management*, 17(3), 504-522.
<https://doi.org/10.1080/09585190500521581>
- Kuvaas, B. (2008). An exploration of how the employee–organization relationship affects the linkage between perception of developmental human resource practices and employee outcomes. *Journal of Management Studies*, 45, 1–25. <https://doi.org/10.1111/j.1467-6486.2007.00710.x>
- Le Fevre, D. M., & Robinson, V. M. J. (2015). The Interpersonal Challenges of Instructional Leadership: Principals' Effectiveness in Conversations About Performance Issues. *Educational Administration Quarterly*, 51(1), 58–95. <https://doi.org/10.1177/0013161X13518218>

-
- Lehmann-Willenbrock, N., & Allen, J. A. (2018). Modeling Temporal Interaction Dynamics in Organizational Settings. *Journal of Business and Psychology, 33*(3), 325-344.
<https://doi.org/10.1007/s10869-017-9506-9>
- Leppänen, I., Hämäläinen, R.,P., Saarinen, E., & Viinikainen, M. (2018). Intrapersonal emotional responses to the inquiry and advocacy modes of interaction: A psychophysiological study. *Group Decision and Negotiation, 27*(6), 933-948.
<http://dx.doi.org.ezproxy.library.bi.no/10.1007/s10726-018-9584-8>
- Levy, P. E., & Williams, J. R. (2004). The social context of performance appraisal: A review and framework for the future. *Journal of management, 30*(6), 881-905. <https://doi.org/10.1016/j.jm.2004.06.005>
- Lewinski, A. A., Anderson, R. A., Vorderstrasse, A. A., & Johnson, C. M. (2019). Developing Methods That Facilitate Coding and Analysis of Synchronous Conversations via Virtual Environments. *International Journal of Qualitative Methods*. <https://doi.org/10.1177/1609406919842443>
- Meinecke, A. L., Lehmann-Willenbrock, N., & Kauffeld, S. (2017). What happens during annual appraisal interviews? How leader–follower interactions unfold and impact interview outcomes. *Journal of Applied Psychology, 102*(7), 1054-1074. <https://doi.org/10.1037/apl0000219>
- Meyer, F., Birkeland, I. K., Emstad, A. B., & Le Fevre, D. M. (2019). Comparing leaders' collaborative problem-solving in conversations in Norway and New Zealand. Paper presented at the AERA, Toronto.
- Miller, V. D., & Gordon, M. E. (Eds.). (2014). *Meeting the challenge of human resource management: A communication perspective*. Routledge
<https://ebookcentral-proquest-com.ezproxy.library.bi.no>
-

-
- Moore, D. A., & Healy, P. J. (2008). The trouble with overconfidence. *Psychological Review, 115*(2), 502–517. <https://doi.org/10.1037/0033-295X.115.2.502>
- Nishii, L. H., & Wright, P. M. (2008). Variability within organizations: Implications for strategic human resources management. In D.B. Smith (Ed.) *The people make the place: Dynamic linkages between individuals and organizations* (pp. 225-248). New York, NY: CRC Press.
- Owens, B. P., Baker, W. E., Sumpter, D. M., & Cameron, K. S. (2016). Relational energy at work: Implications for job engagement and job performance. *Journal of Applied Psychology, 101*(1), 35-49. <https://doi.org/10.1037/apl0000032>
- Peshkin, A. (1988). In Search of Subjectivity—One’s Own. *Educational Researcher, 17*(7), 17–21. <https://doi.org/10.3102/0013189X017007017>
- Podsakoff, P. M., MacKenzie, S. B., Lee, J.-Y., & Podsakoff, N. P. (2003). Common methods biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology, 88*(5), 879–903. <https://doi.org/10.1037/0021-9010.88.5.879>
- Pugh, S. D., Skarlicki, D. P., & Passell, B. S. (2003). After the fall: layoff victims’ trust and cynicism in re-employment. *Journal of Occupational and Organizational Psychology, 76*(2), 201-212. <https://doi.org/10.1348/096317903765913704>
- Roberto, M. (2002). Lessons from Everest: The interaction of cognitive bias, psychological safety, and system complexity. *California Management Review, 45*(1), 136-158. <https://doi.org/10.2307/41166157>
-

-
- Roberts, G. E. (2002). Employee performance appraisal system participation: A technique that works. *Public Personnel Management, 31*(3), 333-342. <https://doi.org/10.1177/009102600203100306>
- Roberts, G. E. & Pregitzer, M. (2007). Why employees dislike performance appraisals. *Regent Global Business Review, (1)*1, 14-21.
- Robinson, V. (2009). Open-to-learning conversations: Background paper. *Module 3: Building Trust in Schools Through Open-to-learning Conversations. First-time Principals Programme.*
https://www.researchgate.net/publication/267411000_Open-to-learning_Conversations_Background_Paper_Introduction_to_Open-to-learning_Conversations/citation/download
- Robinson, V., Meyer, F., Le Fevre, D., & Sinnema, C. (2020). The Quality of Leaders' Problem-Solving Conversations: Truth-Seeking or Truth-Claiming? *Leadership and Policy in Schools.*
<https://doi.org/10.1080/15700763.2020.1734627>
- Romero, D. (2001). *The business of listening: A practical guide to effective listening* (3rd ed.). Crisp Learning.
- Schein, E. H. (2013). *Humble inquiry: the gentle art of asking instead of telling*. Berrett-Koehler Publishers.
- Schmalbach, B., Hennemuth, L. & Echterhoff, G. (2019). A tool for assessing the experience of shared reality: Validation of the German SR-T. *Frontiers in Psychology, 10*(832), 1-14.
<https://doi.org/10.3389/fpsyg.2019.00832>

-
- Sias, P. (2014). Implications of communication research for improving developing policies, procedures, and functions. In V. D Miller, & M. E. Gordon, (Eds.). *Meeting the challenge of human resource management: A communication perspective* (pp. 145-153). Routledge.
<https://ebookcentral-proquest-com.ezproxy.library.bi.no>
- Sinnema, C. E., Robinson, V. M., Ludlow, L., & Pope, D. (2015). How effective is the principal? Discrepancy between New Zealand teachers' and principals' perceptions of principal effectiveness. *Educational Assessment, Evaluation and Accountability*, 27(3), 275-301.
<https://doi.org/10.1007/s11092-014-9209-y>
- Sumelius, J., Björkman, I., Ehrnrooth, M., Mäkelä, K., & Smale, A. (2014). What Determines Employee Perceptions of HRM Process Features? The Case of Performance Appraisal in MNC Subsidiaries. *Human Resource Management*, 53(4), 569-592.
<https://doi.org/10.1057/jibs.2013.27>
- Tan, H., & Lim, A. (2009). Trust in Coworkers and Trust in Organizations. *The Journal of Psychology*, 143(1), 45-66.
<https://doi.org/10.3200/JRLP.143.1.45-66>
- Thames, R. (2015). Organizational Learning. In R. Thames & D. W. Webster (Eds.) *Chasing Change* (pp. 121-134). John Wiley & Sons.
- Tynan, R. (2005). The effects of threat sensitivity and face giving on dyadic psychological safety and upward communication. *Journal of Applied Social Psychology*, 35(2), 223-247.
<https://doi.org/10.1111/j.1559-1816.2005.tb02119.x>