Effects of Supportive Leadership Behaviors on Employee Satisfaction, Engagement, and Performance: An Experimental Field Investigation

Simone Haeckl and Mari Rege*

November 2023

Abstract

We conduct a field experiment in a large corporation to investigate the effects of supportive leadership behaviors on employee satisfaction, engagement, and performance. Treated leaders receive a brief training promoting leadership behaviors that encourage, assure, and value employee efforts. Our experimental design allows us to observe leaders and employees in a subsequent meeting. We find that the leadership training affects the leaders' supportive behaviors and thereby increases employees' self-reported satisfaction and engagement during the meeting by 0.28 and 0.18 standard deviations, respectively. The effect on team performance is 0.13 standard deviations but is not significant.

Keywords: field experiment, leadership training, leadership behaviors, supportive leadership, work satisfaction, engagement, productivity

JEL classification: C93, J24, J28, M53

1 Introduction

In modern work-life, new technology and frequent reorganizations involve constant adaption to new tasks and collaboration with new colleagues (Deming, 2017; Weidmann and Deming, 2021). In these situations, many individuals fail to contribute to their full potential because they are concerned that their colleagues will think that their ideas, reflections, or efforts are irrelevant or stupid (Edmondson, 1999; Newman et al., 2017). In addition, employees might be afraid to ask questions or ask for help because they believe that doing so may embarrass them (Chandrasekhar et al., 2018; Sandvik et al., 2020). Such psychological barriers can come at a large cost for the company due to negative impacts on communication and knowledge sharing, learning behavior, performance, and innovation (see reviews in Edmondson and Lei, 2014; Newman et al., 2017). For a firm's value creation, it seems essential to break down these barriers by creating a supportive work environment where employees feel safe to contribute and learn to the best of their abilities.

To this end, leaders may play a crucial role by encouraging, valuing, and assuring employee efforts to handle the new challenges.[1] There is ample evidence showing that leadership behaviors, i.e., how the leader interacts with the employees, are strong predictors of performance (Bass et al., 2003; Judge and Piccolo, 2004; Liu et al., 2010; Wang et al., 2011), but causal evidence is scarce (Antonakis, 2017; Martin et al., 2020).[2]

We contribute to narrowing this gap in the literature by investigating the causal impacts of leadership behaviors. Specifically, we investigate the effects of supportive leadership behaviors, which we define as behaviors that encourage, assure, and value employee efforts. Our primary outcome variables are work satisfaction and engagement, which are strong predictors of performance and turnover (e.g., Judge et al., 2001; Rich et al., 2010). Additionally, we measure team performance.

Our field experiment was conducted in partnership with a large corporation in which the CEO invited all employees to provide input on a question of strategic importance by participating in a one-hour digital team discussion during working hours. The employees with no responsibility for other personnel were randomly matched with employees with personnel responsibility (leaders) to form teams of three or four members, all from different companies within the corporation to maximize the likelihood of them not knowing each other. This created a real field setting where the employees had to solve a new task

---

[1]See also Englmaier et al. (2021) for a field experiment showing the importance of leadership in non-routine, interpersonal tasks.

[2]Notable exceptions are Alan et al. (2022), Antonakis et al. (2022), Castro et al. (2022), Fest et al. (2021), and Jacobsen et al. (2022).

with people with whom they had never worked before under new leadership. Such a field setting is highly relevant in today's work life, in which teamwork is becoming the norm, and employees constantly have to adapt to new tasks and collaborating partners (Eurofound and Cedefop, 2020; Deming, 2017, Weidmann and Deming, 2021).

Using block randomization, we randomized leaders into treatment and control. Prior to the team meetings, leaders in the control group received a 20-minute leadership training that described the upcoming team tasks and outlined the leader's responsibilities. The training for the treated leaders was identical but included an additional 20-minute insertion designed to experimentally manipulate supportive leadership behaviors by targeting three possible barriers for why leaders may not use supportive behaviors: 1) they do not believe that all employees can contribute to the task at hand: 2) they do not know supportive leadership behaviors matter for employees' contribution; 3) they have limited awareness of how to support employees' contribution. In order to target each of these barriers, the training had three steps: Step 1 was designed to increase the leaders' beliefs in their employees' potential for contributing to the task at hand; Step 2 was intended to increase the leaders' understanding of their role in making all their employees feel supported and engaged in the task at hand; Step 3 provided concrete examples of leadership behaviors that could support employees' satisfaction, engagement, and performance in the task at hand. Across all steps, we promoted the internalization of the training material by following the suggestions of Yeager et al. (2016) to use self-persuasion exercises and asked leaders to share their own reflections and stories.

The teams conducted their team meeting online on Microsoft Teams using scheduled time slots during working hours. We collected survey data assessing leaders and employees before and after the team meetings. Additionally, we collected the corporation's evaluation of each team's input to the question of strategic importance as a measure of team performance. Finally, we recorded each Microsoft Teams meeting, and from these recordings elicited observational data on leadership behaviors and employee engagement.

All the assessments were blind to treatment. Also, employees were blind to treatment. Leaders, of course, experienced their treatment condition but did not know that other leaders had received different leadership training or the purpose of the study (apart from the fact that we were studying interactions in teams and the role of leaders).

Our analyses demonstrate that the treatment increased supportive leadership behaviors during the team meetings (+0.40 sd). Moreover, in groups with treated leaders, employees reported higher work satisfaction (+0.28 sd) and engagement (+0.18 sd). The effect on team performance was 0.13 standard deviations, but was not significant. Using a mediation analysis, we find that the change in employee outcomes was mainly

driven by the increase in supportive behaviors. This evidence is important, as it demonstrates that, even during a one-hour team meeting, the leader's supportive behaviors are important for employee satisfaction and engagement.

Our paper contributes to the existing literature in three important ways. First, large observed differences in firm-level productivity have been a long-standing puzzle, and understanding these differences remains key to understanding the causes of economic growth (Bloom et al., 2013; Gosnell et al., 2020). This has led several scholars to study how management practices are linked to human resource productivity (e.g., Bender et al., 2018; Blader et al., 2019; Bloom et al., 2019, 2013; Bradler et al., 2016; Bruhn et al., 2018; Fryer, 2014; Gosnell et al., 2020; Heinz et al., 2020; Ichniowski et al., 1997; Lazear, 2000; Manthei et al., 2022; Sandvik et al., 2020). In addition, there is growing evidence that leader-specific characteristics, such as social skills, affect employee productivity (e.g., Hoffman and Tadelis, 2021; Lazear et al., 2015). None of these studies, however, investigates how actual leadership behaviors affect productivity. Our paper contributes to this literature by suggesting that supportive leadership behaviors may be important for human resource productivity.

Second, our paper contributes by providing some of the first causal evidence on supportive leadership behaviors. As noted above, studies investigating causal impacts of leadership behaviors using a high-quality experimental design in the field context are scarce (Antonakis, 2017; Martin et al., 2020), but there are a few notable exceptions: Dvir et al. (2002) and Jacobsen et al. (2022) investigate the effects of transformational leadership in the Israel Defense and for school principals, respectively, and find positive effects on performance.[3] Antonakis et al. (2014, 2022) and Fest et al. (2021) demonstrate that charismatic leadership affects performances among temporary employees.[4] Of particular relevance for our paper is the large-scale field experiment by Castro et al. (2022), which is the first RCT investigation of the impact of supportive leadership behavior on psychological safety. The authors find that frequent one-to-one meetings between leaders and employees create a more supportive work environment and increase psychological safety, especially when they are used to talk about employees' needs. We add to this literature by investigating whether supportive leader behaviors can increase employees' satisfaction, engagement, and performance. In contrast to Castro et al. (2022), in which the treatment targets supportive leader behavior *outside* the team interaction context

---

[3]Jacobsen et al. (2022) also investigate the effect of transactional leadership and consider the effect of leadership training in other professions. However, they do not observe "employee" outcomes for leaders in these other professions. In a companion paper An et al. (2020) show that employees whose leaders received training in transformational leadership report to be more satisfied.

[4]The results of Antonakis et al. (2022) have also been replicated by Meslec et al. (2020).

by encouraging one-to-one meetings, our treatment targets leadership behaviors *during* the team interaction.

Another study related to our focus on supportive leadership behavior is Alan et al. (2022). The authors investigate results from a large-scale intervention providing leaders and employees with a training to improve the workplace environment. The authors find that the training reduces toxic competition (measured as antisocial behavior in economic games), increases workplace satisfaction and meritocratic views (based on survey measures), improves employee integration (using a network analysis), and reduces worker turnover (based on admin data). While the intervention, in contrast to our design, is offered to both leaders and employees, the authors show that an improvement in the relationship between leaders and employees is the main driving factor.

Third, a major innovation in this paper is the development of a data-collection procedure in partnership with organizations, which gives us direct access to the leaders, allows us to experimentally manipulate leaders' behaviors, and ensures task comparability across leaders. One reason for the lack of causal field evidence on leadership behaviors is that access to a suitable field is extremely difficult (Martin et al., 2017). A well-powered RCT requires more than a hundred leaders who must be somewhat comparable. Even within organizations, leaders often differ substantially in their job design. To recruit a hundred leaders, several organizations must often be involved, and this makes it a challenging task both to obtain access to the leaders and to make comparisons across leaders. In our paper, we provide a solution to this problem by developing an experimental manipulation and data-collection procedure that benefits the participating organizations.

2 Conceptual Framework

In our field experiment, the employees are asked to solve a new task with people with whom they have never worked before under new leadership. This setting is of high relevance in today's work life, in which teamwork is becoming the norm, and employees constantly have to adapt to new tasks and collaborating partners (Eurofound and Cedefop, 2020; Deming, 2017). In the 2019 version of the European Company Survey, a survey conducted in over 20,000 establishments across Europe, 71% of the surveyed establishments (across country average) stated that teamwork took place in their company. Out of these companies, 28% stated that employees worked in multiple, management-directed teams.

For many individuals it may be stressful to collaborate with new people to solve a new

task. They may fail to thrive and contribute to the team effort because they are worried that the other team members will think their ideas or reflections are irrelevant or stupid (Edmondson, 1999). In this setting, supportive leadership behaviors, i.e. behaviors that encourage, assure, and value employees' effort, may be critical for employee satisfaction, engagement, and performance (Newman et al., 2017).

For clarity, consider an individual contemplating how much to engage in the team task. The individual derives a benefit in terms of increased work satisfaction from engagement, but only if she feels her engagement is perceived as useful by her team members. If she feels her engagement is perceived as non-useful, she experiences a humiliation cost, decreasing work satisfaction. As such, supportive leadership can increase engagement and work satisfaction by assuring employees that no suggestions are stupid and that each idea has value, encouraging employees to think outside the box, encouraging employees to be constructive when criticizing others' ideas, and by valuing employees' engagement by listening carefully to their suggestions.

This leads to the following hypotheses:

Hypothesis 1: *More supportive leadership behaviors will increase work satisfaction.*

Hypothesis 2: *More supportive leadership behaviors will increase engagement.*

Previous literature suggests that both satisfaction and engagement are positively related to performance (e.g., Judge et al., 2001; Rich et al., 2010). In addition, supportive leadership behaviors have been found to relate positively to performance (Derue et al., 2011). Therefore, we form a third hypothesis:

Hypothesis 3: *More supportive leadership behaviors will improve performance.*

In contradiction to Hypothesis 3, important recent evidence suggests that more engagement can reduce performance in problem-solving tasks (Charness et al., 2020). The reason is that people contribute too many ideas, which makes it harder to identify useful contributions. This mechanism may result in a negative impact of our treatment on our team performance measure. However, in contrast to Charness et al. (2020), in our setting, there is no one correct solution and there is a psychological cost of contributing. This leads us to hypothesize that more contributions translate into performance improvement, but it remains an empirical question.

Our conceptual framework is closely related to Edmondson's pioneering paper on psychological safety, defined as a *shared belief held by members of a team that the team is safe for interpersonal risk-taking* (Edmondson, 1999). The paper demonstrates that psychological safety predicts team learning behavior and subsequently team performance.

Later studies have demonstrated that the construct is linked to various important outcomes, such as communication and knowledge sharing, learning behavior, performance and innovation behavior, and employee attitudes (see review in Newman et al., 2017). Moreover, several studies demonstrate that leadership behaviors are predictive of employees' psychological safety (Castro et al., 2022; Edmondson, 1999; Hirak et al., 2012; Leroy et al., 2012; Liu et al., 2014; Nemanich and Vera, 2009; Nembhard and Edmondson, 2006; Newman et al., 2017; Ortega et al., 2014; Walumbwa and Schaubroeck, 2009). To the best of our knowledge, so far, no studies have investigated how leadership behaviors intended to increase psychological safety causally affect employee satisfaction, engagement, and performance.

3 Experimental Design and Empirical Strategy

We conducted our field experiment in partnership with a large corporation over the span of four weeks. In this section, we first describe the experimental context, experimental design, and procedures. Second, we outline the resulting empirical strategy.

3.1 Experimental Context

The corporation we are collaborating with is an industrial and multi-utility company with nearly 2000 employees. It has operations in energy, telecommunication, and infrastructure, selling electricity, internet and TV fiber access, electric car chargers, and solar panels. The employees might have very different work tasks depending on the business area. For example, employees working in infrastructure are more likely to work in the field, while those working in telecommunication or sales are more likely to have office jobs. There is also a lot of variation in skills and training depending on work tasks - from high school graduates with on-the-job training to engineers with advanced university degrees.

The experiment was implemented at the end of the Corona Pandemic. At this point all workers were familiar with Microsoft Teams digital meetings, which we used to implement our experiment.

3.2 Experimental Design and Procedures

The CEO of the corporation invited all employees to participate in a one-hour digital team meeting during work hours to provide input on a question of strategic importance. About 70% of the employees with personnel responsibility (leaders) and 30% of employees

6

with no responsibility for other personnel (employees) signed up for participation and indicated their availability for the team meeting.[5]

The employees were randomly matched to leaders to form teams of three or four members, all from different companies within the corporation, to maximize the likelihood of them not knowing each other. This created a real field setting where the employees were asked to solve a new task with people with whom they had never worked before under new leadership. As discussed in the introduction, this setting is becoming increasingly common in modern work-life. Another advantage of creating teams across companies is that it decreases across-team variation, which helps increase the power. Lastly, creating teams across the companies was also requested by our collaborating partner because they wanted to strengthen the ties between companies within the corporation and the employees' sense of identity with the cooperation.[6]

All meetings were conducted within the same week, but days varied based on availability. The teams received an e-mail providing the time for their one-hour digital team meeting and a link to Microsoft Teams. Leaders received a second e-mail with a link for the digital leadership training, which started one hour prior to their scheduled team meeting. Using block randomization (based on the meeting day), the leaders (and their teams) were randomized into treatment and control. Leaders in the control group received a 20-minute leadership training which described the upcoming team tasks in detail and outlined the responsibilities as team leaders. The training for the treated leaders was identical but included an insertion designed to encourage supportive leadership behaviors. This insertion lasted about 20 minutes and is described below in Section 4.

During the team meeting, the leader led the group through a round of introductions, an icebreaker task, and, most importantly, a team discussion in which the teams collected input to a question of strategic importance. The question was determined by the leadership of our collaborating corporation. Specifically, this question was: "What can our corporation do to become an even more attractive workplace?" Each team was sup-
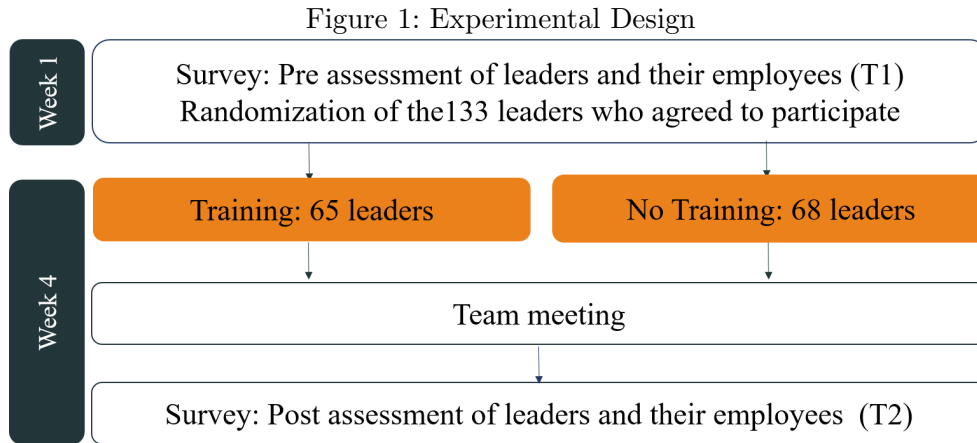
---

[5]Unfortunately, we do not have any information about employees who did not agree to participate. We do, however, see that employees in sub-companies that have more recently joined the corporation are less likely to join. In addition, the sign-up rate differs by the sub-company's business field, indicating that the type of job employees are conducting (field versus office work) affects their decision to participate.

[6]To measure the extent to which we were successful in limiting social ties, we asked each participant whether they knew any of the other participants prior to the meeting. The likelihood of knowing at least one of the persons is 34% in the control group and 20% in the treatment group ($p = 0.002$). However, this variable has no explanatory power for any of our outcome variables, and controlling for it does not affect any of our results.

posed to spend the first 15 minutes brainstorming different suggestions.[7] Thereafter, the team was supposed to spend another 15 minutes choosing and elaborating on the three most promising suggestions on one PowerPoint slide each. These three PowerPoint slides were the teams' input to the CEO and are used as our measure of team performance.

Figure 1: Experimental Design

Week 1	Survey: Pre assessment of leaders and their employees (T1) Randomization of the133 leaders who agreed to participate	
	Training: 65 leaders	No Training: 68 leaders
Week 4	Team meeting	
	Survey: Post assessment of leaders and their employees (T2)	

As illustrated in Figure 1, we also solicited survey measures four weeks prior to the intervention (T1) and directly after the meeting (T2). In addition, leaders were asked to record the meetings, which allowed us to construct additional measures based on observed behavior, as detailed in Section 5.

Notably, all team members were informed that the meetings were recorded. Even if the employees were used to digital meetings, recording a meeting in which they themselves were active was new to most of them. Still, it seemed like the recordings were not causing big concerns for participants as compliance was 100 percent, and we did not receive any complaints. During implementation, a researcher was observing the recording status of all teams in real time. If the leader did not activate the recording within the first 10 minutes of the meeting, the researcher would enter the meeting and ask the leader to activate the recording. Nobody protested, and all leaders complied. It was made clear that the recordings were made for research purposes only, that everybody would remain anonymous, and that recordings would be deleted at the end of the project. Importantly, the request to record was identical across treatment statuses.

In total, 133 leaders and 338 employees signed up for the study. Three work teams were excluded from the analysis, two because of unforeseen absences of leaders or employees

---

[7]We piloted the experimental procedure and measures with five student teams that were discussing what the University could do to become an even better place to study. Due to the small sample size, we did not analyze the data except for checking whether there were questions or complaints.

and one because the leader did not read the instructions and was therefore not able to lead the meeting. This reduces our sample to 130 teams. To ensure a high completion rate in the survey, leaders were asked to share the survey link as the last step of the instructions. In addition, a researcher ensured that the link was sent out to the employees and sent reminders to employees and leaders who had not finished the survey by the end of the day. As a result, only 13 employees in the control group and 16 employees in the treatment group as well as one leader did not finish the end survey. We cannot reject the hypothesis that attrition is balanced based on a Fisher's exact test ($p = 0.437$).

3.3 Empirical Strategy

To investigate treatment effects, we use an intention-to-treat approach, estimating the following OLS model:

$$Y_{i,t} = \alpha + \gamma T_t + \beta X_{i,t} + \epsilon_{i,t}. \tag{1}$$

$Y_{i,t}$ is the outcome measure for employee i in team t; $X_{i,t}$ is a vector of the employee, team, and leader characteristics at baseline (T1); α is the constant term; $\epsilon_{i,t}$ is the error term; T_t is an indicator of the treatment status of the leader of team t; and γ is the estimated treatment effect. For employee-level outcomes, we cluster at the team level. For team performance or leader outcome, we will estimate Equation 1 at the team level.

As we randomize within weekdays, all models control for blocks, i.e., the day on which the meeting took place. Additional control variables in $X_{i,t}$ are described in Section 5. Given that assignment to treatment is random, controlling for baseline characteristics should not affect the treatment estimate but could increase precision. We will estimate all our models with and without the additional controls.

The power calculations in the pre-analysis plan suggested that, for individual-level data, we would be able to detect an effect of 0.29 standard deviations with alpha=0.05 and a power of 80%, assuming 140 leaders (cluster), three employees per cluster, and an intraclass correlation of 0.05. For group-level data, we expected to be able to detect an effect of 0.48 standard deviations with alpha=0.05 and a power of 80%, assuming 140 independent observations at the group level. Such effect sizes are comparable to the effects found in studies with a similar design. For example, Antonakis et al. (2022) find that hearing a charismatic speech prior to working on a task improves performance by 0.46 standard deviations.

In addition to investigating treatment impacts, we conduct a decomposition analysis following Heckman et al. (2013) and Heckman and Pinto (2015) to learn more about the mechanisms underlying the treatment effect. We describe this approach in Section 6.4.

4 Supportive Leadership Treatment

The 20-minute digital leadership training was designed to increase leaders' use of supportive leadership behaviors. The training targeted three possible barriers for why leaders may not use supportive behaviors: 1) they do not believe that all employees can contribute to the task at hand; 2) they do not know that supportive leadership matter for employees' contribution; 3) they have limited awareness of how to support employees' contribution. For leaders to be supportive, all these barriers need to be broken down. Notably, however, it may be that not all these barriers were present in our field context. With access to 130 leaders, we were limited to a field experiment with two arms investigating the effects of the combined treatment and using a mediation analysis to provide suggestive evidence of the most relevant barriers. See Section 7.4 for a more extensive discussion of the combined treatment.

In order to target the three barriers, the training had three steps: Step 1 was designed to increase the leaders' beliefs in their employees' potential for contributing to the task at hand; Step 2 was intended to increase the leaders' understanding of their role in making all their employees feel supported and engaged in the task at hand; Step 3 provided concrete examples of leadership behaviors that could support employees' satisfaction, engagement, and performance in the task at hand. Across all steps, we promoted the internalization of the training material by following the suggestions of Yeager et al. (2016) to use self-persuasion exercises and asked leaders to share their own reflections and stories. Below, we provide some more details on each step. For a full transcript of the training, please see Appendix G.

In Step 1, we promoted a growth mindset – the belief that people can learn and develop with some effort and support from others (Dweck, 2006). Our text was inspired by existing, successful growth mindset interventions (Bettinger et al., 2018; Heslin et al., 2006; Rege et al., 2020; Yeager et al., 2019), but was adapted to promote growth mindset behaviors in the team leadership context. Moreover, we introduced research suggesting that leaders' mindsets affect their management behaviors (Heslin et al., 2006). This research demonstrated that leaders who believe that human potential is generally fixed and not amenable to change will often provide insufficient encouragement and support. We explained how leaders' mindsets could become a self-fulfilling prophecy and provided a context-relevant example on this: "For example, if a leader does not believe that an employee has the potential to contribute to a meeting, this leader might be less willing to support and encourage the employee, which can in of itself make it more difficult for the employee to contribute."

10

We concluded Step 1 by explaining that we were currently developing a leadership training (which was true) and needed help with good real-world examples to improve the training. Then we asked: "Could you please describe a situation in which one of your employees has shown a development and achieved something which many people wouldn't have expected of him or her?" This is a well-known self-persuasion technique from psychology prompting the subjects to see the relevance of and internalize the training material (Yeager et al., 2016).

In Step 2, we explained how many employees feel insecure about their ability to contribute to a group task and are reluctant to voice their thoughts in fear of their ideas being perceived as irrelevant or stupid. The purpose was to increase the leaders' understanding of their role in making all their employees feel supported, safe, and engaged to contribute. We concluded Step 2 by asking the leaders to reflect on their own leadership behaviors. We asked them to think about their own experiences and to describe a meeting in which it felt easy and safe to share their ideas. Again, we said (truthfully) we needed concrete real-world examples for a leadership training that we were in the process of developing.

In Step 3, we provided concrete examples of leadership behaviors that could support employees' work satisfaction and engagement, and increase performance in the task at hand. In particular, we described how leaders could assure, encourage, and value employee efforts, as described in Section 2. Leaders could assure employees by emphasizing that no suggestions are stupid and that it is important to think outside the box. In addition, leaders could encourage constructive criticism and make employees feel it was safe to contribute. Lastly, leaders could make employees feel valued by listening carefully to their suggestions and emphasizing that each idea has its value.

In order to scaffold leaders' internalization of the leadership behaviors, we concluded Step 3 by asking them to help a colleague: "Imagine that a colleague asks you what he or she can do to become a better meeting discussion leader. This colleague has observed that many participants do not dare to speak up in their project meetings. What would you tell him or her?"

It is important to note that by design our change in leadership behavior is likely due to experimenter demand; the leadership training gives very concrete instructions to use supportive leadership behaviors. The research question is how employees, who are blind to treatment, respond to these changes in leadership behaviors and how it affects group productivity. Previous studies have obtained experimental manipulation in leadership behaviors by using actors as leaders in more controlled laboratory settings (see e.g., Antonakis et al., 2022).

5 Measures

We collected outcome, manipulation, and control measures as described below.[8] For multiple-item survey measures, we used the average across all (given) answers as long as at least one of the items was answered. Otherwise, it is reported as missing. We report detailed summary statistics for each measure in Appendix E. For survey measures we also report summary statistics for all items and Cronbach's alpha evaluating internal consistency in Appendix E.

Outcome Measures

Our three primary outcome variables are work satisfaction, engagement, and team performance. Work satisfaction and engagement are strong predictors of productivity and turnover (e.g., Judge et al., 2001; Rich et al., 2010). The team performance was important both for the company and our research question. The company spent a lot of resources on the team meetings, in terms of the time of the employees, and wanted the meetings to result in useful input for the firm. Below, we describe how we measure each of these outcomes.

Satisfaction: To measure *Satisfaction*, we adapted the one-item measure carefully validated by Dolbier et al. (2005). Precisely, participants responded on a 5-point Likert scale to the following question: "All in all, I was very satisfied with the discussion I had with my colleagues about our company's future." We used this single-item measure, which performs similarly to a multi-item measure, in order to keep our survey short.[9]

Employee Engagement: For employee engagement, we have two different outcome measures. Our main measure of interest is *Engagement*, measured in a survey right after the group task. There are several validated measures of work engagement. However, the items in these measures refer back to work in general. We wanted to measure work engagement in the group discussion. To capture work engagement in the group discussion, we constructed a measure based on three questions, all rated on a 5-point Likert scale: "During the group discussion I..." 1) saw myself as an important part

---

[8]All definitions follow the pre-registered pre-analysis plan, with the exception of one minor change. The external raters evaluated the leadership practices indicating whether or not the leader showed the respective practice rather than rating each practice on a scale. This change was made to simplify the work of the raters.

[9]See Wanous et al. (1997) for a discussion of the appropriateness of using single-item measures to elicit work satisfaction.

of the discussion, 2.) did my best to come up with good ideas, and 3) experienced the discussion as interesting. Internal consistency was considered adequate (Cronbach's alpha = 0.73; see Table E.2 in the Appendix).

In addition, we also elicited an observational measure of engagement based on the transcriptions of the video recordings. The external raters were trained and instructed to construct *Engagement (video)* defined as the number of times the person contributing least contributed to the discussion. The measures are weakly positively correlated ($\rho = 0.16$, $p = 0.081$, 124 teams). We will focus on the survey measure in the main part of the paper as the survey measure is available at the individual level while the experimental measure is only observed at the group level. We discuss how the results change when using the experimental measure in the appendix (see Table A.10).

Performance: To measure *Performance*, the output created by each group (PowerPoint presentation with suggestions) is evaluated in terms of value for the strategic question of the company by the Human Resources department of the company. Precisely, four employees working in HR rated each of the ideas independently on a scale from one to five based on originality and relevance. These employees are blind to treatment and have no information about the team composition, i.e., they do not know which employees and leaders are part of each team. More generally, they do not even know which employees and leaders signed up to participate in the project. After the individual evaluation, they discussed the performance of each team and agreed on a rating.[10] Performance is the average score across all three ideas. In addition, we consider different definitions and look at subcategories in an exploratory analysis in Appendix C. The HR department took the most promising ideas and included them in the company's strategy discussion.

Manipulation Measures

Leader Supportive Behavior: We elicited two measures of leadership behaviors during the group discussion. *Supportive behavior* was measured using the following self-developed five-item scale: To what extent did the leader of the group discussion: 1) Encourage you to think outside the box? 2) Make you feel safe and secure? 3) Listen to the suggestions you were making? 4) Provide ideas him or herself? 5) Make you feel comfortable with participating in the discussion? Employees answered this question at the end of the group discussion using a 5-point Likert scale. We take the average across answers of employees as our proxy for the supportive behavior of the leader. As long

---

[10]A detailed description of each category is provided in the appendix in Table A1.

as at least one employee answered, we use his or her answer as a proxy for the leader's behavior. Internal consistency was considered adequate (Cronbach's alpha = 0.79; see Table E.4 in the appendix).

In addition, we used the video material to create a second measure of leader behavior. Two external reviewers, who were trained, worked independently, and were blind to treatment, indicated for each of the behaviors suggested in the treatment whether or not the leader engaged in it.[11] After an independent evaluation, they discussed the behavior of each leader and agreed on a rating. *Supportive behavior (video)* is the sum of the supportive behaviors the leader engaged in. Both measures of leader behavior are significantly positively correlated ($\rho = 0.22$, $p = 0.014$, 129 teams). We present results from the survey measure in the main part of the paper and demonstrate that the manipulation checks are robust to using the video measure in the appendix (see Table A.3).

Leader Mindset: *Leader mindset* was measured right after the group task using the short version of the implicit theory measure developed by Chiu et al. (1997). This short measure consists of the following three questions: 1) The kind of person someone is, is something basic about them, and it can't be changed very much; 2) People can do things differently, but the important parts of who they are can't really be changed; and 3) Everyone is a certain kind of person, and there is not much they can really change about that. Leaders answered these questions on a 6-point Likert scale. Internal consistency was considered adequate (Cronbach's alpha = 0.92; see Table E.5 in the appendix).

Control Variables

We used administrative data to collect information on the participants' gender and seniority, i.e., number of years in the company.[12] We also received information on the company structure, allowing us to identify the leaders' level, i.e., a dummy with the value

---

[11]Precisely, the reviewers indicated whether or not the leaders engaged in each of the following behaviors: 1) Emphasize that each group member's ideas are important because they all have their own experience and it is important to come up with a variety of ideas. 2) Acknowledge that some might feel nervous to contribute in the group discussion, but clarify that there is no need to be shy and that it is important that all of them contribute. 3) Emphasize that there are no stupid suggestions. 4) Look into the camera when others speak. 5) Write down their suggestions when they see it. 6) Ask clarification questions. 7) Emphasize that there is also room for more radical ideas. 8) Share one very radical and one more realistic idea yourself to show the group members that any type of contribution is welcome. 9) Allow group members to be critical but honor the potential of each idea. 10) Emphasize that it is critical that they all speak up in order to select and elaborate on the three best ideas.

[12]We have added a random number drawn from a mean 0 discrete uniform distribution within a reasonable range to seniority in order to maintain anonymity.

one if they are leading other leaders, and used pre-existing work teams to evaluate the leaders' coaching behavior in a survey.[13] Employees for whom the leader has personnel responsibility evaluated the leader on a 5-point Likert scale using the 10-item coaching behavior scale developed by Heslin et al. (2006). We take the average across employees' answers as our measure of a leader's coaching behavior. As long as at least one employee answered, we use their answer as a proxy for the leader's behavior. However, we do not observe employees for all leaders. There are two reasons why the leaders' employees do not assess their behaviors: 1) they did not participate in the study or 2) they are themselves a leader and, therefore, received the leader survey, not the employee survey. This lack of information on higher-level leaders is a drawback of the design of our survey, but it was necessary to keep the survey time for leaders. In the regressions, we use a dummy to account for missing observations.

6 Results

The result section is structured as follows. We first provide descriptive statistics and show that our control variables are balanced across treatment and control. Second, we test whether the treatment was successful in manipulating the leaders' mindsets and behaviors. Third, we investigate the effects of the treatment on employee satisfaction, engagement, and performance using the intention-to-treat model in Equation 3. Fourth, we decompose the treatment effect using a mediation analysis.[14] Fifth, we investigate heterogeneous treatment effects, and lastly, we discuss long-term effects.

6.1 Descriptive Statistics and Balance Test

Table 1 presents our balance table. We include only the 130 teams that participated in the group discussion and the 302 employees who finished the end survey.[15]

Team size refers to the number of people participating in the team meeting. All other control variables are based on administrative data from the company (gender, seniority,

---

[13]We also elicited additional control variables, which we use in additional robustness checks in Appendix D.

[14]For readability, we do not show coefficients of control variables in the tables. The extended version of the tables is presented in Appendix A.

[15]In total, 133 leaders and 338 employees signed up for the study. Three work teams are excluded from the analysis, two because of unforeseen absences of leaders or employees and one because the leader did not read the instructions and was therefore not able to lead the meeting. This reduces our sample to 130 teams. In addition, 13 employees in the control group and 16 employees in the treatment group did not finish the end survey which reduces the employee sample to 302 observations. We cannot reject the hypothesis that attrition is balanced based on a Fisher's exact test ($p = 0.437$).

Table 1: Balance table

Variable	(1) Control	(2) Treatment	(3) Difference	(4) Obs.
Team-level measures				
Team size	3.612	3.476	-0.129	130
	(0.602)	(0.692)	(0.111)	
Female leader (indicator)	0.269	0.302	0.032	130
	(0.447)	(0.463)	(0.081)	
Seniority (leader)	10.687	9.968	-0.731	130
	(8.419)	(8.720)	(1.525)	
Leader level	0.164	0.238	0.073	130
	(0.373)	(0.429)	(0.071)	
Coaching behavior	3.866	3.696	-0.169	83
	(0.618)	(0.691)	(0.146)	
Observations	67	63	130	
Individual-level measures				
Female employee (indicator)	0.333	0.393	0.058	302
	(0.473)	(0.490)	(0.056)	
Seniority (employee)	9.037	8.314	-0.589	302
	(8.676)	(9.319)	(1.035)	
Observations	162	140	302	
F-statistic testing for joint significance (Team level)			1.17	
(*p*-value)			(0.321)	
F-statistic testing for joint significance (Individual level)			1.41	
(*p*-value)			(0.198)	

Notes: See Section 5 for a detailed description of control variables. Columns provide the mean (and standard deviation) for the work teams in the control group (column 1), and the treatment group (column 2), and the estimated coefficient (robust standard error) from regressing each covariate against treatment status controlling for block fixed effects (column 3). * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

and leader level) except for coaching behavior which was elicited in the baseline survey (see also Section 5). The column labeled "Control" presents the means of the 67 teams or 162 employees in the control group, and the column labeled "Treamtent" presents the means of the 63 teams or 140 employees in the treatment group, respectively. We have included standard deviations in parentheses. The column labeled "Difference" shows the mean difference between the control and the treatment group based on ordinary least squares regressions of the treatment dummy on the various control variables including block fixed effects. For coaching behavior the data is only available for 83 leaders. As can be seen from the table, we do not detect any significant differences between the treatment and control group, and do not find that the control variables jointly have

explanatory power for the treatment assignment, confirming that randomization was successful.

6.2 Manipulation Checks

Before we discuss treatment effects on our main outcome variables, we test whether the manipulation was successful, i.e., if the treatment successfully motivated the leader to engage in supportive behaviors as reported by the employees. Table 2 shows OLS regressions at the group level with supportive behavior, as reported by the employees, as the dependent variable.

Table 2: Manipulation Checks

| | Supportive behavior | | Leader mindset | |
	(1)	(2)	(3)	(4)
Treatment	0.372**	0.397**	0.523***	0.590***
	(0.173)	(0.172)	(0.186)	(0.187)
	[0.04]	[0.04]	[0.01]	[0.01]
Observations	129	129	127	127
Adjusted R^2	0.012	0.042	0.049	0.050
Block FE	YES	YES	YES	YES
Controls	NO	YES	NO	YES

Notes: OLS regression with block-fixed effects and robust standard errors in parentheses; numbers in brackets indicate family-wise rejection probabilities adjusting for multiple-hypotheses testing using the Westfall and Young procedure (Jones et al., 2019). The dependent variable is the proxy for the extent to which the leader engaged in supportive leadership practices as reported by the work team in columns (1) & (2) and the leader's self-reported mindset in columns (3) & (4). Dependent variables are normalized relative to the distribution of the control group. The sample size under 130 is due to missing observations of the dependent variable. Controls include leader gender, team size, seniority of the leader and average seniority of the team members, the leader's level and a proxy for coaching behavior by the leader as reported by his or her employees. We include a dummy for missing observations for coaching behavior in columns (2) & (4) to keep the sample constant. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

In columns (1) & (2), we can see that employees in teams whose leaders received the treatment are around 0.4 standard deviations more likely to state that their leader engaged in supportive behaviors during the group discussion. Comparing the estimates we can see that it is robust to including control variables. We replicate columns (1) & (2) of Table 2 using supportive leadership (video) as the dependent variable in Table A.3 in the appendix and find similar results. We conclude that the treatment successfully increased supportive behavior.

As described in Section 4, the leader training also promoted leaders' beliefs in their team members' capacity to contribute to the group task (growth mindset). In columns (3) & (4), we find that the treatment increases the leaders' reported growth mindset

by around 0.6 standard deviations. The estimate is robust to the inclusion of control variables.

The items measuring growth mindset are closely tied to the leadership training and are self-reported. As such, the measure may capture an experimenter demand effect. We, therefore, interpret the positive effects as evidence that the leaders, at the very least, paid attention to the treatment material. In contrast, the measure of supportive leadership is reported by the employees, indicating that there was indeed a change in the leaders' behavior. The leadership training gives very concrete instructions to be a supportive leader, and our measure asks employees about these behaviors. Consequently, we refer to the evidence in Table 2 as a manipulation check, i.e., we investigate whether we were able to break down the barriers described in Section 4. In addition, we conduct an exploratory analysis using AI to classify leaders' behavior and find that treated leaders are significantly more likely to be classified as supportive (see Appendix B).

6.3 Effect of Treatment on Employee Satisfaction, Engagement, and Performance

Figure 2 shows the distribution of our main outcome variables by treatment. The light, orange bars and lines represent the treatment group the dark, grey lines represent the control group. Starting with employee satisfaction (left panel), we see that, overall, the employees were fairly satisfied with the discussion. No employee strongly disagrees (1) with the statement, "All in all, I was very satisfied with the discussion I had with my colleagues about our company's future." Still, we see a shift towards the right in the treatment group. The treatment shifts the level of agreement of the median employee from "agree" (4) to "strongly agree" (5). This shift in distributions is also statistically significant using non-parametric tests ($p = 0.004$, Mann-Whitney test).

Moving one panel to the right, we show the distribution of our measure of engagement, which is the weighted average of three items, all rated on a five-point Likert scale (see Section 5). In general, employees report having been quite engaged in the discussions, with 80% agreeing, on average, to the three engagement statements. Again, we see a shift to the right, that is, a shift towards more agreement to having been engaged. This shift in distributions is also marginally statistically significant using non-parametric tests ($p = 0.073$, Mann-Whitney test).

Lastly, the right panel shows the distribution of the team's performance evaluated by the human resources specialists of the cooperating corporation (who were blind to treatment as well as team composition). The average score the HR specialists give is 2.68 out of 5 points, and it ranges from 1 to 4.33 points. While the average score is

Figure 2: Treatment effects on main outcome variables
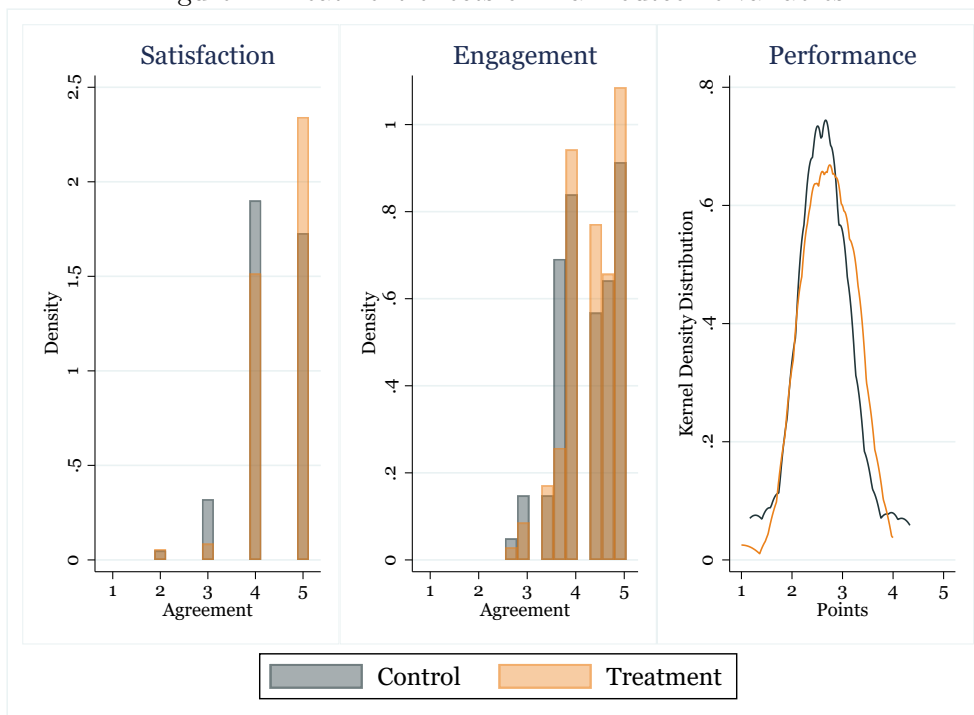
Notes: Histograms of main the main outcome variables for the treatment group (orange) and control group (gray). The left panel shows the distribution of work satisfaction (N = 302), the middle panel shows the distribution of employee engagement (N = 302), and the right panel shows the distribution of performance (N = 130).

with 2.72 points slightly higher in the treatment group than in the control group (2.65 points), this difference is not statistically significant ($p = 0.392$, Mann-Whitney test).

In Table 3, we investigate whether the treatment has beneficial effects on the employees and their performance using OLS regressions. We use individual-level data for work satisfaction and engagement (302 observations) and group-level data for performance (130 observations).[16] Individual-level data is clustered at the work-team level.

In columns (1) & (2), work satisfaction during the group meeting, as self-reported by the employees, is the dependent variable. Having a leader who received the treatment increases employee work satisfaction during the group work by about 0.28 standard deviations. This increase is statistically significant at the 5 percent level in all specifications also when adjusting for multiple-hypotheses testing (numbers in brackets). In the appendix, we provide additional results from a multinomial logistic regression in Table A.7 to take into account that we measure satisfaction using a one-item measure with a

---

[16]In Table A.6 in the appendix, we show that the results are robust to accounting for differences in the number of answers per team using a weighted regression.

Table 3: Treatment effects on employee satisfaction, engagement, and group performance

	Satisfaction		Engagement		Performance	
	(1)	(2)	(3)	(4)	(5)	(6)
Treatment	0.296**	0.278**	0.187*	0.176	0.121	0.131
	(0.119)	(0.125)	(0.109)	(0.113)	(0.157)	(0.155)
	[0.04]	[0.06]	[0.21]	[0.25]	[0.49]	[0.49]
Observations	302	302	302	302	130	130
Adj. R^2	0.018	0.017	0.001	-0.009	0.029	0.085
Block FE	YES	YES	YES	YES	YES	YES
Controls	NO	YES	NO	YES	NO	YES

Notes: OLS regression with block-fixed effects. Numbers in parentheses indicate standard errors; numbers in brackets indicate family-wise rejection probabilities adjusting for multiple-hypotheses testing using the Westfall and Young procedure (Jones et al., 2019). In columns (1) - (4) we use individual-level data from 302 employees who completed the survey after the work task and cluster standard errors by team, in columns (5) & (6) we use group level from data 130 groups and report robust standard errors. The dependent variable is the employees' self-reported satisfaction with the group task in columns (1) & (2) and the employees' self-reported engagement in columns (3) & (4). In columns (5) & (6) the dependent variable is a compound measure of output quality with equal weights on relevance and originality. Dependent variables are normalized relative to the distribution of the control group. Controls include leader gender, team size, seniority of the leader and average seniority of the team members, the leader's level, and a proxy for coaching behavior by the leader as reported by his or her employees. We include a dummy for missing observations for coaching behavior in columns (2), (4), and (6) to keep the sample constant. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

5-point Likert scale. Moreover, we address potential concerns about the robustness of findings based on the cardinalization of an ordinal scale by showing that participants are more likely to at least agree or strongly agree that they were satisfied with the meeting using a linear probability model. The results are in line with our findings provided in Table 3.

In columns (3) & (4), employee engagement, as self-reported by the employees, is the dependent variable. While we find a positive and consistent treatment effect of around 0.18 standard deviations, we barely have sufficient power to find effects of this size (see also Section 3 for the power analysis). Also, adding additional controls does not significantly improve precision, as the control variables have limited explanatory power (see also Table A.5 in the appendix), and we find no statistically significant results when controlling for multiple hypotheses testing. Using a linear probability model in Table A.9, however, we show that employees in the treatment group are significantly more likely to at least agree or strongly agree that they were engaged in the meeting.

In addition to the survey measure presented in Table 3, we also pre-registered an observational measure of engagement based on the number of contributions of the person contributing the least, which is reported in Table A.10 in the appendix. We do not find any significant treatment effects using the observed measure. To gain deeper insights into employees' behavior during the meetings, we also conducted an exploratory

analysis, presented in Appendix B, where we investigate additional potential measures of engagement, such as the total number of contributions, the number of useful contributions, and the number of words spoken, and we have calculated Herfindahl-indices to better understand how (un)balanced the discussion was. We do not find significant treatment differences in any of these measures. This discrepancy suggests that while the intervention was able to increase the average perceived engagement, it was not strong enough to change the contribution behavior in the teams. This conclusion is in line with additional exploratory analysis also reported in Appendix B, showing that leaders in the treatment teams speak more and appear to be more engaging as classified by research assistants blind to treatment. In addition, we have used AI to classify the emotions employees show during the team discussion. We find suggestive evidence that employees of treated leaders appear to be happier during the meetings, which could explain why they report to be more engaged.[17]

Finally, we test whether the treatment also improves employee performance in columns (5) & (6). To that end, we use performance based on ratings by human resources specialists of the cooperating company as the dependent variable. While we consistently find positive effects in the size of 0.13 standard deviations, this effect is too small to reach statistical significance given our sample size. Again, adding additional control variables does not increase precision significantly, as the explanatory power of the control variables is limited (see Table A.5).

In sum, we find a strong positive effect of our treatment on work satisfaction (Hypothesis 1) and weaker evidence of a positive effect on work engagement (Hypothesis 2). While it looks like there is also an improvement in performance (Hypothesis 3), the effect is much smaller than for work satisfaction and not detectable given our sample size. Considering the theory of change underlying the treatment it is not surprising that we find smaller effects for performance because while supportive behaviors are expected to directly encourage satisfaction and engagement, the effect on performance is a downstream outcome resulting from increased engagement.

6.4 Decomposing Treatment Effects

In the following, we decompose the treatment effect of each outcome variable into the effect mediated through the leaders' mindset, the effect mediated through supportive

---

[17]We would like to note here that the ability of AI to identify emotions correctly should not be over-estimated and has clear limitations, especially in culturally-diverse samples and depending on the context and video conditions. Given the homogeneity of our sample and the comparable setting, we still believe that it helps us to better understand how the treatment affected employees in our setting.

leader behavior, and the residual treatment effect.[18] This part of the analysis is exploratory and has not been preregistered.

Following Heckman et al. (2013) and Heckman and Pinto (2015), we assume that outcomes (Y_i) of individual i are a linear function of observed mediators (Θ_i^j) and the control variables (X_i).[19]

$$Y_i = \alpha T_i + \sum_{j \in J} \beta^j \Theta_i^j + \gamma X_i + \epsilon_i - \tag{2}$$

Figure 3: Decomposing Treatment Effects
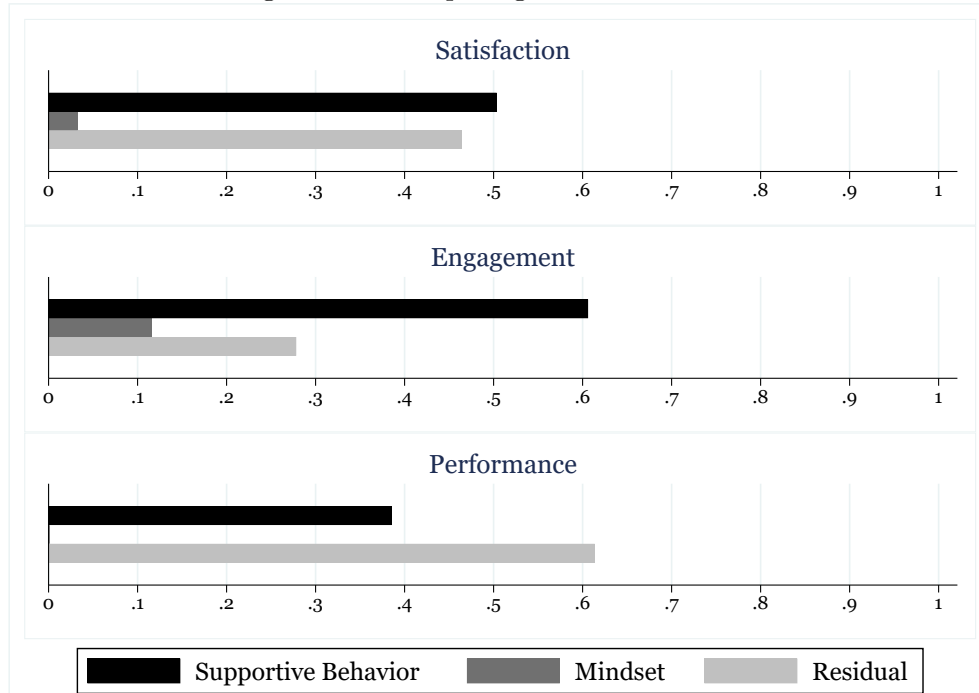
Notes: Percent of treatment effect explained by mediators. We only include the blocking variable as a control variable in the regression and cluster standard errors at the team level.

To be able to interpret the results of the mediation analysis as causal, we have to assume sequential ignorability (Imai et al., 2010). This assumption consists of two conditions: Condition i) requires that potential outcomes are independent of treatment status; and Condition ii) requires independence of observed and treatment-induced, unobserved

---

[18]In addition, we tested whether leader's mindset affects their behavior but we did not find evidence for such a mediation effect (see Table A.4 in the appendix).

[19]For simplicity, we assume that the mapping of the mediators and control variables on the outcome is not affected by treatment: This assumption is supported by a likelihood-ratio test showing no significant improvement of the model fit when allowing for mappings to vary by treatment.

mediators. While condition i) holds by design in a randomized experiment, condition ii) requires the additional assumption of the exogeneity of mediators conditional on the treatment and controls.

As shown in Figure 3, we find that supportive leadership behaviors are the main mediator. Precisely 50% of the treatment effect on satisfaction, 61% of the effects on engagement, and 39% of the effect on performance are mediated by leadership behaviors. However, only the effects on satisfaction and engagement are statistically significant. Mindset does not have a significant mediation effect on any of the outcome variables, and the estimate is close to 0 for satisfaction and performance as the dependent variable and 14% with engagement as the dependent variable.

Using supportive leadership behaviors, as stated by the employees as the mediator, not only captures whether the leader behaved in a supportive way but also whether the employees perceived the behavior as supportive. While this is an advantage, as the perception might be important for satisfaction and engagement, the analysis also suffers from the common variance problem as the mediators and the outcome variable are measured in the same survey. As a robustness check, we repeated the mediation analysis using the average of supportive leader behavior reported by the other team members as an alternative mediator. Using this measure, the share of the treatment effect that is explained by supportive behavior decreases to 14% for satisfaction ($p = 0.086$) and to 10% for engagement ($p = 0.323$).

6.5 Heterogeneous Treatment Effects

In addition to the average treatment effects reported above, it is also interesting to investigate heterogeneous treatment effects across leader characteristics. Given our sample size, we have, however, limited power to conduct such subgroup analysis and, therefore, consider the evidence provided below as suggestive. We have pre-registered the following subgroups: leader's gender, leader tenure (median split), coaching behavior at baseline (median split), and leader mindset at baseline (median split). We focus on gender differences in this section and present detailed results for the other groups in Appendix A.

We investigate gender differences in treatment effects because previous literature suggests gender differences in leadership styles. Specifically, they find that female leaders are more likely to have features of transformational leadership (Bass et al., 1996; Eagly and Carli, 2003). This is especially true for the individual consideration aspect of transformational leadership (Eagly and Carli, 2003), which is related to our definition of supportive leadership behavior. In addition, a recent study by An et al. (2020) found

that female leaders tend to be more democratic and participatory, while male leaders are more independent and assertive. Given that our treatment encourages leaders to be supportive, the effects might be larger for male leaders because they may be less inclined to engage in supportive behaviors without the treatment.

Table 4: Gender differences in treatment effects

	(1) Supportive Behavior	(2) Mindset	(3) Satisfaction	(4) Engagement	(5) Performance
Treatment	0.655***	0.375*	0.496***	0.398***	0.048
	(0.197)	(0.215)	(0.135)	(0.122)	(0.197)
	[<0.01]	[0.08]	[<0.01]	[0.01]	[0.79]
Female leader	0.543**	-0.325	0.571***	0.349**	-0.201
	(0.255)	(0.314)	(0.150)	(0.146)	(0.217)
Treated x Female l.	-0.891**	0.716*	-0.781***	-0.800***	0.281
	(0.373)	(0.420)	(0.246)	(0.242)	(0.336)
	[0.04]	[0.08]	[<0.01]	[<0.01]	[0.39]
Constant	0.795	0.078	-0.048	0.278	-0.390
	(0.535)	(0.583)	(0.451)	(0.397)	(0.601)
Observations	129	127	302	302	130
Adjusted R^2	0.078	0.067	0.043	0.019	0.082
Block FE	YES	YES	YES	YES	YES
Controls	YES	YES	YES	YES	YES

Notes: OLS regression with block-fixed effects. Numbers in parentheses indicate standard errors; numbers in brackets indicate family-wise rejection probabilities adjusting for multiple-hypotheses testing using the Westfall and Young procedure (Jones et al., 2019). In columns (1), (2) and (5) we use group level from data 130 groups and report robust standard errors. In columns (3) and (4), we use individual-level data from 302 employees who completed the survey after the work task and cluster standard errors by team. The dependent variable is the proxy for the extent to which the leader engaged in supportive leadership practices as reported by the work team in column (1), the leader's self-reported mindset in column (2), the employees' self-reported satisfaction with the group task in column (3), the employees' self-reported engagement in column (4) and a compound measure of output quality with equal weights on relevance and originality in column (5). Dependent variables are normalized relative to the distribution of the control group. The sample size under 130 is due to missing observations of the dependent variable. Controls include team size, seniority of the leader and average seniority of the team members, the leader's level, and a proxy for coaching behavior by the leader as reported by his or her employees. We include a dummy for missing observations for coaching behavior. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

In Table 4, we show the manipulation check in columns (1) and (2) and the treatment effect on the main outcome variables in columns (3) - (5). We see that the treatment effect on supportive behaviors is significantly larger for male leaders than for female leaders. The same holds true for the effects on satisfaction and engagement. We find reverse results for leader mindset, which might partly be driven by the fact that mindset is a self-reported measure, and women might be more prone to provide the "desired" answers (Quidt et al., 2018).

These results suggest that the treatment was more effective for male than for female

leaders. However, it is important to mention that the results for female leaders are based on a fairly small sample, as more men than women have leadership positions in our sample.[20]

Concerning the heterogeneity analyses across the other subgroups, we do not find conclusive effects. We find that the treatment improves performance for experienced leaders and significantly more so than for inexperienced ones. However, we do not see heterogeneous effects for any of the other outcomes or manipulation checks (see Table A.11). We do not find evidence for heterogeneous treatment effects based on baseline coaching behavior or mindset (see Tables A.12 and A.13).

6.6 Long-Term Effects

As the main objective of this study was to investigate how an exogenous change in leaders' supportiveness affects employees' job satisfaction, engagement and team performance in the one-hour meeting immediately after the intervention, we did not design the study to measure the long-term effects of the intervention. Still, as the experiment is part of a large project investigating leadership behavior, we do also have survey data from leaders and employees one month after the intervention. In this survey, we have asked leaders about their mindset and leadership behaviors using the Multifactor Leadership Questionnaire, MLQ, (Bass and Avolio, 1997) to measure transformative leadership style. In addition, we asked the leaders' employees about their leadership style using the MLQ as well as a 10-item coaching behavior scale developed by Heslin et al. (2006). 109 leaders (84 %) responded, out of which one only partially answered the survey.[21]

Table 5 shows that both the leaders' self-reported mindset (column 1) and transformative leadership behavior (column 2) are significantly higher in the treatment than in the control group, suggesting that leaders at least remember some of the training's content a month later. The second result, however, is not robust to MHT correction. In columns (3) and (4), we look at the leaders' behavior as reported by their real employees, which reduces the sample to 78 leaders.[22] While it is reassuring that the point estimate for the employee-reported behavior is similar to the leader's self-report (comparing columns 2 and 3), we do not find a significant treatment difference.

---

[20]We observe 37 (18 in control and 19 in treatment) female leaders and 93 (49 in control and 44 in treatment) male leaders.

[21]Attrition is balanced between treatment and control ($p = 0.346$, Fisher's exact test).

[22]We were able to match employee answers to slightly more control than treatment leaders (69 compared to 52 %, $p = 0.073$, Fisher's exact test). However, this slight imbalance was already present in the baseline and is therefore unlikely caused by treatment.

Table 5: Long-term effects on leader behavior and mindset

	(1) Growth Mindset	(2) Transformative Leadership	(3) Transformative Leadership(E)	(4) Coaching (E)
Treatment	0.558***	0.259**	0.185	0.132
	(0.178)	(0.129)	(0.238)	(0.147)
	[0.01]	[0.18]	[0.57]	[0.57]
Observations	109	108	78	78
Adjusted R^2	0.580	0.536	0.394	0.681
Block FE	YES	YES	YES	YES
Controls	YES	YES	YES	YES

Notes: OLS regression with block-fixed effects. Numbers in parentheses indicate robust standard errors; numbers in brackets indicate family-wise rejection probabilities adjusting for multiple-hypotheses testing using the Westfall and Young procedure (Jones et al., 2019). In columns (1) & (2) we use leaders' self-reported behaviors of the 109 leaders who answered both the baseline and the endline survey, one of these leaders only partially. In columns (3) & (4), we use averages of what employees say about their leaders. We have received answers on 78 leaders from their employees. All regressions control for baseline measures of the dependent variable. In addition, we control for leader gender, team size, seniority of the leader and average seniority of the team members, the leader's level, and a proxy for coaching behavior by the leader as reported by his or her employees. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

7 Discussion

7.1 Selection into the Study

Participation in the study was voluntary. We are, therefore, looking at the selected sample of employees and leaders who were willing to agree to participate in the group discussions. As it was very important for our cooperation partner to get input from as many employees as possible, they implemented engagement measures such as advertising the project on the company's Microsoft Teams pages and sharing a video of the CEO promoting the project. Around 70% of the leaders and 35% of the employees signed up, summing up to a total sign-up rate of around 39%.

As we, unfortunately, do not have any information on employees and leaders who did not participate, we cannot say much about the characteristics of these employees compared to our sample. We do, however, observe two interesting patterns. First, we can investigate how integration into the corporation affects participation. The corporation consists of several sub-companies, two of which were merged with the corporation just very shortly before our study. We see that within these sub-companies, the participation rate is only 29% which is substantially lower than the overall participation rate of 39%. This difference indicates that employees who are less integrated into the corporation were less likely to participate.

Second, we can examine whether the type of job a person usually works in affects the likelihood of participation. The corporation distinguishes between the following areas: energy, telecommunication, infrastructure, and others. Depending on the business area, the employees might have very different workdays. For example, employees working in infrastructure are more likely to work in the field, while those working in telecommunication are more likely to have office jobs. Depending on the structure of one's workday, it might be more or less feasible and natural to participate in a one-hour online meeting. Our participation rates show that participation is lowest in sub-companies working in infrastructure (38%) and highest in sub-companies in telecommunication (64%).

Lastly, it seems likely that personal characteristics and preferences to participate in group discussions and contribute to the company's development drove the decision to participate. In all, we expect that our voluntary sign-up mechanism resulted in a sample of participants that are more likely to feel attached to the company, have an office job, and care about the company's development than the remaining employees. In addition, employees in our sample probably do not have a strong aversion to participating in online group discussions with strangers. Given the nature of our intervention, it is likely that these selection effects limited the scope to improve employees' performance, suggesting that we are looking at a lower bound of effects. However, we do not have data to investigate this conjecture.

7.2 Survey Measures

Two of our three main outcome variables are based on survey questions. Using survey questions is a quick, cheap, and easily implementable way to collect measures of individual's opinions, preferences, or psychological states (Duckworth and Yeager, 2015; Stantcheva, 2023). In field settings, it is often not possible to use economic experiments due to time and organizational constraints. In addition, collecting information on real-life decisions is often not feasible, leaving survey questions as the only choice to measure outcomes.

However, survey measures also have some clear limitations. One corporation with survey measures is the lack of incentives to answer truthfully, allowing respondents to be sensitive to experimenter demand effects, potentially biasing the treatment estimates. Following suggestions by, e.g., Stantcheva (2023), we limit the scope for experimenter demand effects by keeping employees blind to treatment. In fact, they are not even aware that there are different treatments, making it unlikely that the treatment group responds differently to the control group because of experimenter demand effects.

A second concern is the relevance of survey measures for real-life decisions. There

27

is ample evidence that self-reported survey answers are good predictors of real-life outcomes. For example, self-reported job satisfaction has been found to predict quitting behavior (Green, 2010), and survey measures of work engagement are positively correlated with performance (e.g., Judge et al., 2001; Rich et al., 2010). Similarly, Falk et al. (forthcoming) developed a survey module to measure a set of economic preferences and advocate that this module is a good alternative to experimental measures.

Lastly, answers to survey questions might be biased due to social desirability, acquiescence bias, response order bias, moderacy bias, or extreme response bias (Stantcheva, 2023). These biases might affect our measures of the outcomes, for example, if a large share of employees thinks it is desirable to report to be very engaged and therefore reports a higher level of engagement than their truly perceived level. However, randomization ensures that, in expectation, the share of employees having this bias is the same in treatment and control, allowing us to obtain an unbiased measure of the treatment effect.

7.3 Counterfactual

Leaders both in the control and treatment groups received the training during the hour prior to the online meeting. The only difference in the training was the 20-minute insertion on supportive leadership behaviors. We did not want to lengthen the training of the control group with useless placebo material for ethical reasons. Moreover, adding material with valuable content could have changed the behavior of the control group confounding the treatment differences. As a consequence, the training was 20 minutes shorter for controlled leaders. This could have led to them being more relaxed and rested when implementing the meeting. Alternatively, it could have led to them to lose their focus on the task at hand, which could in itself cause a treatment difference. This is a possible limitation of our study. Notably, however, the sizable impacts on supportive behaviors suggest that a critical treatment contrast is in supportive behaviors.

7.4 Experimenter Demand and Combined Treatment

In this paper, we investigate how employees who are blind to treatment are affected by supportive leadership behaviors. Therefore, we designed a leadership training that provides very concrete instructions on how to use supportive leadership behaviors. As a result, by design our change in leadership behavior is likely due to experimenter demand. However, the major contribution of this paper is not how to change leaders' behavior but rather to investigate how employees respond to supportive behaviors. Previous studies

have obtained experimental manipulation in leadership behaviors by using actors as leaders in more controlled laboratory settings (see, e.g., Antonakis et al., 2022).

Our treatment experimentally manipulated supportive leadership behaviors by targeting three barriers: 1) they do not believe that all employees can contribute to the task at hand (leader's mindset): 2) they do not know supportive leadership behaviors matter for employees' contribution; 3) they have limited awareness on how to support employees' contribution. It may be that not all these barriers were present in our field context. For example, even if a leader is aware of supportive behaviors, she will not necessarily use these behaviors if she does not believe that the employees have the potential to contribute to the task at hand or does not believe that leadership support matters for the employees' contribution.

If the focus of the paper were on leadership training, the optimal design would have been a multi-arm RCT where we could test how important it is to break down the different barriers alone and in combination. In the field context, with limited access to comparable leaders, obtaining sufficient power is very difficult. Since the focus of our paper was to investigate how employees respond to supportive behaviors, and we only had 130 leaders, we chose an experimental design with two arms investigating the effects of the combined treatment. Nevertheless, we gained some insights into the relevance of the different barriers in a suggestive mediation analysis. It suggests that barrier 1 is not a critical barrier in our context. However, the evidence suggests that employee-observed leadership behavior is a mediator. Our data do not allow us to investigate if this mediation is due to targeting barriers 2 or 3, or both.

7.5 Relationship between Engagement and Performance

In Section 2, we hypothesize that our treatment will improve team performance by encouraging all team members to contribute. We do not find evidence in support of this hypothesis. While we argue that this might be because we do not have enough power to detect effects on the team level, another potential explanation is that more engagement does not lead to better performance on our team performance measure. The reason is that too many ideas may make it hard to identify the most useful input for the firm. As discussed in Section 2, a recent paper by Charness et al. (2020) suggests that team performance suffers from too much communication. The paper shows that teams perform better in a problem-solving task if there is a (small) cost to communication because it reduces the number of unproductive messages sent. Our treatment does the exact opposite by reducing the psychological barriers of contributing. Due to lack of

power, we cannot provide evidence that reducing the psychological barrier improved or worsened team performance.

8 Conclusion

In a constantly changing workplace, employees frequently have to adapt to new tasks and collaborate with new people. In these situations, many individuals fail to contribute to their full potential because they are concerned that their colleagues will think that their ideas, reflections, or efforts are irrelevant or stupid (Edmondson, 1999; Newman et al., 2017). For employees to remain engaged and productive in these settings, leaders can play a crucial role by encouraging, valuing, and assuring their effort to handle the new challenges. We investigated this hypothesis in partnership with a large corporation that provided access to a real field setting where the employees had to solve a new task with people with whom they had never worked before, under new leadership.

Our field experiment randomly treated half of the leaders with a brief training promoting leadership behaviors that encourage, assure, and value employee efforts. This training was integrated as a 20-minute extension of the training that all leaders received. Both leaders and employees were blind to treatment.

We found that the treatment was successful in experimentally manipulating supportive leadership behaviors during the team meetings (+0.40 sd). Moreover, in groups with treated leaders, employees reported higher work satisfaction (+0.28 sd) and engagement (+0.18 sd). A mediation analysis suggested that these effects were indeed driven by the change in supportive leadership behavior. This finding is relevant because a large literature suggests that satisfaction and engagement are strong predictors of performance (e.g., Judge et al., 2001; Rich et al., 2010). In addition, job satisfaction is an important predictor of turnover (Green, 2010). This suggests that supportive leadership should be valued both when recruiting leaders and in leadership education and training programs.

We measured team performance during the one-hour meeting and found that the treatment effect was 0.13 standard deviations, but not significant. Since performance was a downstream measure, i.e., the consequence of a change in satisfaction and engagement, it is not surprising that the results are smaller than those for satisfaction and engagement. In addition, we observe performance only at the team level. The lack of power to measure downstream results, which is a result of practical constraints in recruitment, is a limitation of our study. Future research should be sufficiently powered to more carefully investigate impacts on downstream outcomes. Moreover, we need research that can investigate the cumulative effects of supportive leadership on performance.

Our field experiment provides some of the first causal evidence on the effect of supportive leadership behaviors. In addition, a major innovation in this paper is the development of a data-collection procedure in partnership with organizations, which gives us direct access to the leaders, allows us to experimentally manipulate leaders' behaviors, and ensures task comparability across leaders. It is possible to create a win-win situation by partnering with organizations that need input from all employees on a question of strategic importance. We hope this can inspire other researchers to help fill the large research gap, where causal evidence on the effect of leadership behaviors is scant.

References

Alan, Sule, Gozde Corekcioglu, and Matthias Sutter (Sept. 2022). "Improving Workplace Climate in Large Corporations: A Clustered Randomized Intervention*". *The Quarterly Journal of Economics* 138.1, pp. 151–203.

An, Seung-Ho, Kenneth J. Meier, Jacob Ladenburg, and Niels Westergård-Nielsen (2020). "Leadership and Job Satisfaction: Addressing Endogeneity With Panel Data From a Field Experiment". *Review of Public Personnel Administration* 40.4, pp. 589–612.

Antonakis, John (2017). "On Doing Better Science: From Thrill of Discovery to Policy Implications". *The Leadership Quarterly* 28.1, pp. 5–21.

Antonakis, John, Samuel Bendahan, Philippe Jacquart, and Rafael Lalive (2014). "Causality and Endogeneity: Problems and Solutions". *The Oxford Handbook of Leadership and Organizations*. Oxford library of psychology. New York, NY, US: Oxford University Press, pp. 93–117.

Antonakis, John, Giovanna d'Adda, Roberto A. Weber, and Christian Zehnder (2022). ""Just Words? Just Speeches?" On the Economic Value of Charismatic Leadership". *Management Science* 68.9, pp. 6355–6381.

Bass, B. M. and B. J. Avolio (1997). *Full Range Leadership Development: Manual for the Multifactor Leadership Questionnaire*. Redwood City, CA: Mindgarden.

Bass, Bernard M., Bruce J. Avolio, and Leanne Atwater (1996). "The Transformational and Transactional Leadership of Men and Women". *Applied Psychology* 45.1, pp. 5–34.

Bass, Bernard M., Bruce J. Avolio, Dong I. Jung, and Yair Berson (2003). "Predicting Unit Performance by Assessing Transformational and Transactional Leadership". *Journal of Applied Psychology* 88.2, pp. 207–218.

Bender, Stefan, Nicholas Bloom, David Card, John Van Reenen, and Stefanie Wolter (2018). "Management Practices, Workforce Selection, and Productivity". *Journal of Labor Economics* 36.S1, 371–S409.

Bettinger, Eric, Sten Ludvigsen, Mari Rege, Ingeborg F. Solli, and David Yeager (2018). "Increasing Perseverance in Math: Evidence from a Field Experiment in Norway". *Journal of Economic Behavior & Organization* 146, pp. 1–15.

Blader, Steven, Claudine Gartenberg, and Andrea Prat (2019). "The Contingent Effect of Management Practices". *The Review of Economic Studies* 87.2, pp. 721–749.

Bloom, Nicholas, Erik Brynjolfsson, Lucia Foster, Ron Jarmin, Megha Patnaik, Itay Saporta-Eksten, and John Van Reenen (2019). "What Drives Differences in Management Practices?" *American Economic Review* 109.5, pp. 1648–83.

Bloom, Nicholas, Benn Eifert, Aprajit Mahajan, David McKenzie, and John Roberts (2013). "Does Management Matter? Evidence from India". *The Quarterly Journal of Economics* 128.1, pp. 1–51.

Bradler, Christiane, Robert Dur, Susanne Neckermann, and Arjan Non (2016). "Employee Recognition and Performance: A Field Experiment". *Management Science* 62.11, pp. 3085–3099.

Bruhn, Miriam, Dean Karlan, and Antoinette Schoar (2018). "The Impact of Consulting Services on Small and Medium Enterprises: Evidence from a Randomized Trial in Mexico". *Journal of Political Economy* 126.2, pp. 635–687.

Castro, Silvia, Florian Englmaier, and Maria Guadalupe (2022). *Fostering Psychological Safety in Teams: Evidence from an RCT*. SSRN Discussion Paper.

Chandrasekhar, Arun G, Benjamin Golub, and He Yang (2018). *Signaling, Shame, and Silence in Social Learning*. Working Paper 25169. National Bureau of Economic Research.

Charness, Gary, David J. Cooper, and Zachary Grossman (2020). "Silence is Golden: Team Problem Solving and Communication Costs". *Experimental Economics* 23.3, pp. 668–693.

Chiu, Chi-yue, Carol S Dweck, Jennifer Yuk-yue Tong, and Jeanne Ho-ying Fu (1997). "Implicit Theories and Conceptions of Morality." *Journal of Personality and Social Psychology* 73.5, p. 923.

Deming, David J. (2017). "The Growing Importance of Social Skills in the Labor Market". *The Quarterly Journal of Economics* 132.4, pp. 1593–1640.

Derue, D. Scott, Jennifer D. Nahrgang, Ned Wellman, and Stephen E. Humphrey (2011). "Trait and Behavioral Theories of Leadership: An Integration and Meta-Analytic Test of Their Relative Validity". *Personnel Psychology* 64.1, pp. 7–52.

Dolbier, Christyn L., Judith A. Webster, Katherin T. McCalister, Mark W. Mallon, and Mary A. Steinhardt (2005). "Reliability and Validity of a Single-item Measure of Job Satisfaction". *American Journal of Health Promotion* 19.3, pp. 194–8.

Duckworth, Angela L. and David Scott Yeager (2015). "Measurement Matters: Assessing Personal Qualities Other Than Cognitive Ability for Educational Purposes". *Educational Researcher* 44.4, pp. 237–251.

Dvir, Taly, Dov Eden, Bruce J. Avolio, and Boas Shamir (2002). "Impact of Transformational Leadership on Follower Development and Performance: A Field Experiment". *The Academy of Management Journal* 45.4, pp. 735–744.

Dweck, Carol (2006). *Mindset*. London: Robinson.

Eagly, Alice H and Linda L Carli (2003). "The Female Leadership Advantage: An Evaluation of the Evidence". *The Leadership Quarterly* 14.6, pp. 807–834.

Edmondson, Amy (1999). "Psychological Safety and Learning Behavior in Work Teams". *Administrative Science Quarterly* 44.2, pp. 350–383.

Edmondson, Amy C. and Zhike Lei (2014). "Psychological Safety: The History, Renaissance, and Future of an Interpersonal Construct". *Annual Review of Organizational Psychology and Organizational Behavior* 1.1, pp. 23–43.

Englmaier, Florian, Stefan Grimm, Dominik Grothe, David Schindler, and Simeon Schudy (2021). *The Value of Leadership: Evidence from a Large-scale Field Experiment*. CESifo Working Paper 9273.

European Foundation for the Improvement of Living and Working Conditions, European Centre for the Development of Vocational Training (2020). *European Company Survey, 2019.* http://doi.org/10.5255/UKDA-SN-8691-1.

Falk, Armin, Anke Becker, Thomas Dohmen, David Huffman, and Uwe Sunde (forthcoming). "The Preference Survey Module: A Validated Instrument for Measuring Risk, Time, and Social Preferences". *Management Science* 0.0, null.

Fest, Sebastian, Ola Kvaløy, Petra Nieken, and Anja Schöttner (2021). "How (Not) to Motivate Online Workers: Two Controlled Field Experiments on Leadership in the Gig Economy". *The Leadership Quarterly* 32.6, p. 101514.

Fryer Roland G., Jr. (2014). "Injecting Charter School Best Practices into Traditional Public Schools: Evidence from Field Experiments". *The Quarterly Journal of Economics* 129.3, pp. 1355–1407.

Gosnell, Greer K., John A. List, and Robert D. Metcalfe (2020). "The Impact of Management Practices on Employee Productivity: A Field Experiment with Airline Captains". *Journal of Political Economy* 128.4, pp. 1195–1233.

Green, Francis (2010). "Well-being, Job satisfaction and Labour Mobility". *Labour Economics* 17.6, pp. 897–903.

Heckman, James, Rodrigo Pinto, and Peter Savelyev (2013). "Understanding the Mechanisms through Which an Influential Early Childhood Program Boosted Adult Outcomes". *American Economic Review* 103.6, pp. 2052–86.

Heckman, James J. and Rodrigo Pinto (2015). "Econometric Mediation Analyses: Identifying the Sources of Treatment Effects from Experimentally Estimated Production Technologies with Unmeasured and Mismeasured Inputs". *Econometric Reviews* 34.1-2, pp. 6–31.

Heinz, Matthias, Sabrina Jeworrek, Vanessa Mertins, Heiner Schumacher, and Matthias Sutter (Apr. 2020). "Measuring the Indirect Effects of Adverse Employer Behaviour on

Worker Productivity: A Field Experiment". *The Economic Journal* 130.632, pp. 2546–2568.

Heslin, Peter A., Don Vandewalle, and Gary P. Latham (2006). "Keen to Help? Managers' Implicit Person Theories and Their Subsequent Employee Coaching". *Personnel Psychology* 59.4, pp. 871–902.

Hirak, Reuven, Ann Chunyan Peng, Abraham Carmeli, and John M. Schaubroeck (2012). "Linking Leader Inclusiveness to Work Unit Performance: The Importance of Psychological Safety and Learning from Failures". *The Leadership Quarterly* 23.1, pp. 107–117.

Hoffman, Mitchell and Steven Tadelis (2021). "People Management Skills, Employee Attrition, and Manager Rewards: An Empirical Analysis". *Journal of Political Economy* 129.1, pp. 243–285.

Ichniowski, Casey, Kathryn Shaw, and Giovanna Prennushi (1997). "The Effects of Human Resource Management Practices on Productivity: A Study of Steel Finishing Lines". *The American Economic Review* 87.3, pp. 291–313.

Imai, Kosuke, Luke Keele, and Dustin Tingley (2010). "A General Approach to Causal Mediation Analysis." *Psychological Methods* 15.4, p. 309.

Jacobsen, Christian Bøtcher, Lotte Bøgh Andersen, Anne Bøllingtoft, and Tine Louise Mundbjerg Eriksen (2022). "Can Leadership Training Improve Organizational Effectiveness? Evidence from a Randomized Field Experiment on Transformational and Transactional Leadership". *Public Administration Review* 82.1, pp. 117–131.

Jones, Damon, David Molitor, and Julian Reif (Aug. 2019). "What do Workplace Wellness Programs do? Evidence from the Illinois Workplace Wellness Study". *The Quarterly Journal of Economics* 134.4, pp. 1747–1791.

Judge, Timothy, Carl Thoresen, Joyce Bono, and Gregory Patton (2001). "The Job Satisfaction–Job Performance Relationship". *Psychological Bulletin* 127, pp. 376–407.

Judge, Timothy A and Ronald F Piccolo (2004). "Transformational and Transactional Leadership: A Meta-analytic Test of Their Relative Validity". *Journal of Applied Psychology* 89.5, p. 755.

Lazear, Edward P. (2000). "Performance Pay and Productivity". *The American Economic Review* 90.5, pp. 1346–1361.

Lazear, Edward P., Kathryn L. Shaw, and Christopher T. Stanton (2015). "The Value of Bosses". *Journal of Labor Economics* 33.4, pp. 823–861.

Leroy, Hannes, Bart Dierynck, Frederik Anseel, Tony Simons, Jonathon R. B. Halbesleben, Deirdre McCaughey, Grant T. Savage, and Luc Sels (2012). "Behavioral Integrity for

Safety, Priority of Safety, Psychological Safety, and Patient Safety: A Team-Level Study". *Journal of Applied Psychology* 97.6, pp. 1273–1281.

Liu, Jiayan, Oi-Ling Siu, and Kan Shi (2010). "Transformational Leadership and Employee Well-Being: The Mediating Role of Trust in the Leader and Self-Efficacy". *Applied Psychology* 59.3, pp. 454–479.

Liu, Songbo, Jia Hu, Yuhui Li, Zhen Wang, and Xiaoshuang Lin (2014). "Examining the Cross-Level Relationship between Shared Leadership and Learning in Teams: Evidence from China". *The Leadership Quarterly* 25.2, pp. 282–295.

Manthei, Kathrin, Dirk Sliwka, and Timo Vogelsang (2022). "Talking About Performance or Paying for It? A Field Experiment on Performance Reviews and Incentives". *Management Science* 0.0, null.

Martin, Robin, Olga Epitropaki, and Laurie O'Broin (2017). "Methodological Issues in Leadership Training Research: In Pursuit of Causality". *Methodological Challenges and Advances in Managerial and Organizational Cognition* 2, pp. 73–94.

Martin, Robin, David J. Hughes, Olga Epitropaki, and Geoff Thomas (2020). "In pursuit of causality in leadership training research: A Review and Pragmatic Recommendations". *The Leadership Quarterly* 32.5, p. 101375.

Meslec, Nicoleta, Petru L. Curseu, Oana C. Fodor, and Renata Kenda (2020). "Effects of Charismatic Leadership and Rewards on Individual Performance". *The Leadership Quarterly* 31.6, p. 101423.

Nemanich, Louise A. and Dusya Vera (2009). "Transformational Leadership and Ambidexterity in the Context of an Acquisition". *The Leadership Quarterly* 20.1, pp. 19–33.

Nembhard, Ingrid M. and Amy C. Edmondson (2006). "Making it Safe: the Effects of Leader Inclusiveness and Professional Status on Psychological Safety and Improvement Efforts in Health Care Teams". *Journal of Organizational Behavior* 27.7, pp. 941–966.

Newman, Alexander, Ross Donohue, and Nathan Eva (2017). "Psychological Safety: A Systematic Review of the Literature". *Human Resource Management Review* 27.3, pp. 521–535.

Ortega, Aida, Piet Van den Bossche, Miriam Sánchez-Manzanares, Ramon Rico, and Francisco Gil (2014). "The Influence of Change-Oriented Leadership and Psychological Safety on Team Learning in Healthcare Teams". *Journal of Business and Psychology* 29.2, pp. 311–321.

Quidt, Jonathan de, Johannes Haushofer, and Christopher Roth (2018). "Measuring and Bounding Experimenter Demand". *American Economic Review* 108.11, pp. 3266–3302.

Rege, Mari, Paul Hanselman, Ingeborg Solli, Carol Dweck, Sten Ludvigsen, Eric Bettinger, Robert Crosnoe, Chandra Muller, Gregory Walton, Angela Duckworth, and David Yeager (2020). "How Can We Inspire Nations of Learners? Investigating Growth Mindset and Challenge-seeking in Two Countries". *American Psychologist* 76.5, 755–767.

Rich, Bruce Louis, Jeffrey A. Lepine, and Eean R. Crawford (2010). "Job Engagement: Antecedents and Effects on Job Performance". *The Academy of Management Journal* 53.3, pp. 617–635.

Sandvik, Jason J., Richard E. Saouma, Nathan T. Seegert, and Christopher T. Stanton (Apr. 2020). "Workplace Knowledge Flows". *The Quarterly Journal of Economics* 135.3, pp. 1635–1680.

Stantcheva, Stefanie (2023). "How to Run Surveys: A Guide to Creating Your Own Identifying Variation and Revealing the Invisible". *Annual Review of Economics* 15.1, null.

Walumbwa, Fred O. and John Schaubroeck (2009). "Leader Personality Traits and Employee Voice Behavior: Mediating Roles of Ethical Leadership and Work Group Psychological Safety". *Journal of Applied Psychology* 94.5, pp. 1275–1286.

Wang, Gang, In-Sue Oh, Stephen H. Courtright, and Amy E. Colbert (2011). "Transformational Leadership and Performance Across Criteria and Levels: A Meta-Analytic Review of 25 Years of Research". *Group & Organization Management* 36.2, pp. 223–270.

Wanous, John P., Arnon E. Reichers, and Michael J. Hudy (1997). "Overall Job Satisfaction: How Good are Single-item Measures?" *Journal of Applied Psychology* 82.2, pp. 247–252.

Weidmann, Ben and David J. Deming (2021). "Team Players: How Social Skills Improve Team Performance". *Econometrica* 89.6, pp. 2637–2657.

Yeager, David S., Carissa Romero, Dave Paunesku, Christopher S. Hulleman, Barbara. Schneider, Cintia Hinojosa, Hae Y. Lee, Joseph O'Brien, Kate Flint, Alice Roberts, Jill Trott, Daniel Greene, Gregory M. Walton, and Carol S. Dweck (2016). "Using Design Thinking to Improve Psychological Interventions: The Case of the Growth Mindset During the Transition to High School". *Journal of Educational Psychology* 108.3, pp. 374–391.

Yeager, David S. et al. (2019). "A National Experiment Reveals Where a Growth Mindset Improves Achievement". *Nature* 573.7774, pp. 364–369.

Online Appendix to "Effects of Supportive Leadership Behaviors on Employee Satisfaction, Engagement, and Performance: An Experimental Field Investigation"

Simone Haeckl and Mari Rege[*]

November 2023

[*]University of Stavanger, Department of Economics;
simone.haeckl-schermer@uis.no; mari.rege@uis.no;

Appendix

A. Additional figures and tables

Table A.1: Scales used by the evaluators to rate the quality of ideas.

Originality	
5	The proposed idea is novel, unique, exciting, brings up a new topic.
4	The proposed idea is unusual, exciting, brings up a new aspect.
3	The proposed idea is interesting.
2	The proposed idea is interesting but has already been discussed in the company (is not new).
1	The proposed idea is common, mundane, boring.
Relevance	
5	We will definitely include this idea in the strategy discussion.
4	We might include this idea in the strategy discussion.
3	We will not include this concrete idea but the general topic in the strategy discussion.
2	We will probably not consider this idea or the general topic in the strategy discussion.
1	We will definitely not consider this idea or the general topic in the strategy discussion.

Table A.2: Manipulation Checks (extended)

	Supportive behavior		Leader Mindset	
	(1)	(2)	(3)	(4)
Treatment	0.372**	0.397**	0.523***	0.590***
	(0.173)	(0.172)	(0.186)	(0.187)
Female Leader		0.098		0.039
		(0.194)		(0.220)
Team size		-0.254*		0.088
		(0.129)		(0.141)
Seniority (leader)		0.005		0.012
		(0.088)		(0.089)
Seniority (team)		-0.106		0.045
		(0.081)		(0.089)
High level leader		-0.010		-0.166
		(0.214)		(0.250)
Coaching		0.023		-0.134
		(0.088)		(0.114)
Coaching missing		-0.462**		-0.404*
		(0.180)		(0.223)
Constant	-0.022	0.973*	0.086	-0.090
	(0.239)	(0.524)	(0.309)	(0.590)
Observations	129	129	127	127
Adjusted R^2	0.012	0.042	0.049	0.050

Notes: OLS regression with block-fixed effects and robust standard errors in parentheses. The dependent variable is the proxy for the extent to which the leader engaged in supportive leadership practices as reported by the work team in columns (1) & (2) and the leader's self-reported mindset in columns (3) & (4). Dependent variables are normalized relative to the distribution of the control group and independent variables are normalized using the entire sample, except for team size which is not standardized. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

Figure A.1: Distribution of main variables
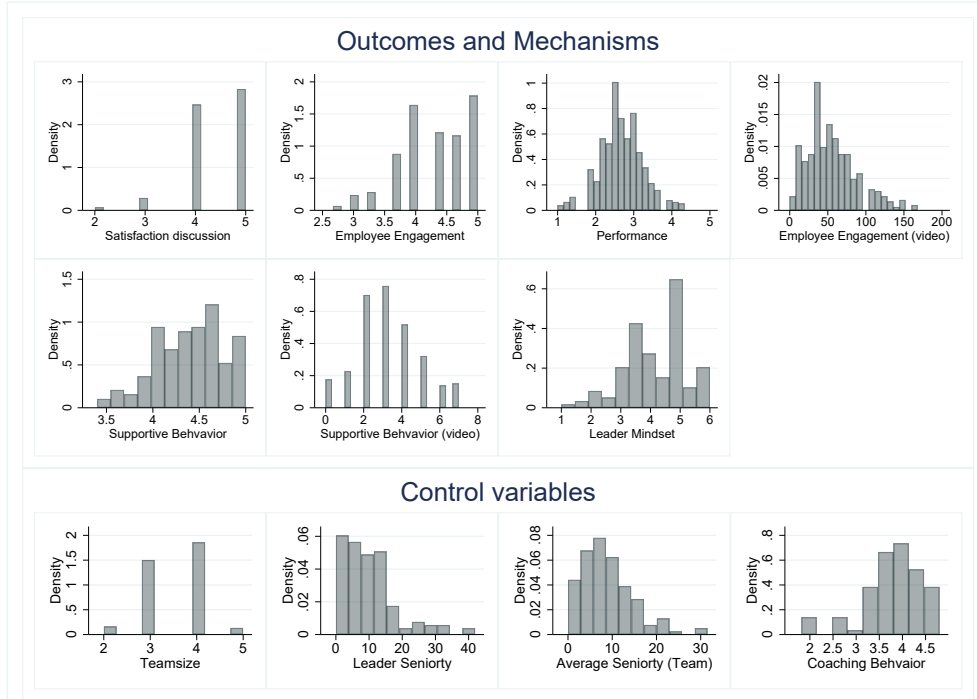
Notes: Histograms of main variables. Outcomes and mechanisms are measured ex-post while control variables are measured ex-ante. Coaching Behavior is based on the rating the leaders received from their real employees.

Table A.3: Manipulation Check using Supportive behavior (video)

	Leader Behavior	
	(1)	(2)
Treatment	0.747***	0.883***
	(0.201)	(0.214)
Observations	130	130
Adjusted R^2	0.101	0.153
Block FE	YES	YES
Controls	NO	YES

Notes: OLS regression with block-fixed effects and robust standard errors in parentheses. The dependent variable is the number of supportive leadership practices the leader engaged in based on the evaluation of the video. The dependent variable is normalized relative to the distribution of the control group. Controls include leader gender, team size, seniority of the leader and average seniority of the team members, the leader's level, and a proxy for coaching behavior by the leader as reported by his or her employees. We include a dummy for missing observations for coaching behavior in columns (2), (4), and (6) to keep the sample constant. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

Table A.4: Mindset as a Mediator

Total treatment effect on supportive behavior	0.372**	
	(0.169)	
Direct treatment effect on supportive behavior	0.383**	97%
	(0.170)	
Effect mediated by mindset	-0.011	3%
	(0.046)	

Notes: Decomposition of the treatment effect on supportive leader behaviors into the effect mediated by mindset and the direct effect. Values in parentheses indicate standard errors. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

Table A.5: Treatment effects on employee satisfaction, engagement, and group performance - extended

	Satisfaction		Engagement		Performance	
	(1)	(2)	(3)	(4)	(5)	(6)
Treatment	0.296**	0.278**	0.187*	0.176	0.121	0.131
	(0.119)	(0.125)	(0.109)	(0.113)	(0.157)	(0.155)
Female leader		0.181		-0.051		-0.059
		(0.137)		(0.134)		(0.165)
Team size		-0.065		-0.101		0.006
		(0.114)		(0.103)		(0.152)
Seniority (leader)		-0.032		0.065		-0.058
		(0.057)		(0.054)		(0.068)
High level leader		0.008		0.197		0.631***
		(0.139)		(0.131)		(0.199)
Female		0.196*		0.040		
		(0.108)		(0.119)		
Seniority (individual)		-0.033		-0.007		-0.023
		(0.049)		(0.054)		(0.071)
Coaching behavior		-0.038		-0.033		0.193**
		(0.056)		(0.059)		(0.092)
Coaching missing		-0.124		-0.099		-0.176
		(0.133)		(0.117)		(0.179)
Constant	-0.001	0.158	0.107	0.497	-0.418*	-0.455
	(0.177)	(0.460)	(0.107)	(0.417)	(0.250)	(0.593)
Observations	302	302	302	302	130	130
Adjusted R^2	0.018	0.017	0.001	-0.009	0.029	0.085
Block FE	YES	YES	YES	YES	YES	YES

Notes: OLS regression with block-fixed effects. In columns (1) - (4), we use individual-level data from 302 employees who completed the survey after the work task and cluster standard errors at the team level, in columns (5) & (6), we use group-level data from 130 groups and report robust standard errors. The dependent variable is the employees' self-reported satisfaction in columns (1) & (2) and employees' self-reported engagement in columns (3) & (4). In columns (5) & (6), the dependent variable is a compound measure of performance with equal weights on relevance and originality. Dependent variables are normalized relative to the distribution of the control group and independent variables are normalized using the entire sample, except for team size which is not standardized. Values in parentheses indicate standard errors. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

Table A.6: Treatment effects on employee satisfaction and engagement - weighted

| | Satisfaction | | Engagement | |
	(1)	(2)	(3)	(4)
Treatment	0.294**	0.266*	0.170	0.140
	(0.134)	(0.141)	(0.120)	(0.129)
Observations	302	302	302	302
Adjusted R^2	0.022	0.031	0.004	-0.001
Block FE	YES	YES	YES	YES
Controls	NO	YES	NO	YES

Notes: Probability-wrighted OLS regression with block-fixed effects. Observations are weighted by the inverse of the number of team members who responded, i.e., each team's responses are weighted equally. We use individual-level data from 302 employees who completed the survey after the work task and cluster standard errors at the team level. The dependent variable is the employees' self-reported satisfaction in columns (1) & (2) and employees' self-reported engagement in columns (3) & (4). Dependent variables are normalized relative to the distribution of the control group and independent variables are normalized using the entire sample, except for team size which is not standardized. Values in parentheses indicate standard errors. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

Table A.7: Treatment effect on satisfaction (predicted probabilities)

	(1) Disagree	(2) Neither agree nor disagree	(3) Agree	(4) Strongly agree
Treatment	0.01	0.02	0.39	0.58
	(0.007)	(0.010)	(0.042)	(0.043)
Control	0.02	0.10	0.45	0.43
	(0.012)	(0.031)	(0.037)	(0.038)
Observations	302	302	302	302
Block FE	YES	YES	YES	YES
Controls	YES	YES	YES	YES

Notes: Multinomial logistic regression with Satisfaction as the dependent variable. Coefficients show probabilities for each answer on the 5-point Likert scale for treatment and control. As no participant indicated that they "strongly disagree" this category could not be estimated. Controls include leader gender, team size, seniority of the leader and average seniority of the team members, the leader's level, and a proxy for coaching behavior by the leader as reported by his or her employees. We include a dummy for missing observations for coaching behavior. Values in parentheses indicate standard errors clustered at the team level.

Table A.8: Treatment effect on the probability of being very satisfied

| | (1) | (2) | (3) | (4) |
	at least agree	at least agree	strongly agree	strongly agree
Treatment	0.057*	0.061*	0.145**	0.133**
	(0.029)	(0.033)	(0.059)	(0.059)
Constant	0.861***	0.845***	0.485***	0.679***
	(0.055)	(0.134)	(0.063)	(0.197)
Observations	302	302	302	302
Adjusted R^2	0.005	-0.010	0.022	0.028
Block FE	YES	YES	YES	YES
Controls	NO	YES	NO	YES

Notes: Linear probability model. The dependent variable has value 1 if satisfaction is greater or equal to 4 in columns (1) and (2) or if satisfaction is greater than 4 in columns (3) and (4). Controls include leader gender, team size, seniority of the leader and average seniority of the team members, the leader's level, and a proxy for coaching behavior by the leader as reported by his or her employees. We include a dummy for missing observations for coaching behavior. Standard errors are clustered at the team level and presented in parentheses. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

Table A.9: Treatment effect on the probability of being very engaged

| | (1) | (2) | (3) | (4) |
	at least agree	at least agree	strongly agree	strongly agree
Treatment	0.120**	0.113**	0.176**	0.159*
	(0.047)	(0.051)	(0.079)	(0.083)
Constant	0.808***	0.937***	2.647***	3.005***
	(0.053)	(0.181)	(0.064)	(0.306)
Observations	302	302	302	302
Adjusted R^2	0.017	0.005	0.011	0.010
Block FE	YES	YES	YES	YES
Controls	NO	YES	NO	YES

Notes: In columns (1) and (2), the dependent variable has the value 1 if the index of engagement (average over all three items) is greater than or equal to 4 and 0 otherwise. In columns (3) & (4), the dependent variable has the value 3 if all items have a value greater than or equal to 4, value 2 if only two items have a value greater than or equal to 4, value 1 if only one item has a value greater than or equal to 4, and 0 otherwise. Controls include leader gender, team size, seniority of the leader and average seniority of the team members, the leader's level, and a proxy for coaching behavior by the leader as reported by his or her employees. We include a dummy for missing observations for coaching behavior. Standard errors are clustered at the team level and presented in parentheses. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

Table A.10: Treatment effects on engagement using the observational measure

	Engagement (video)	
	(1)	(2)
Treatment	0.163	0.072
	(0.175)	(0.146)
Observations	125	125
Adjusted R^2	0.022	0.366
Block FE	YES	YES
Controls	NO	YES

Notes: OLS regression with block-fixed effects and robust standard errors in parentheses. The dependent variable is the number of contributions in the meeting of the least active employee. The dependent variable is normalized relative to the distribution of the control group. Controls include leader gender, team size, seniority of the leader and average seniority of the team members, the leader's level, and a proxy for coaching behavior by the leader as reported by his or her employees. We include a dummy for missing observations for coaching behavior. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

Table A.11: Differences in treatment effects by leader seniority

	(1) Supportive Behavior	(2) Mindset	(3) Satisfaction	(4) Engagement	(5) Performance
Treatment	0.516**	0.666**	0.239	0.154	0.394**
	(0.247)	(0.262)	(0.177)	(0.133)	(0.189)
New leader	0.120	-0.000	0.094	-0.036	0.296
	(0.244)	(0.264)	(0.168)	(0.152)	(0.233)
Treated x New lead.	-0.277	-0.166	0.084	0.038	-0.603*
	(0.351)	(0.387)	(0.247)	(0.233)	(0.340)
Constant	0.931*	-0.081	0.115	0.504	-0.539
	(0.553)	(0.613)	(0.470)	(0.421)	(0.595)
Observations	129	127	302	302	130
Adj. R^2	0.039	0.045	0.017	-0.017	0.099
Block FE	YES	YES	YES	YES	YES
Controls	YES	YES	YES	YES	YES

Notes: OLS regression with block-fixed effects. Numbers in parentheses indicate standard errors. In columns (1), (2) and (5) we use group level from data 130 groups and report robust standard errors. In columns (3) and (4), we use individual-level data from 302 employees who completed the survey after the work task and cluster standard errors by team. To test for heterogeneous treatment effect by leader seniority we do a median split at 8 years of experience. 72 leaders are classified as new leaders. The dependent variable is the proxy for the extent to which the leader engaged in supportive leadership practices as reported by the work team in column (1), the leader's self-reported mindset in column (2), the employees' self-reported satisfaction with the group task in column (3), the employees' self-reported engagement in column (4) and a compound measure of output quality with equal weights on relevance and originality in column (5). Dependent variables are normalized relative to the distribution of the control group. The sample size under 130 is due to missing observations of the dependent variable. Controls include team size, leader gender, average seniority of the team members, the leader's level, and a proxy for coaching behavior by the leader as reported by his or her employees. We include a dummy for missing observations for coaching behavior. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

Table A.12: Differences in treatment effects by leader coaching behavior

	(1) Supportive Behavior	(2) Mindset	(3) Satisfaction	(4) Engagement	(5) Performance
Treatment	0.342	0.121	0.158	0.065	0.311
	(0.335)	(0.280)	(0.222)	(0.223)	(0.273)
Coaching leader	0.065	-0.826***	-0.074	-0.008	0.408
	(0.302)	(0.299)	(0.180)	(0.187)	(0.269)
Treated x Coach. l.	-0.080	0.970**	-0.036	-0.078	-0.259
	(0.423)	(0.444)	(0.293)	(0.283)	(0.432)
Constant	0.660	0.231	0.179	0.363	-0.844
	(0.690)	(0.736)	(0.548)	(0.518)	(0.792)
Observations	84	83	197	197	84
Adj. R^2	-0.044	0.158	-0.027	-0.052	0.074
Block FE	YES	YES	YES	YES	YES
Controls	YES	YES	YES	YES	YES

Notes: OLS regression with block-fixed effects. Numbers in parentheses indicate standard errors. In columns (1), (2) and (5) we use group level from data 130 groups and report robust standard errors. In columns (3) and (4), we use individual-level data from 302 employees who completed the survey after the work task and cluster standard errors by team. The sample is significantly reduced as we only observe baseline coaching behavior of 84 leaders. This leads to 84 observations for group-level data (columns 1,2 and 5) and 197 observations for individual-level regressions (columns 3 and 4). To test for heterogeneous treatment effect by leader seniority we do a median split at a coaching score of 3.85. 43 leaders are classified as coaching leaders. The dependent variable is the proxy for the extent to which the leader engaged in supportive leadership practices as reported by the work team in column (1), the leader's self-reported mindset in column (2), the employees' self-reported satisfaction with the group task in column (3), the employees' self-reported engagement in column (4) and a compound measure of output quality with equal weights on relevance and originality in column (5). Dependent variables are normalized relative to the distribution of the control group. The sample size under 130 is due to missing observations of the dependent variable. Controls include team size, leader gender, leader seniority and average seniority of the team members, and the leader's level. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

Table A.13: Differences in treatment effects by leader growth mindset

	(1) Supportive Behavior	(2) Mindset	(3) Satisfaction	(4) Engagement	(5) Performance
Treatment	0.450*	0.403*	0.409**	0.237	0.167
	(0.240)	(0.222)	(0.172)	(0.154)	(0.204)
Growth mind. leader	0.173	0.982***	0.145	0.193	-0.060
	(0.241)	(0.241)	(0.170)	(0.154)	(0.227)
Treated x Growth l.	-0.120	0.396	-0.281	-0.123	-0.079
	(0.337)	(0.298)	(0.233)	(0.215)	(0.320)
Constant	0.893	-0.714	0.116	0.428	-0.409
	(0.538)	(0.545)	(0.459)	(0.421)	(0.611)
Observations	129	127	302	302	130
Adj. R^2	0.030	0.358	0.015	-0.010	0.072
Block FE	YES	YES	YES	YES	YES
Controls	YES	YES	YES	YES	YES

Notes: OLS regression with block-fixed effects. Numbers in parentheses indicate standard errors. In columns (1), (2) and (5) we use group level from data 130 groups and report robust standard errors. In columns (3) and (4), we use individual-level data from 302 employees who completed the survey after the work task and cluster standard errors by team. To test for heterogeneous treatment effect by leader's growth mindset we do a median split at a score of 4. 59 leaders are classified as new having a growth mindset. The dependent variable is the proxy for the extent to which the leader engaged in supportive leadership practices as reported by the work team in column (1), the leader's self-reported mindset in column (2), the employees' self-reported satisfaction with the group task in column (3), the employees' self-reported engagement in column (4) and a compound measure of output quality with equal weights on relevance and originality in column (5). Dependent variables are normalized relative to the distribution of the control group. The sample size under 130 is due to missing observations of the dependent variable. Controls include team size, leader gender, leader seniority and average seniority of the team members, the leader's level, and a proxy for coaching behavior by the leader as reported by his or her employees. We include a dummy for missing observations for coaching behavior. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

B. Exploratory analysis: Additional engagement measures

The following analysis is clearly exploratory, which must be taken into consideration when interpreting any of the following results. To learn more about the intensity of the discussion, we have looked at the following additional measures: the total number of contributions (in addition to only looking at the contributions of the least active worker), the number of useful contributions (defined as contributions that are related to the topic and help with the development of ideas and coded by a research assistant, blind to treatment), the number of words spoken, and Herfindahl-indices to understand better how (un)balanced the discussion was.

To address questions about what leaders and employees say during the discussion, we have asked a research assistant (blind to treatment) to classify each contribution based on whether they fall into one of the following categories: new idea, elaborate on previous idea, rephrased idea, and who came up with the ideas that were eventually chosen by the team to be forwarded to the HR department.

We do not find statistically significant differences in any of the additional measures on the team level (see Tables B.1, B.3 and B.4).[1] When considering employees and leaders separately, we observe that the number of ideas contributed by the employees is slightly lower in the treatment than in the control group (column 2 of Table B.4). In addition, we find evidence that leaders in the treatment group are more engaged in the discussion as the number of words spoken by the leader is significantly higher in the treatment than in the control group (see column 6 in Table B.2). Interestingly these additional words spoken do not translate into significantly more ideas contributed, rephrased, or elaborated by the leader (see columns 3, 6, and 9 in Table B.4). As a next step, we identified who came up with the ideas that were sent to the HR department and how many people were involved in the development of each of these ideas (as classified by a research assistant blind to treatment). Table B.5 shows that we do not find significant treatment differences between the number of ideas that were developed as a group effort (column 1), the number of different people whose ideas were chosen to be sent to the HR department (column 2), or the number of ideas contributed by the leader (column 3).

---

[1]For simplicity of interpretation, we use OLS regressions in the tables below. The results are robust to using logistic regressions for binary variables and Poisson regressions for count data instead.

Table B.1: Contributions per person (all participants)

	(1)	(2)	(3)	(4)	(5)	(6)
	Nr contr	Nr contr	Nr Useful	Nr Useful	Nr Words	Nr Words
Treatment	8.31	4.71	0.11	-0.30	117.66*	67.83
	(5.32)	(4.85)	(0.50)	(0.42)	(68.62)	(61.15)
Constant	70.77***	131.22***	8.09***	16.66***	1,282.55***	2,453.50***
	(7.11)	(15.56)	(0.57)	(1.38)	(85.59)	(249.86)
Observations	414	414	414	414	414	414
Adj. R^2	0.09	0.22	0.03	0.12	0.01	0.08
Block FE	YES	YES	YES	YES	YES	YES
Controls	NO	YES	NO	YES	NO	YES

Notes: OLS regression with block-fixed effects. Numbers in parentheses indicate standard errors clustered by team. We include data of the 125 teams for which we have a transcript of the meeting and only include the 414 employees and leaders who also finished the endline survey, to keep the sample constant. The dependent variable is the number of contributions to the group discussion per person in columns (1) & (2), the number of useful contributions per person as classified by a research assistant in columns (3) & (4), and the number of words said during the group discussion per person in columns (5) & (6). We show all results with and without controls. Controls include team size, leader gender, leader seniority and average seniority of the team members, the leader's level, and a proxy for coaching behavior by the leader as reported by his or her employees. We include a dummy for missing observations for coaching behavior. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

Table B.2: Contributions per person by employees and leaders

	Nr contributions		Nr Useful		Nr Words	
	(1)	(2)	(3)	(4)	(5)	(6)
	Employees	Leader	Employees	Leader	Employees	Leader
Treatment	2.20	8.99	-0.90	1.13	-81.71	383.65***
	(4.783)	(6.451)	(0.574)	(0.769)	(69.98)	(125.68)
Constant	129.68***	110.47***	17.58***	14.99***	2,203.76***	2,359.04***
	(16.681)	(19.415)	(1.851)	(2.598)	(246.986)	(475.812)
Observations	289	125	289	125	289	125
Adj. R^2	0.240	0.244	0.087	0.218	0.076	0.155
Block FE	YES	YES	YES	YES	YES	YES
Controls	YES	YES	YES	YES	YES	YES

Notes: OLS regression with block-fixed effects. In columns (1), (3), and (5), we use individual-level data from 289 employees who completed the survey after the work task and for whose teams we have a transcript and cluster standard errors by team. In columns (2), (4), and (6), we use group level from data 125 teams for which we have a transcript of the meeting and report robust standard errors. The dependent variable is the number of contributions to the group discussion per person in columns (1) & (2), the number of useful contributions per person as classified by a research assistant in columns (3) & (4), and the number of words said during the group discussion per person in columns (5) & (6). Controls include team size, leader gender, leader seniority and average seniority of the team members, the leader's level, and a proxy for coaching behavior by the leader as reported by his or her employees. We include a dummy for missing observations for coaching behavior. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

Table B.3: Contributions by team

	(1) Nr Contr	(2) Nr Useful	(3) Nr Words	(4) HI Contr	(5) HI Words	(6) min Nr Words
Treatment	8.75	-1.55	125.90	0.00	0.01	106.47*
	(17.488)	(1.572)	(204.886)	(0.014)	(0.010)	(64.133)
Constant	181.757***	23.924***	3,589.853***	0.586***	0.628***	2,218.278***
	(50.899)	(4.449)	(752.508)	(0.057)	(0.035)	(247.492)
Observations	125	125	125	125	125	125
Adj. R^2	0.154	0.042	0.033	0.547	0.478	0.440
Block FE	YES	YES	YES	YES	YES	YES
Controls	YES	YES	YES	YES	YES	YES

Notes: OLS regression with block-fixed effects. We use group level from data 125 teams for which we have a transcript of the meeting and report robust standard errors. The dependent variable is the number of contributions to the group discussion in column (1), the number of useful contributions as classified by a research assistant in column (2), the number of words spoken in column (3), the Herfidahl index for contributions in column (4), the Herfindahl index for number of words spoken in column (5) and the minimal number of words spoken in column (6). Controls include team size, leader gender, leader seniority and average seniority of the team members, the leader's level, and a proxy for coaching behavior by the leader as reported by his or her employees. We include a dummy for missing observations for coaching behavior. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

Table B.4: Type of contribution

	Nr Ideas			Nr Rephrase			Nr Elaborate		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	All	Employees	Leader	All	Employees	Leader	All	Employees	Leader
Treatment	-0.288	-0.440*	0.141	-0.094	-0.234	0.152	-0.052	-0.093	0.099
	(0.183)	(0.242)	(0.400)	(0.235)	(0.213)	(0.594)	(0.204)	(0.276)	(0.312)
Constant	6.652***	7.050***	5.965***	2.854***	1.236*	3.990*	4.695***	5.409***	3.522***
	(0.647)	(0.924)	(1.232)	(0.859)	(0.740)	(2.212)	(0.670)	(0.891)	(1.046)
Observations	414	289	125	414	289	125	414	289	125
Adj. R^2	0.109	0.078	0.168	-0.006	-0.005	-0.047	0.087	0.087	0.114
Block FE	YES	YES	YES	YES	YES	YES	YES	YES	YES
Controls	YES	YES	YES	YES	YES	YES	YES	YES	YES

Notes: OLS regression with block-fixed effects with standard errors clustered by team except for columns (3), (6), and (9) where we use robust standard errors. In columns (1), (4), and (7), we use individual-level data from 414 employees and leaders for whose teams we have a transcript. We split the data into employee contributions (columns 2, 5, and 8) and leader contributions (columns 3, 6, and 9). The dependent variable is the number of ideas contributed in columns (1) - (3), the number of times an idea was rephrased in columns (4) -(6), and the number of times an idea was elaborated in columns (7) - (9). Controls include team size, leader gender, leader seniority and average seniority of the team members, the leader's level, and a proxy for coaching behavior by the leader as reported by his or her employees. We include a dummy for missing observations for coaching behavior. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

Table B.5: Development of ideas that have been reported to the HR

	(1) Nr of ideas elaborated together	(2) Nr of people's ideas'	(3) Nr of ideas leader
Treatment	-0.041	0.171	0.175
	(0.153)	(0.107)	(0.135)
Constant	0.875*	0.886***	1.250**
	(0.471)	(0.297)	(0.481)
Observations	125	125	125
Adj. R^2	-0.008	0.098	0.098
Block FE	YES	YES	YES
Controls	YES	YES	YES

Notes: OLS regression with block-fixed effects. We use group level from data 125 teams for which we have a transcript of the meeting and report robust standard errors. The dependent variable is the number of ideas that were reported to the HR which were developed as a group effort, i.e., several people elaborated on the idea (as rated by a research assistant blind to treatment) in column (1), how many different people came up with an idea that was sent to HR in column (2), and the number of ideas that were sent to HR that was contributed by the leader in column (3). Controls include team size, leader gender, leader seniority and average seniority of the team members, the leader's level, and a proxy for coaching behavior by the leader as reported by his or her employees. We include a dummy for missing observations for coaching behavior. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

As none of these measures could explain why employees report being more engaged in the meetings after the leader was treated, we have tried to learn more about what type of behavior the leader is showing and how this affects the atmosphere in the meetings. To this end, we have first asked a research assistant to indicate whether or not they perceived the leader as engaging. Overall, 80% of the leaders were classified as engaging. This share is with 87% significantly higher in the treatment group than in the control group, where only 74% of leaders are classified as engaging ($p = 0.055$, Chi^2-test). Second, we have used machine learning to classify the behavior of the leader in the meetings. To this end, we have used Chat GTP to classify the behavior of the leader in each of the meetings using the following prompt:"I have posted x parts of a meeting above. The leader of the meeting is person y. I would like you to answer two questions about the meeting. Based on the things the leader said, how would you describe the leadership style? How would you describe the leader's behavior based on the things the leader said?" We identify the ten most frequently used adjectives by Chat GTP to describe the behavior of the leader in both the treatment and control. Next, we create a dummy variable for each of these adjectives with value one if the description of the leader's behavior included this adjective. As shown in Table B.6, we find that the behavior of treated leaders is significantly more likely to be described as supportive (the odds ratios are approximately twice as high compared to the control group). In addition, we tried to explain the treatment status of the leader by the adjectives used. We find that being classified as supportive is the only adjective that is significantly correlated with

treatment status at the 5 percent level. We, therefore, conclude that the treated leader was behaving in a more supportive manner.

Table B.6: Leader behavior as classified by Chat GTP (Odds ratios)

	(1) Supportive Behavior	(2) Treatment Status
Treatment	2.283**	
	(0.891)	
Actively		1.615
		(0.713)
Positive		1.088
		(0.461)
Supportive		2.322**
		(0.990)
Inclusive		1.328
		(0.598)
Collaborative		0.523
		(0.224)
Organized		0.941
		(0.430)
Respectful		0.739
		(0.334)
Open		1.624
		(0.683)
Engaged		0.686
		(0.338)
Productive		1.015
		(0.537)
Constant	0.752	0.602
	(0.962)	(0.288)
Observations	125	125
Pseudo R^2	0.074	0.066
Block FE	YES	YES
Controls	YES	YES

Notes: Odds ratios based on logistic regressions with block-fixed effects. We use group level from data 125 teams for which we have a transcript of the meeting and report robust standard errors. The dependent variable is a dummy variable with value one if Chat GTP used the word " supportive" to describe the leader's behavior in column (1) and a dummy variable with value one if the leader was in the treatment group in column (2). Controls include team size, leader gender, leader seniority and average seniority of the team members, the leader's level, and a proxy for coaching behavior by the leader as reported by his or her employees. We include a dummy for missing observations for coaching behavior. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

As a final step, we have used the face reading software Microsoft Azure to classify

emotions the participants show in the meetings.[2] To that end, we have created video snippets of every 10 seconds of the first minutes of the team meetings, i.e., the time before the participants start the icebreaker task, as at this point most team leaders start sharing their screen, which does not allow the usage of the face reading software any longer. The software scans each of these snippets, identifies how many people are in the image, and classifies the emotions these people show into the following categories: anger, contempt, disgust, fear, happiness, neutral, sadness, and surprise. The sum of all emotion scores within one screen is normalized to one. As a next step, research assistants match the ID each person received based on their location on the screen to the ID each participant has in our data set. Figure B.1 below shows the average value for each of these emotions assigned to each person.

Figure B.1: Person average score for each emotion
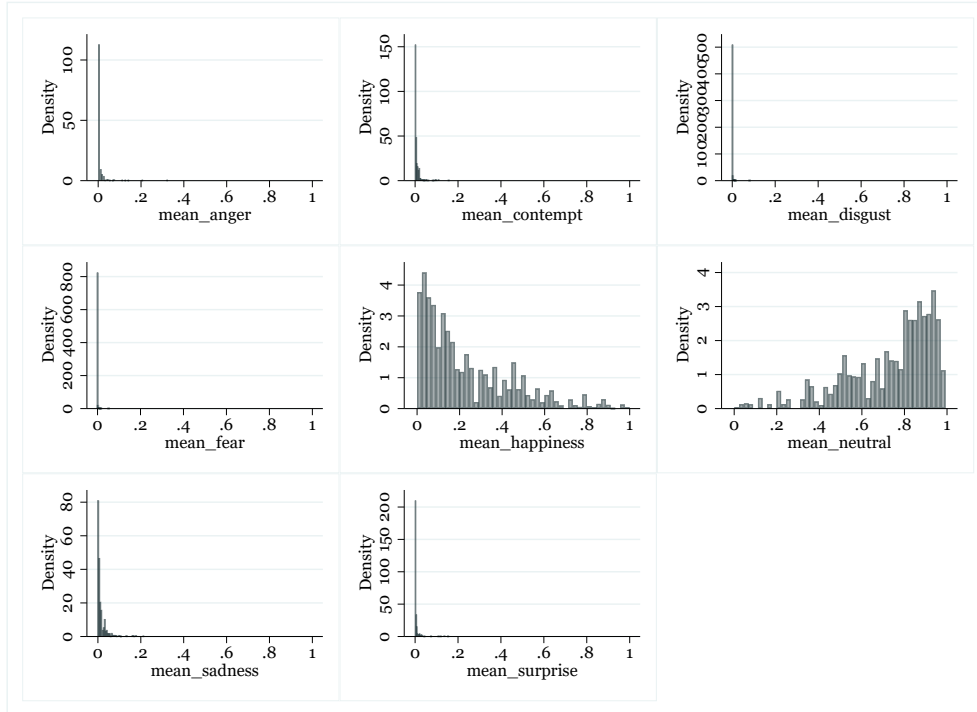
Notes: Histograms of per person average emotion scores for anger, contempt, disgust, fear, happiness, neutral, sadness, and surprise. Each emotion score ranges from zero to one and within each screen and for each person the scores are normalized to sum up to one.

Given that we find the most variation in the measure for happiness, indicating that the software performs better in detecting variation in happiness compared to the other

---

[2]Similar face-reading software has, for example, been used to investigate the relationship between emotions and philanthropy (Fiala and Noussair, 2017), risk-taking (Nguyen and Noussair, 2014), and trust (Kugler et al., 2020).

emotions, we focus on happiness in our analysis. Table B.7 shows that employees whose leaders have been treated receive a marginally significantly higher rating on happiness.[3] We interpret this as suggestive evidence that the treatment increased leaders' ability to generate a good meeting atmosphere.

Table B.7: Employee happiness as classified by the face reading software

	(1)	(2)
	Happiness	Happiness
Treatment	0.056*	0.063**
	(0.031)	(0.031)
Constant	0.192***	0.275**
	(0.049)	(0.108)
Observations	259	259
Adjusted R^2	-0.001	0.145
Block FE	YES	YES
Controls	NO	YES

Notes: OLS regression with block-fixed effects. We use individual-level data from 259 employees and 103 teams for which the software was able to identify emotions and report standard errors clustered by team. The dependent variable is happiness as rated by the face-reading software ranging from 0 to 1. Controls include team size, leader gender, leader seniority and average seniority of the team members, the leader's level, and a proxy for coaching behavior by the leader as reported by his or her employees. We include a dummy for missing observations for coaching behavior. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

C. Performance measures

In the main part of the paper, we use the average quality of the three ideas the teams submitted as the performance measure. In Table C.1, we instead consider each idea separately. We find treatment effects that are similar in size, however, the coefficients remain insignificant. The same picture emerges when we use more complex models and account for measurement errors such as rater bias.[4]

While we consider the compound measure in the main part of the experiment, Table C.2 shows the treatment effects on originality and relevance separately. We find suggestive evidence that the treatment helps the teams find more original ideas rather than more relevant ones.

---

[3]Note that the sample size is reduced because of the additional requirements for the videos in order for the software to be able to classify emotions. The analysis is based on the remaining 103 teams.

[4]Results are available from the authors upon request.

Table C.1: Treatment effect on performance using individual ideas

	Performance	
	(1)	(2)
Treatment	0.096	0.109
	(0.128)	(0.124)
Observations	388	388
Teams	130	130
Adjusted R^2	0.030	0.079
Block FE	YES	YES
Controls	NO	YES

Notes: OLS regression with block-fixed effects and standard errors clustered at the team level in parentheses. The dependent variable is the compound measure of performance with equal weights on relevance and originality for each idea. Each team submitted three ideas except for two teams which only submitted two ideas. The dependent variable is normalized relative to the distribution of the control group. Controls include leader gender, team size, seniority of the leader and average seniority of the team members, the leader's level and a proxy for coaching behavior by the leader as reported by his or her employees. We include a dummy for missing observations for coaching behavior to keep the sample constant. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

Table C.2: Treatment effect on idea originality and relevance

	(1)	(2)
	Originality	Relevance
Treatment	0.113*	0.049
	(0.062)	(0.054)
Observations	130	130
Adjusted R^2	0.091	0.183
Block FE	YES	YES
Controls	YES	YES

Notes: OLS regression with block-fixed effects and robust standard errors in parentheses. In column (1), we use the average originality score as the dependent variable, and in column (2) we use the average relevance score as the dependent variable. Dependent variables are normalized relative to the distribution of the control group. Controls include leader gender, team size, seniority of the leader and average seniority of the team members, the leader's level, and a proxy for coaching behavior by the leader as reported by his or her employees. We include a dummy for missing observations for coaching behavior to keep the sample constant. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

D. Deviations from the Pre-Analysis Plan

In this appendix, we will discuss all the changes we made with respect to the pre-analysis plan. We first discuss variables that were preregistered but are not included in the main part of the paper. To be more precise, we elicited a second measure of performance capturing creativity which we discuss in Section D.1 and we elicited a more extensive set of control variables which we include as additional controls in Section D.2. Lastly, we describe minor changes due to errors in the pre-registration.

D.1. Creativity

We also elicited a second measure of performance, capturing the groups' creativity. At the beginning of the meeting, participants were asked to solve an icebreaker task. The task is based on the "Unusual Uses Task" developed by Guilford (1967) and performance on the task has been used as a proxy for creativity. In this task, the groups have three minutes to think about as many "unusual uses" of an ordinary item such as, for example, paper. Creativity is a compound measure of originality, i.e., the infrequency of an answer, flexibility, i.e., the number of distinct categories the answers came from, and validity, i.e., whether the answer actually describes a potential use of the item (see Table D.1). We follow the procedure of Bradler et al. (2019) when aggregating the three categories. Participants repeated the task three times, using three different items. We also used the same items as Bradler et al. (2019), namely paper, tin can, and cord. We used the average score over all three items as our measure of creativity. The order of the items was fixed across treatments. We decided not to include the measure in the main part of the experiment for the following reason. The task was presented to the participants as an icebreaker activity and was not incentivized. The reason for this unconventional design choice was to keep the setting as similar to a normal work meeting as possible. Unfortunately, the video material shows that the teams did not take the task very seriously and many had problems executing it. Consequently, we do not consider the measure reliable but, for the sake of completeness, provide the results below. Table D.1 shows the distribution of the measure and the subcategories. We observe the score for 122 out of 130 teams as eight teams did not submit any response. Table D.2 shows the treatment effect on creativity. We do not find a significant effect and the coefficients do not have the expected sign.

Table D.1: Creativity

	Mean	SD	N
Score Creativity task	23.55	6.44	122
Validity	33.14	9.18	122
Flexibility	26.19	6.54	122
Originality	11.32	4.81	122

Notes: Descriptive statistics for the creativity score and its components.

Table D.2: Treatment effect on creativity

| | Creativity score | |
	(1)	(2)
Treatment	-0.208	-0.148
	(0.158)	(0.154)
Observations	122	122
Adjusted R^2	-0.011	0.140
Block FE	YES	YES
Controls	NO	YES

Notes: OLS regression with block-fixed effects and robust standard errors in parentheses. The dependent variable is the creativity score and is normalized relative to the distribution of the control group. Controls include leader gender, team size, seniority of the leader and average seniority of the team members, the leader's level, and a proxy for coaching behavior by the leader as reported by his or her employees. We include a dummy for missing observations for coaching behavior. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

D.2. Additional Control Variables

In addition to the control variables included in the main part of the paper, we also elicited the following variables: the leaders' and employees' mindset and transformational leadership style, as reported by the leader's employees.

These variables were collected during the baseline survey. Employees' and leaders' mindset is based on an 8-item implicit person theory measure developed by Levy and Dweck (1997) using a 6-point Likert scale. For the transformational leadership style, employees for whom the leader has personnel responsibility evaluated the leader on a 5-point Likert scale using the 20 items measuring transformational leadership in the Multifactor Leadership Questionnaire (Bass and Avolio, 1997).

To show that our sample is balanced also considering these additional variables, Table D.3 presents results from the ordinary least squares regressions of the treatment dummy on the additional control variables including block fixed effects. We do not find significant differences between the two groups.

Table D.4 and Table D.5 replicate Table 2 and Table 3, respectively, including the additional controls. The results do not change, and the control variables have limited explanatory power, which is why we did not include them in the main part of the paper.

Table D.3: Balance table (additional controls)

Variable	(1) Control	(2) Treatment	(3) Difference	(4) Obs.
Leader Mindset	3.993	3.933	-0.066	130
	(0.776)	(0.863)	(0.145)	
Average Mindset (team)	3.722	3.743	0.022	130
	(0.311)	(0.273)	(0.052)	
Transformational leadership	3.897	3.742	-0.153	83
	(0.351)	(0.629)	(0.112)	
Observations	67	63	130	

Notes: The columns provide the mean (and standard deviation) for the work teams in the control group (column 1), and the treatment group (column 2), and the estimated coefficient (robust standard error) from regressing each covariate against treatment status controlling for block fixed effects (column 3). * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

D.3. Necessary corrections

There are a few errors in the pre-analysis plan that we list here:

- Satisfaction was measured on a 5-point Likert scale

- The coaching behavior scale developed by Heslin et al. (2006) consists of 10 items.

- Mindset at baseline was measured based on Levy and Dweck (1997).

Table D.4: Manipulation Checks (extended)

	Supportive behavior			Leader Mindset		
	(1)	(2)	(3)	(4)	(5)	(6)
Treatment	0.372**	0.317*	0.393**	0.523***	0.589***	0.627***
	(0.173)	(0.176)	(0.173)	(0.186)	(0.144)	(0.139)
Female Leader		0.076	0.050		0.237	0.249
		(0.185)	(0.192)		(0.164)	(0.168)
Team size		-0.241*	-0.241*		0.036	0.022
		(0.145)	(0.142)		(0.141)	(0.146)
Seniority (leader)		-0.025	-0.017		0.037	0.030
		(0.091)	(0.094)		(0.070)	(0.073)
Leader Mindset		-0.020	-0.038		0.668***	0.649***
		(0.088)	(0.087)		(0.084)	(0.084)
Seniority (team)		-0.060	-0.084		-0.006	-0.015
		(0.079)	(0.081)		(0.068)	(0.076)
Creat. Mind.		0.131	0.128		-0.128	-0.121
		(0.103)	(0.101)		(0.095)	(0.095)
High level leader			-0.031			-0.083
			(0.214)			(0.156)
Coaching			-0.026			-0.176*
			(0.119)			(0.093)
Transform. lead			0.063			0.103
			(0.129)			(0.093)
Leadership miss.			-0.459**			-0.306*
			(0.178)			(0.164)
Constant	-0.022	0.853	0.991*	0.086	-0.324	-0.171
	(0.239)	(0.570)	(0.564)	(0.309)	(0.546)	(0.569)
Observations	129	129	129	127	127	127
R^2	0.012	0.018	0.039	0.049	0.454	0.468
Block FE	YES	YES	YES	YES	YES	YES

Notes: OLS regression with block-fixed effects and robust standard errors in parentheses. The dependent variable is the proxy for the extent to which the leader engaged in supportive leadership practices as reported by the work team in columns (1) - (3) and the leader's self-reported mindset in columns (4) - (6). Dependent variables are normalized relative to the distribution of the control group and independent variables are normalized using the entire sample. A sample size of less than 130 is due to missing observations of the dependent variable. Group characteristics include leader gender, team size, seniority of the leader and average seniority of the team members, the leader's mindset at baseline, and the average group mindset at baseline. Leadership style includes a proxy for coaching behavior by the leader and transformational leadership style as reported by his or her employees. We include a dummy for missing observations for coaching behavior and leadership style. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

Table D.5: Treatment effects on employee satisfaction, engagement and group performance - extended

	Satisfaction			Engagement			Performance		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Treatment	0.296**	0.249**	0.269**	0.187*	0.160	0.163	0.121	0.109	0.131
	(0.119)	(0.118)	(0.125)	(0.109)	(0.113)	(0.113)	(0.157)	(0.157)	(0.156)
Female Leader		0.147	0.155		-0.054	-0.057		-0.001	-0.064
		(0.130)	(0.134)		(0.131)	(0.132)		(0.186)	(0.178)
Team size		-0.054	-0.066		-0.111	-0.114		-0.026	0.006
		(0.116)	(0.121)		(0.106)	(0.108)		(0.175)	(0.158)
Seniority (l.)		-0.044	-0.053		0.061	0.054		-0.088	-0.064
		(0.060)	(0.060)		(0.055)	(0.052)		(0.072)	(0.071)
Leader Mindset		-0.038	-0.044		-0.036	-0.037		-0.051	-0.042
		(0.049)	(0.049)		(0.052)	(0.053)		(0.081)	(0.079)
Seniority (t.)		-0.037	-0.047		-0.053	-0.077		0.040	-0.023
		(0.053)	(0.052)		(0.048)	(0.049)		(0.070)	(0.074)
Creative Mind. (t.)		0.098*	0.095*		0.062	0.056		0.020	0.003
		(0.057)	(0.057)		(0.054)	(0.055)		(0.080)	(0.078)
High level l.			0.016			0.227*			0.625***
			(0.132)			(0.130)			(0.206)
Coaching			-0.084			-0.059			0.162
			(0.063)			(0.081)			(0.135)
Trans. lead.			0.074			0.052			0.043
			(0.075)			(0.089)			(0.129)
Lead. miss.			-0.130			-0.121			-0.182
			(0.131)			(0.113)			(0.181)
Constant	-0.001	0.119	0.207	0.107	0.527	0.562	-0.418*	-0.292	-0.432
	(0.177)	(0.462)	(0.474)	(0.107)	(0.425)	(0.432)	(0.250)	(0.680)	(0.606)
Observations	302	302	302	302	302	302	130	130	130
Adjusted R^2	0.018	0.028	0.022	0.001	-0.003	-0.007	0.029	-0.005	0.064
Block FE	YES	YES	YES	YES	YES	YES	YES	YES	YES

Notes: OLS regression with block-fixed effects and standard errors in parentheses. In columns (1) - (6), we use individual-level data from 302 employees who completed the survey after the work task and cluster standard errors at the team level, in columns (7) - (9), we use group-level data of 130 groups and report robust standard errors. The dependent variable is the employees' self-reported satisfaction with the group tasks in columns (1) - (3) and employees' self-reported engagement in columns (4) - (6). In columns (7) - (9), the dependent variable is a compound measure of performance with equal weights on relevance and originality. Dependent variables are normalized relative to the distribution of the control group and independent variables are normalized using the entire sample. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

E. Summary statistics

Table E.1: Satisfaction

	Mean	SD	N
Satisfaction discussion	4.42	0.66	302

Notes: Descriptive statistics for the survey measure for satisfaction with the group task.

Table E.2: Engagement

	Mean	SD	N	Cronbach's Alpha
Engagement	4.30	0.57	302	0.73
Engagement1	4.13	0.80	302	
Engagement2	4.28	0.73	302	
Engagement3	4.49	0.60	302	

Notes: Descriptive statistics for the survey measure for employee engagement in the group task and its components Engagement1-3.

Table E.3: Performance

	Mean	SD	N
Performance	2.68	0.57	130
Originality	2.56	0.64	130
Relevance	2.81	0.59	130

Notes: Descriptive statistics for the measure for performance based on the rating by the HR department and its components Originality and Relevance.

Table E.4: Leader behavior

	Mean	SD	N	Cronbach's Alpha
Leader behavior	4.38	0.51	302	0.79
To what extent did the leader...				
encourage you to think outside the box?	3.88	0.86	302	
make you feel safe and secure?	4.37	0.73	302	
listen to the suggestions you were making?	4.70	0.54	302	
provide ideas him or herself?	4.21	0.82	302	
make you feel comfortable to participate in the discussion?	4.74	0.50	302	

Notes: Descriptive statistics for the survey measure for leader behavior in the group task elicited from the employees and its components.

Table E.5: Mindset

	Mean	SD	N	Cronbach's Alpha
Leader Mindset ex post	4.14	1.07	127	0.92
Mindset 1	4.15	1.21	127	
Mindset 2	3.94	1.17	127	
Mindset 3	4.33	1.07	127	

Notes: Descriptive statistics for the survey measure for leader mindset elicited after the group task and its components.

Table E.6: Observational Measures

	Mean	SD	N
Engagement (video)	58.29	34.80	126
Supportive behavior (video)	4.43	1.90	130

Notes: Descriptive statistics for observational measures.

Table E.7: Coaching behavior

	Mean	SD	N	Cronbach's Alpha
Coaching behavior	3.84	0.83	324	0.93
Coaching 1	3.60	1.00	321	
Coaching 2	3.54	1.06	319	
Coaching 3	3.61	1.01	322	
Coaching 4	3.45	1.06	318	
Coaching 5	3.93	1.10	320	
Coaching 6	3.96	1.05	321	
Coaching 7	3.78	1.10	316	
Coaching 8	4.23	1.00	319	
Coaching 9	3.98	1.07	318	
Coaching 10	4.33	0.94	319	

Notes: Descriptive statistics for the measure for coaching behavior and its components elicited from the employees in the baseline survey.

Table E.8: MLQ

	Mean	SD	N	Cronbach's Alpha
Transformational Leadership	3.82	0.45	329	0.94
Tramsformative Leadership 1	3.42	0.89	291	
Transformative Leadership 2	3.12	0.99	311	
Transformative Leadership 3	3.85	0.82	319	
Tramsformative Leadership 4	3.99	0.79	320	
Transformative Leadership 5	3.64	0.99	294	
Transformative Leadership 6	3.85	0.87	320	
Transformative Leadership 7	3.72	0.92	314	
Transformative Leadership 8	3.27	1.01	315	
Transformative Leadership 9	4.03	0.86	305	
Transformative Leadership 10	4.42	0.75	321	
Tramsformative Leadership 11	4.30	0.77	324	
Tramsformative Leadership 12	4.04	0.82	306	
Tramsformative Leadership 13	4.11	0.84	322	
Tramsformative Leadership 14	3.77	0.97	313	
Tramsformative Leadership 15	3.92	0.92	314	
Tramsformative Leadership 16	3.64	0.83	312	
Transformative Leadership 17	3.63	0.94	314	
Transformative Leadership 18	3.56	0.84	311	
Transformative Leadership 19	3.92	0.85	318	
Tramsformative Leadership 20	4.08	0.78	316	

Notes: Descriptive statistics for the transformative leadership measure elicited from the employees in the baseline survey and its components.

F. Instructions (translated from Norwegian)

Welcome!

Thank you for contributing to *companyname*'s future.

Please provide your email address, so we know which team you belong to (only the researchers will have access to this information to merge the data).

Enter your email address:

[   ]

Confirm your email address

[   ]

We will now explain to you how the group discussion you will have in Microsoft Teams will be structured.

It is important that you read these instructions carefully.

IMPORTANT!
We will guide you through the group discussion step by step on this page during the meeting. It is therefore important that you do not close this browser window after reading the instructions, but continue to click through the instructions during the meeting. You ae not done before you have finished the short survey at the end.

The steps of the meeting are:

- At the beginning of the group discussion: **Activate recording**
- When the meeting has started:
    - Introduce yourself to the other participants and also let them introduce themselves. (Please also make sure that they have activated their cameras.)
- Briefly introduce the agenda of the meeting
- Ice breaker task
- Group discussion:
    - Phase 1: Brainstorm
    - Phase 2: Prioritize and make recommendations
- Post discussion survey for group leaders and members

27

We will now provide more details for each step.

At the beginning of the group discussion: Activate recording. If you are working for Viken or Signal, please ask one of the other participants to start the recording for you.

Click on the three dots and then select „Start Inspilling"
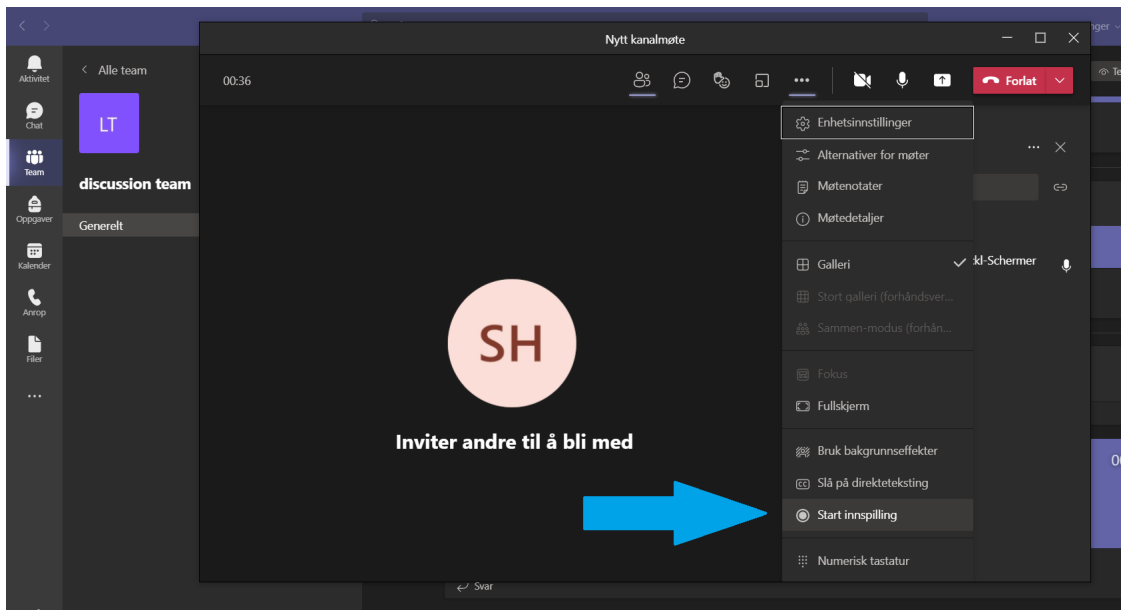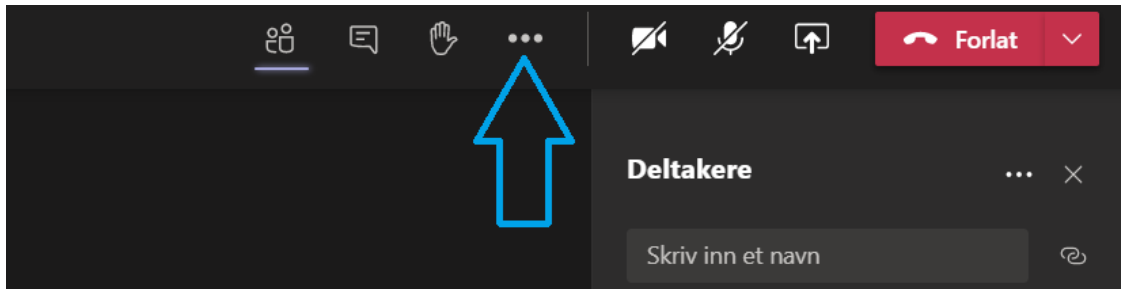
If you have been successful, you will see a red dot in the upper left corner of your meeting screen.
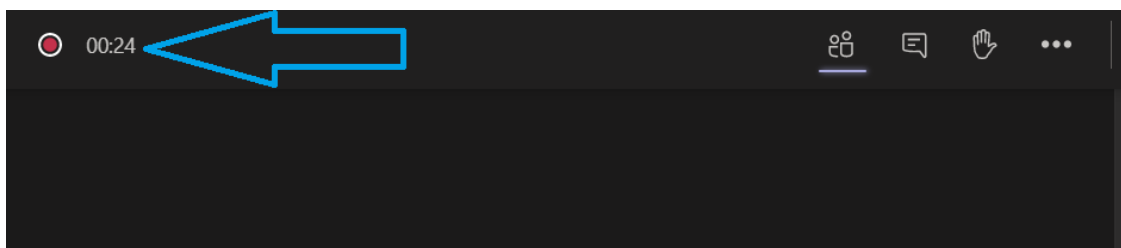
When the meeting has started, please introduce yourself to the two other participants and also let them introduce themselves. Also, provide a short overview of the meeting agenda.

After your introduction we have a short ice breaker task prepared for you and your group. To work on this task, you have to open this browser window again.

Here is a short preview for you:

Task Description:
The ice breaker task consists of three rounds. In each round, you will be asked to think about unusual uses of an object. Please enter as many of these "uses" as your group can come up with in the survey form that you will see on this page. You have three minutes for each of the objects.

Example of the Task:
Please list as many, as different, and as unusual uses for a rubber tire as you can think of. Do not restrict yourself to a specific size of a tire. You can also list uses that require several tires. Do not restrict yourself to uses you are familiar with, but think of as many new uses as possible! After three minutes the next task will appear.

Example Answers:
Please describe the possible use in a few words if necessary. Using the example of the rubber tire: "sled" or "flower box" are clear answers, whereas "target" would require further explanation, such as "ball game with tire as target. Try to think of original answers: An answer is considered (very) original if only (very) few people think of it. Furthermore, try to think about different categories: Using the example of the rubber tire: "car tire" and "bicycle tire" belong to the category "tires as wheels" and the answer "swing seat" is a different category (category "toys").

After completing the ice breaker task, please start the discussion with your group members.

During the group discussion, we would like you to discuss *companyname*'s future with a focus on a specific question, we
we tell you more about when the meeting has started. As a meeting leader you are also expected to participate in the discussion.

You have 30 minutes to discuss. First, please take 15 minutes to collect ideas.
Please collect all your ideas on one PowerPoint slide (just a simple list is enough). After you have prepared a list of potential ideas, please take the next 15 minutes to choose the 3 ideas you find most promising. Elaborate on each of these ideas. Make a new slide for

each idea and describe it in 3 bullet points.

To open a PowerPoint slide in Microsoft Teams click on "Filer".
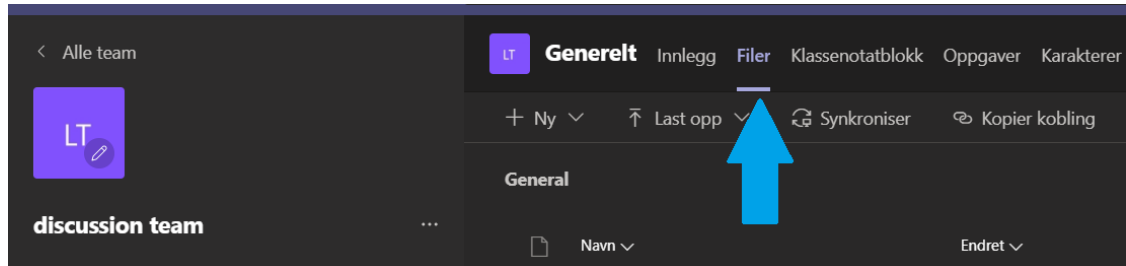
Next click on "Ny" and select "Power Point-presentasjon"
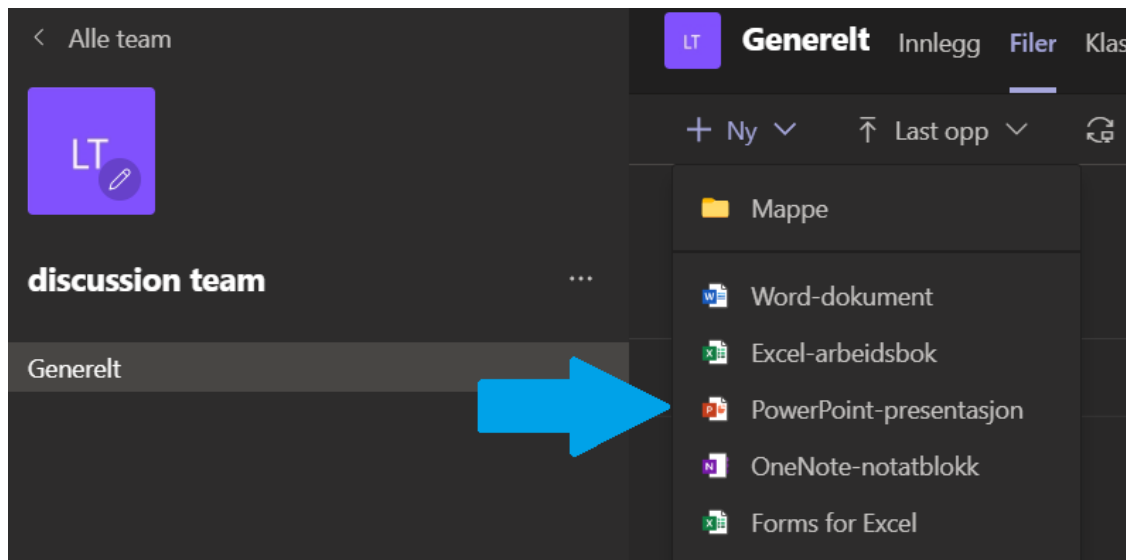
Enter "Ideas Team X" with X being your Team number as the file name. Your changes to the Power Point Presentation will be saved automatically.

Before you continue, please try if you are able to open a Power Point Presentation and save it in Microsoft Teams.

Thank you for completing the introduction. The meeting will start soon
Here we have summarized again the outline of the group discussion.
(Download the summary so you can access it during the meeting.)

- Start with a short round of introductions. Also, briefly explain the agenda of the meeting.
- Open your browser and do the "ice-breaking" tasks with your group.
- Explain the topic of the group discussion to your group. We will display the topic of the discussion in this browser when you reach this part.
- Open a PowerPoint Presentation in Teams.
- First 15 minutes: Brainstorm and list all ideas that your group has (on the first PowerPoint slide)
- Second 15 minutes: Pick 3 ideas and describe each of them on one PowerPoint slide. Be brief - 3 bullet points are enough.

Please note that all the information will be repeated here once your meeting has started, so you can see the instructions again step by step. (Click "Meeting has started" when the meeting has started.)

Start the meeting at the set time. We hope that everyone will show up as agreed upon. However, please start the meeting nonetheless, if someone does not show up, as long as at least one person shows up.

Good luck!

Great, your meeting has started.

Now please make sure you **activated the recording.** If you are working for Viken or Signal, please ask one of your group members to start the recording for you.
Please also check that **everyone has turned on their cameras**.

If you have forgotten how to start the recording you can look at the instructions again <u>here</u>.

☐   recording is activated

☐   all cameras are turned on

Introduce yourself to the two other participants and also let them introduce themselves.
Afterwards, briefly explain the meeting agenda.

Breaking the ice

You can now start the ice breaker task. Please read the following out loud to your group members:

"The ice breaker task consists of three rounds. In each round, you will be asked to think about unusual uses of an object. Please enter as many of these "uses" as your group can come up with in the survey form that you will see on this page. You have three minutes for each of the objects.

Example of the Task:
Please list as many, as different, and as unusual uses for a rubber tire as you can think of. Do not restrict yourself to a specific size of a tire. You can also list uses that require several tires. Do not restrict yourself to uses you are familiar with, but think of as many new uses as possible! After three minutes the next task will appear."

If you and your group members are ready please press "**Ready**" to see the first object.

Round one:

In this round your object is **Paper**.

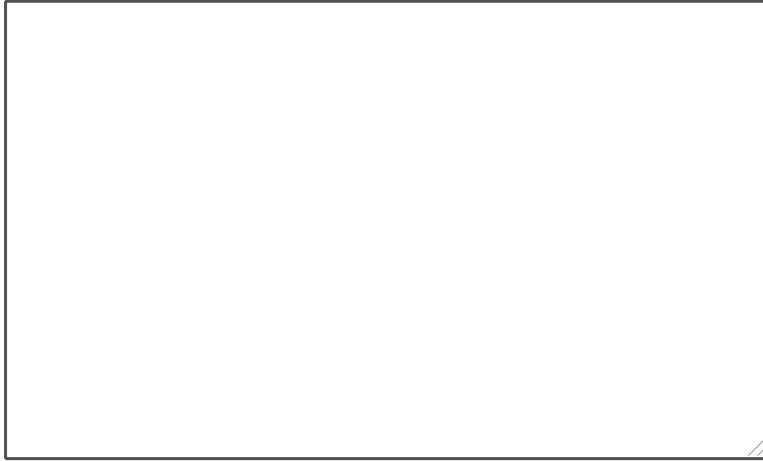Please list the "uses" your group could think of in the box below

```

```

0 3 0 0

Round two:

In this round your object is **Tin Can**.

Please list the "uses" your group could think of in the box below

0 3 0 0

Round three:

In this round your object is **Cord**.

Please list the "uses" your group could think of in the box below

0 3 0 0

Great!

You have now completed the ice breaker task.

Now you can start the discussion about *companyname*'s future.

Please read the following instructions out loud:

"*companyname* has an ambitious goal in its Strategy 2030 that says that we want to be among the top ten employers. We would like you to discuss what *companyname* can do to become an even more attractive workplace. The meeting leader is also expected to participate in the discussion.

You have 30 minutes to discuss. Please take 15 minutes to brainstorm and collect ideas. Write down all ideas in one PowerPoint slide (a simple list is enough). After you have prepared a list of potential ideas, take the next 15 minutes to choose the 3 ideas you find most promising. Elaborate on each of these ideas. Make a new PowerPoint slide for each idea and describe it in three bullet points.."

If you are done, please press "**Start discussion**"

Discussion

What can *companyname* do to become an even more attractive workplace?

Please use the first 15 minutes to brainstorm ideas. Collect your ideas on the first slide of your PowerPoint Presentation. The researchers will not forward this first slide to *companyname*'s strategy discussion.

1 5 0 0

Please use the next 15 minutes to choose the 3 best ideas and generate a PowerPoint slide for each idea. Describe each idea in three bullet points. These slides will be used in *companyname*s strategy discussion.

IMPORTANT: Remember to come back to this website after you are done with the discussion and the PowerPoint presentation, so you can receive the link to the final survey for leaders and employees.

1 5 0 0

You have now finished the group discussion. We kindly ask each of your to take 5 minutes to respond to the survey.
Please share the following link with your group members by posting it in the chat in Teams (NOTICE: You as a leader, should not follow the same link).

 link

The link leads to a short survey to provide feedback on the group discussion. You can see the survey for leaders if you press continue.

G. Supportive Leadership Treatment Instructions

In this part, we will discuss some insights from research about how to lead a good meeting.

We will also ask you to share some of you own experiences. We are working on the development of a leadership training and need good practical examples to further develop the program.

The importance of mindset

Recent psychological and management research has reached some interesting conclusions.

Leading developmental psychologist, Carol Dweck from Stanford University has, together with
colleagues, found the following:

- people who believe that abilities can grow and develop have a tendency to set higher goals, to persist for longer when they encounter difficulties, and to experience more success than those who believe that talent and ability are given once and for all.

Slightly simplified, Dweck says that there are two different ways of thinking:

- We can believe abilities and skills can develop through hard work, good strategies, and help from others. We call this way of thinking a growth mindset.
- We can believe that the malleability of abilities is limited. We call this way of thinking a fixed mindset.

The importance of a <u>leader's</u> mindset

Inspired by Dweck's mindset research, management researchers have investigated how a leader's mindset affects his or her management practices. This research suggests that a leader with a fixed mindset will often provide less mentoring and support to his or her employees compared to a leader with a growth mindset.

In this way, a leader's mindset can become a self-fulfilling prophecy. If a leader believes that an employee is not able to contribute to a meeting, this leader might be less willing to support and encourage the employee, which can in itself make it more difficult for the employee to contribute.

We are currently developing a leadership training and need good practice examples to improve the training.

Could you please describe a situation in which one of your employees has shown development and achieved something which many people wouldn't have expected of him or her?

```
[                                                                    ]
[                                                                    ]
[                                                                    ]
[                                                                    ]
```

The importance of a leader's support and encouragement

In the type of group discussion you are about to lead, it is important to encourage everyone to contribute. Everyone has different experiences. Therefore, it is important that each participant shares his or her opinion. Some can be reluctant to come with suggestions because they do not feel that they have anything to contribute or they are afraid that their suggestion is somehow stupid.

Research on mindset has shown that a leader who believes that everyone can contribute, will work hard to ensure that everyone is included in the discussion. Such a leader will support and encourage the participants to contribute.

How can we encourage participants to contribute?

As mentioned earlier, we are working on developing a leadership training and for that we need good and realistic examples.

Please think about your own experience and describe a meeting in which you felt that it was easy to share your ideas.

```
[                                                                    ]
[                                                                    ]
[                                                                    ]
[                                                                    ]
```

What was special about this meeting? What did the other participants do to make you feel safe to contribute?

[ ]

Here are some examples of what other people have said:

Mari M.: I feel more comfortable in group discussions when I know that other people struggle as well. That makes it less scary to say something that others might find stupid.

Vegard H.: I need to feel understood. It is important to me that I notice that people try to understand what I am trying to say.

Here is what research suggest are good leadership practices:
(Download the summary of good meeting leadership, so you have it available during the meeting)

Create a safe environment

- Emphasize that each participant's ideas are important because they all have their own experiences and it is important that the group comes up with a variety of ideas.
- Acknowledge that some might feel nervous in this type of group discussion, but clarify that there is no need to be shy and that it is important that all of them contribute. Emphasize that there are no stupid suggestions.

Be an active listener

- Look into the camera when others speak and signal them that you are listening (e.g., by nodding and confirming that you understand what they are saying).
- Write down their suggestions.
- Ask clarification questions.

When you are collecting ideas in the upcoming group discussion:

Show group members that it is okay to think outside the box

- Emphasize that there is also room for more radical ideas
- Consider sharing one very radical idea yourself to show the group members that any type of contribution is welcome

When you are choosing and elaborating on ideas in the upcoming group discussion:

Encourage constructive criticism and prioritize

- Allow group members to be critical but honor the potential of each idea.
- Emphasize that it is crucial that they speak their mind in order to select and elaborate on the three best ideas.

What would be your advice?

Imagine that a colleague asks you what he or she can do to become a better meeting leader. This colleague has observed that many participants do not dare to speak up in the project meetings he or she leads. What would you tell him or her?

You are now done with the introduction about good leadership.

We hope that our tips from the research frontier will help you.

Meeting has started

Thank you for completing the introduction. The meeting will start soon
Here we have summarized again the outline of the group discussion.
(Download the summary so you can access it during the meeting.)

- Start with a short round of introductions. Also, briefly explain the agenda of the meeting.
- Open your browser and do the "ice-breaking" tasks with your group.
- Explain the topic of the group discussion to your group. We will display the topic of the discussion in this browser when you reach this part.
- Open a PowerPoint Presentation in Teams.