Journal of Economic Behavior and Organization 188 (2021) 18–35

Contents lists available at ScienceDirect

Journal of Economic Behavior and Organization

journal homepage: www.elsevier.com/locate/jebo

Forced distribution rating systems and team collaboration

Linda Loberg[a], Stephan Nüesch[a], Johann Nils Foege[b,*]

[a] University of Münster, Business Management Group, Schlossplatz 3, 48143 Münster, Germany
[b] Leibniz University Hannover, Innovation Management Group, Schloßwender Str. 7, 30159 Hannover, Germany

ARTICLE INFO

Article history:
Received 1 April 2020
Revised 21 April 2021
Accepted 21 April 2021
Available online 1 June 2021

Keywords:
Performance appraisal
Forced distribution
Collaboration
Knowledge sharing
Team performance

ABSTRACT

This study provides three real-effort experiments on how a forced distribution rating system (FDRS) influences team collaboration. In the first and the second experiment, we examine the performance implications of an FDRS in a card sequencing task (1) when working alone and (2) when working in a team. In the third experiment, we test how an FDRS affects knowledge sharing within teams. Our findings show that an FDRS increases the speed of completing the card sequencing task when working alone and decreases the speed of completing the card sequencing task when working in a team. Beyond that, we find that an FDRS also significantly decreases knowledge sharing within teams. As the FDRS was perceived as unfair in collaborative settings but not when working alone, we provide evidence on the role of perceived justice concerning the effects of an FDRS and shed light on the psychological and economic consequences of introducing an FDRS in environments where team collaboration is essential for success.

1. Introduction

Teams are at the center of how work is accomplished in modern business (Kozlowski, 2017). In team settings, individual contributions are likely to combine in a non-additive manner due to the manifold interdependencies in team collaboration (Alchian and Demsetz, 1972). As individual team members' contributions are typically not quantifiable, they are often evaluated based on subjective ratings by the supervisor (Angelovski et al., 2016). Subjective performance evaluations, however, suffer from biases such as the leniency bias, i.e. the tendency to give inflated evaluations, or the centrality bias, i.e. the tendency to give undifferentiated evaluations (Prendergast, 1999; Berger et al., 2013; Machegiani et al., 2016). To counteract these two biases, firms can introduce a forced distribution rating system (FDRS). An FDRS is a performance appraisal system that forces supervisors to rate their employees according to a given distribution. In the famous case of General Electric, a 20/70/10 distribution was implemented, meaning that 20% of employees had to be rated as top performers, 70% as so-called "vitals," or average performers, and 10% as poor performers (Mulligan and Bull Schaefer, 2011).

This study examines the influence of an FDRS on team collaboration in three real-effort experiments. Proponents of FDRSs (Grote, 2005; Welch and Byrne, 2003) argue that by increasing the accuracy of performance evaluations and by setting strong incentives for good performers, FDRSs stimulate employee motivation and thus increase performance. While internal

* Corresponding author.
E-mail addresses: linda.loberg@uni-muenster.de (L. Loberg), stephan.nueesch@wiwi.uni-muenster.de (S. Nüesch), nils.foege@wa.uni-hannover.de, nils.foege@rub.de (J.N. Foege).

https://doi.org/10.1016/j.jebo.2021.04.035
0167-2681/

comparisons and pay disparities are manifestations of a pay-for-performance culture that encourages enhanced employee efforts (Becker and Huselid, 1992; Ehrenberg and Bognanno, 1990a,b; Lazear and Rosen, 1981), they may also discourage employees' cooperation and collaboration, decrease team cohesion, and increase individualism and selfishness (Chambers and Baker, 2020; Lazear, 1989; Levine, 1991; Pfeffer and Langton, 1993; Shaw et al., 2002; Siegel and Hambrick, 2005).

Empirical evidence on the effectiveness of FDRSs is scarce. Several studies test how *endogenous* ratings or wage disparities affect performance; in these studies, the rating or wage disparity can be chosen by the supervisor and is not predetermined by the experimenters (e.g., Pfeffer and Langton, 1993; Shaw et al., 2002; Siegel and Hambrick, 2005; Franck and Nüesch, 2011). Several other studies compare how different *predetermined exogenous* ratings or wage disparities affect performance; in these studies, the rating or wage disparity cannot be chosen by the supervisor and is predetermined by the experimenters (e.g., Beersmaa et al., 2003; Harbring and Irlenbusch, 2008, 2011; Danilov et al., 2019). However, even if the disparity were exactly the same, the motivation and incentive effects are likely to be different between endogenous and predetermined exogenous disparity (Fehr et al., 2013). To the best of our knowledge, only one study, Berger et al. (2013), compares the effects of a forced distribution rating system with disparity that is freely chosen and thus endogenous. Berger et al. (2013) conducted a laboratory experiment on an FDRS's effect on subjects' performance in which subjects had to solve a task individually. Their results indicate higher productivity if the supervisor has to use an FDRS, but that an FDRS leads to sabotage and thus lower performance if such an option is granted to the participants. Whether an FDRS is used or not, team members do much more than sabotage each other, and firms are increasingly relying on teamwork instead of individual effort (Kozlowski and Bell, 2013), as, under the best of circumstances, interdependencies in teams facilitate knowledge transfer, foster creativity, and create synergies (Kozlowski, 2017; van Knippenberg and Mell, 2016). For this reason, our research examines the effects of an FDRS on different kinds of team collaboration.

To test the influence of an FDRS, we conducted three different real-effort experiments. In the first experiment, the participants performed a card sequencing task alone. In the second experiment, they performed the same card sequencing task in a team. We find that an FDRS increases the speed of accomplishing the task when working alone and decreases the speed when working in a team. In the third experiment, we investigated the effect of an FDRS on intra-team knowledge sharing when conducting a prediction task. Each team member received different background information and could communicate with the other team members through a chatroom that was invisible to the supervisor. We find that an FDRS significantly decreases the number of words and the number of task-relevant statements written in the chatroom, and, as a consequence, significantly diminishes prediction accuracy. The number of informal statements, however, was unaffected by the FDRS. The three experiments show that an FDRS has detrimental consequences on team collaboration.

2. Literature review and behavioral predictions

FDRSs are relative performance appraisal systems that force supervisors to compare employees to each other and rate them (Stewart et al., 2010). Typically, each employee must be assigned to one of three categories: top performers, vitals, and poor performers (Mulligan and Bull Schaefer, 2011). The prescribed distribution of employees across ratings often approximates a bell curve, as in the case of GE, with its 20/70/10 system, or Ford, with its 10/80/10 system (Moon et al., 2016). This classification has important consequences for employees: While top performers get high bonuses, poor performers receive no bonuses, and can even lose their jobs if their ratings remain low (Berger et al., 2013; Scullen et al., 2005).

2.1. Forced distribution rating systems when working alone

An FDRS counteracts rater biases such as the leniency bias and the centrality bias (Berger et al., 2013; Schleicher et al., 2008). The leniency bias is supervisors' tendency to rate their employees too mildly to avoid giving bad feedback or causing conflict in the team by rating a team member as a poor performer (Blume et al., 2009). The centrality bias is supervisors' tendency to give compressed ratings around a norm rather than to differentiate between good and bad performances (Prendergast, 1999). As FDRSs are meant to force supervisors to be more honest and direct, FDRSs are supposed to improve rating accuracy and thus employee motivation by rewarding truly good performances (Grote, 2005; Guralnik et al., 2004; Murphy and Cleveland, 1995). This creates a high-performance culture and promotes honest feedback (Grote, 2005; Guralnik et al., 2004; Welch and Byrne, 2003).

Deutsch (1985) argues that the equity rule, according to which everyone is rewarded proportionally to their contribution, is the dominant principle of distributive justice in situations with a strong focus on individual productivity. The equity rule may be contrasted with the equality rule, under which all team members receive the same reward. Justice judgment theory's (Leventhal, 1976) concept of the equity rule is derived from Adams (1963) equity theory, which proposes that employees perceive they are being treated fairly if they receive rewards that match their contribution relative to their co-workers' contributions. By counteracting the supervisor's centrality bias, an FDRS creates rating differentials that may help to implement the equity rule and thus increases perceived fairness. Thus, the employees know that if they perform better than their peers, they will also receive a higher bonus than their peers, which increases perceived fairness compared to a situation in which all receive the same or a very similar bonus (Strom et al., 2014). However, it is important to look carefully at the details of an FDRS. When an FDRS regulates not only the percentages of the team members that have to be rated as top, medium, and low performers, but also the specific bonuses attached to the ratings (as in our experiments), it is uncertain whether the predetermined bonus disparity will increase perceived equity by aligning the reward-contribution ratios within the team. Thus, when employees work alone, according to the justice judgment theory (Deutsch, 1985; Greenberg, 1987; Leventhal, 1976), the effect of an FDRS with specific bonuses on perceived justice, and thus effort, is unclear.

According to Lazear and Rosen's (1981) tournament theory, an FDRS should have a positive effect on productivity when individuals work alone. By counteracting the centrality bias, an FDRS introduces so-called tournaments in the form of internal comparisons and, thus, provides workers with incentives to be more productive than their peers (Berger et al., 2013; Harbring and Irlenbusch, 2003). Evidence suggests that such tournaments also increase performance when task interdependencies are low, as in our working alone condition (e.g., Becker and Huselid, 1992; Ehrenberg and Bognanno, 1990a,b).

Hypothesis 1. *An FDRS increases performance when working alone compared to a non-forced rating system.*

2.2. Forced distribution rating systems when working in a team

One response of organizations to the growing complexity, competition, and uncertainty in today's business environments is the increased prevalence of work teams (Kozlowski and Ilgen, 2006; Salas et al., 2016; van Knippenberg and Mell, 2016). Teams are often able to provide more rapid, complex, and innovative solutions to organizational problems than individuals are (Kozlowski and Ilgen, 2006; Salas et al., 2016). The work that teams do is typically characterized by a high degree of task interdependence (Kozlowski and Bell, 2013), i.e. team members work in close collaboration. Task interdependence is defined as the team members' need to share materials, information, and expertise to complete their tasks (van der Vegt and van de Vliert, 2002). Team members have to coordinate their joint activities (Katz-Navon and Erez, 2016), leading to frequent spillover effects from an individual's accomplishment into the work processes of other team members. Thus, the team's output cannot be broken down to individual team members' contributions (Alchian and Demsetz, 1972).

Through their tournament-like structures and the imposed salary dispersion, FDRSs can induce competition and undermine cooperation and collaboration among team members (Garcia and Tor, 2007; Moon et al., 2016) and may even lead to sabotage behaviors, such as impeding other employees' work, as has been found in tournament studies (Harbring and Irlenbusch, 2011). Lazear (1989) and Levine (1991) predict that members of teams with higher salary dispersion will act more selfishly and behave less cooperatively than members of teams with compressed wage structures. Berger et al. (2013) found that an FDRS decreased performance when study participants had the opportunity to sabotage their team members. We argue that the negative consequences of an FDRS on team performance go beyond sabotage, and that team members generally collaborate less efficiently under an FDRS.

Levine (1991) argues that pay disparity decreases cohesiveness, and refers to Deutsch (1988), who summarized the evidence by noting that "when efficiency requires efficient cooperation, almost any movement towards a democratic egalitarian structure increases effectiveness" (Levine, 1991, p. 239). Leventhal (1976) and Deutsch (1985) argue that in contexts that require collaboration and good social relations, the equality rule – that all team members receive the same reward – is the dominant principle of distributive justice. As an FDRS makes it impossible for each team member to receive the same reward, it creates feelings of injustice in collaborative settings, which can result in lower motivation and reduced willingness to cooperate.

Hypothesis 2. *An FDRS decreases performance when working in a team compared to a non-forced rating system.*

2.3. Forced distribution rating systems and knowledge sharing

Bartol and Srivastava (2002) define knowledge sharing as "individuals sharing organizationally relevant information, ideas, suggestions, and expertise with one another." While knowledge sharing can happen on many levels, individuals are the organizations' prime knowledge-movers (Nonaka, 1994). As effective knowledge management is crucial to gain a sustainable competitive advantage (Grant, 1996), organizations need to facilitate knowledge exchange between employees. Knowledge sharing positively relates to important team and organizational outcomes. Mesmer-Magnus and Dechurch (2009), for instance, provide meta-analytical evidence for a stable positive relationship between information sharing and team performance. They further show that information sharing also positively relates to team cohesion, decision effectiveness, and satisfaction.

We argue that the decision to share or not to share knowledge is a result of a process of balancing the pros and cons of that behavior (Nelson, 2016). On the one hand, employees who share their knowledge may gain a good reputation and expert status (Nelson, 2016), or just a warm glow, a feeling of having done the right thing (Leszczyc and Rothkopf, 2010). They might also rely on the strong norm of reciprocity, expecting to get something back for giving (Falk and Fischbacher, 2006). Humans' need to belong, an extremely pervasive motivation to form and maintain interpersonal relationships (Baumeister and Leary, 1995), might also play a role. Being in contact with one's co-workers, helping them, and exchanging information and thoughts creates mutual trust and fosters the bonds between individuals (Bierly et al., 2009).

On the other hand, employees may not have enough time to share their knowledge, or may not perceive the usefulness of doing so (Cabrera and Cabrera, 2002; KPMG 2000; Riege, 2005). Beyond that, revealing personal insights might make employees vulnerable, as they could potentially lose their advantage in knowledge and thereby their capability to outperform their co-workers (Cabrera and Cabrera, 2002).

We expect that employees' motivation to collaborate and to share their knowledge also depends on the work environment, and, in particular, on the reward system. We argue that an FDRS, a competitive reward system, intensifies the disadvantages of knowledge sharing. Knowledge is typically transferred between co-workers in informal situations, for example, during a coffee break and/or in personal one-on-one communication. This activity is typically unobserved by the supervisor. While supervisors can judge the outcome of employees' behavior and team processes, it is almost impossible for supervisors to correctly assess the individual inputs, such as the effort each team member invested to share knowledge. As a consequence, under a competitive reward system, employees may change their contributory behavior and may choose to reduce their effort in collaborative actions and, instead, increase their effort in other activities that are more visible to the supervisor.

Justice judgment theory predicts that the equality rule is the dominant principle of distributive justice in such team contexts where collaboration is beneficial: that is, that every team member should be rewarded equally (Deutsch, 1985; Leventhal, 1976). An FDRS, however, impedes the application of the equality rule, and leads to ratings that will be perceived as unfair. As a consequence, employees may change their contributory behavior and may choose to share less knowledge, investing their effort in other activities that are more visible to the supervisor instead.

Previous research supports this reasoning, providing evidence that a competitive reward system decreases knowledge sharing, while a cooperative reward system enhances collaboration. Fey and Furu (2008) examine the effect of endogenous wage and rating disparity on the intensity of knowledge sharing within firms. They find that incentive pay based on corporate results leads to more knowledge sharing between multinational units. The experiment by Ferrin and Dirks (2003) shows that an exogenously introduced cooperative reward system increases knowledge sharing and trust, whereas an exogenously introduced competitive reward system decreases trust and knowledge sharing. In the cooperative reward condition, the participants were told that their reward would be based on the performance of their dyad. In the competitive reward condition, the participants were told that their reward would be based on their performance relative to their partner's performance. Thus, in the cooperative reward condition, both individuals of the dyad received the same reward, whereas, in the competitive reward condition, rewards could differ between the partners of a dyad. Our experiment compares knowledge sharing in the non-forced rating system treatment, in which the supervisor was able to freely decide how to distribute the bonus pool within the team and was free to use either cooperative or competitive remuneration, with knowledge sharing in the FDRS treatment, in which the supervisor was forced to distribute the bonus pool among the team members according to a predetermined competitive distribution. Hence, we hypothesize:

Hypothesis 3. *An FDRS decreases intra-team knowledge sharing compared to a non-forced rating system.*

3. Design of the three experiments

We conducted three experiments to test the influence of an FDRS on team collaboration. Each including two treatments, one in which the distribution of the bonus pool by a supervisor followed an FDRS and one in which the supervisor was free on how to distribute the bonus pool – a non-forced rating system. In all three experiments, subjects were randomly assigned a fixed role, either employee or supervisor. Three employees and one supervisor formed a team and engaged in the same type of task for 10 rounds. The subjects were recruited from a large German university using the ORSEE online recruiting system (Greiner, 2015). The first and second experiments took place in May 2017, the third in June 2018. We chose a between-subject design where the subjects were randomly assigned to one of the two treatments. Each subject participated in only one of the three experiments.

The experimenters followed a standardized procedure written down in a manual. Instructions were distributed and read aloud to make sure that everyone in the team had the same information. Thereafter, participants had to answer four control questions to ensure that everyone understood the procedure and incentive scheme before the experiment could start.[1] The first and second experiments included face-to-face interactions and did not involve a computer or other electronic device. The third experiment was computerized using the software z-Tree (Fischbacher, 2007) and participants remained anonymous to each other throughout the experiment.

3.1. Description of the tasks

In the first and the second experiment, the participants conducted a card sequencing task based on the work of Bachrach et al. (2006). The card sequencing task required putting cards in sequences as quickly as possible, following certain rules. During the task, the employees were allowed to talk to each other, whereas the supervisor who stayed in the same room had to remain silent. Thus, the supervisor could easily observe the speed at which the card sequencing task was conducted.

In the first experiment, which we call "working alone," each employee of a group of three simultaneously received one pile of cards and worked on it independently. It was not explicitly forbidden to help other team members after completing all sequences of one's own pile. However, no such behavior was observed by the experimenter.

In the second experiment, which we call "working in a team," the team of three together received three piles, one pile after another. They needed to finish the first before receiving the second and the second before receiving the third. The experimenter did not give any instructions on how the team should organize itself to conduct the task. It was not explicitly

[1] The experiments were conducted in German. The instructions in Appendix I and Appendix II are a translation into English of the original instructions.

forbidden for each team member to work on a single pile, one after another, while the other two waited. However, no such behavior was observed by the experimenter. Thus, the employees were forced to work together to complete the task as quickly as possible.

In the third experiment, the employees' task was to make predictions in 10 rounds of 10 min each. Each team of employees chatted through connected computer terminals. All test subjects were in the same room but did not know who else was assigned to their team. In the first two experiments, the supervisor could observe the behavior of the three employees, while in this third experiment, he/she could not. The employees answered questions without having access to the Internet. Each employee received three pieces of background information. Two of the three pieces of background information were given to all team members whereas the third piece of information differed between team members. By combining all pieces of information, the employees could correctly solve the task by means of calculating the correct answer. For example, in one task, the employees had to answer the question: What was the population of Brazil in 2016? All three employees received the following two pieces of information. (1) Germany had a population of 83 million. (2) The population of the USA was 240 million higher than that of Germany. Each of the following pieces of information was given to only one of the three employees: (1) The population of Brazil was 112 million higher than that of Egypt. (2) The population of Egypt was 49 million higher than that of Spain. (3) The population of Germany was 36 million higher than that of Spain. The correct answer, 208 million, could be calculated by combining the pieces of information. The participants knew that the supervisor could not read or send any messages. After each employee had given a prediction (an answer) to the presented question, the employees' predictions and the correct answer were displayed on the supervisor's screen.

3.2. Description of incentives and FDRS manipulation

In the first and the second experiment, the supervisor had to distribute a bonus pool of 45 euros among the three employees in each round. In the *FDRS* treatments, the supervisor was forced to assign 30 euros to one employee, 15 euros to another, and 0 euros to the third. In the *non-forced rating system* treatments, the supervisor was able to freely decide how to distribute the bonus pool of 45 euros among the three team members. After each round, each employee learned how many euros he/she had been assigned by the supervisor in that particular round but did not learn how much the other employees had been assigned.

In the first and second experiment, the supervisor's payoff was dependent on the speed of the employees in solving the card sequencing task and the number of incorrect sequences. In the first experiment, the supervisorś payoff was based on the average of the individual times that the three employees needed to solve the task and the average of the number of incorrect sequences. In the second experiment, the supervisor's payoff was based on the time the team needed to solve the task and on the number of incorrect sequences of the team.[2]

To avoid wealth effects, one round was randomly chosen to determine the payoff at the end of the experiment. The shorter the time, the more money the supervisor received. To calculate the supervisor's payoff, the exact (average) time was rounded down to the next whole minute. For every minute that this time was less than 20 min, the supervisor received one euro. In both of these experiments, at the end of the session, before paying out the supervisor and the employees, the experimenter analyzed the accuracy of the preserved card sequences to measure the number of incorrect sequences. The supervisor and the employees lost 50 cents per card sequence that did not meet the criteria defined in the task. Whereas the speed in completing the task was well observed by the supervisor, the presence of incorrect sequences was less visible and measured only ex-post by the experimenter. The fact that employees knew that incorrect sequences negatively influenced their final payoff reduced or eliminated their incentives to increase the speed at the expense of quality. Indeed, robustness analyses show that an FDRS does not influence the frequency of incorrect sequences when working alone or when working in a team. All subjects received a show-up fee of 12 euros and each member of the team with the quickest (average) time in one round with correct sequences won a 10-euro Amazon voucher.[3]

In the third experiment, the employees' predictions and the correct answer were displayed on the supervisor's screen after each employee had given a prediction to the presented question. The supervisor then had to assess the employees' performance by assigning euros to the team members. In the third experiment, the supervisor had to distribute a bonus pool of 60 euros among the three employees in each round. In the *FDRS* treatment, the supervisor was forced to assign 40 euros to one employee, 20 euros to another, and 0 to the third. In the *non-forced rating system* treatment, the supervisor was able to freely decide how to distribute the bonus pool of 60 euros among the three team members. After the supervisor had assigned the money, the prediction of every employee of the team and the correct value were displayed to the employees. They also learned the individual number of euros they had been assigned by the supervisor.

The supervisor's payoff in the third experiment depended on the accuracy of the employees' predictions. The better the predictions, the more money the supervisor received. To calculate the supervisor's payoff, the average of the relative deviations of employees' predictions in absolute terms was calculated. For every percentage point that this average was less than 25%, the supervisor received 1 euro. In Experiment 3, the bonus pool was 60 euros, not 45 euros as in Experiments 1

[2] The supervisor's payoff was contingent on the employees' performance, as we aimed at replicating what happens in real life, where supervisors' rewards typically depend on their employees' performance.

[3] As the Amazon voucher was granted in both the non-forced distribution rating system and the FDRS treatments in the first two experiments, it served as an additional incentive to perform well in all four treatments and is therefore not expected to confound our treatment effects.

Table 1
Description of the three experiments.

	Experiment 1	Experiment 2	Experiment 3
Task	Card sequencing task when working alone Card sequencing task based on the work of Bachrach et al. (2006). Each of the three employees simultaneously received one pile of cards and worked on it independently	Card sequencing task when working in a team Card sequencing task based on the work of Bachrach et al. (2006). The three employees received three piles together, one after another, and needed to finish the first before receiving the second, and needed to finish the second before receiving the third	Prediction task Each of the three employees answered questions like "What was the population of Brazil in 2016?" without having access to the Internet. Each employee received three pieces of background information. Two of the three pieces of background information were given to all team members whereas the third piece of information differed between team members. Employees were able to communicate with each other through a chatroom, whose content was invisible to the supervisor
FDRS treatment	The supervisor was forced to assign 30 euros to one employee, 15 euros to another, and 0 euros to the third	The supervisor was forced to assign 30 euros to one employee, 15 euros to another, and 0 euros to the third	The supervisor was forced to assign 40 euros to one employee, 20 euros to another, and 0 euros to the third
Non-forced rating system treatment	The supervisor was able to freely decide how to distribute the bonus pool of 45 euros among the three employees	The supervisor was able to freely decide how to distribute the bonus pool of 45 euros among the three employees	The supervisor was able to freely decide how to distribute the bonus pool of 60 euros among the three employees
Availability of an individual performance measure	Yes	No	Yes
Performance measure	Time the employee needed to complete all sequences of his/her pile	Time the team needed to complete all sequences of the three piles given to the team	Prediction error in%
Supervisorś incentives	The mean of the three times was rounded down to the next whole minute. For every minute that this time was less than 20 min, the supervisor received one euro. The supervisor lost 50 cents per card sequence that did not meet the criteria defined in the task	The time the team needed was rounded down to the next whole minute. For every minute that this time was less than 20 min, the supervisor received one euro. The supervisor lost 50 cents per card sequence that did not meet the criteria defined in the task	For every percentage point that the employees' average prediction deviated from the correct value less than 25%, the supervisor received 1 euro
Teams	20	20	20
Employees	60	60	60
Supervisors	20	20	20
Rounds	10	10	10
Average age	23.7 years	23.9 years	22.8 years
Percentage male participants	45%	55%	50%
Average duration of the experiment	1h 55 min	2 h 5 min	2 h 30 min
Average payoff of the supervisors	24.98 euros (incl. 12 euros show-up fee)	24.52 euros (incl. 12 euros show-up fee)	30.07 euros (incl. 15 euros show-up fee)
Average payoff of the employees	26.19 euros (incl. 12 euros show-up fee)	26.56 euros (incl. 12 euros show-up fee)	35.00 euros (incl. 15 euros show-up fee)

and 2, and all subjects received a show-up fee of 15 euros, not 12 euros as in Experiments 1 and 2, to provide a consistent hourly rate. All details of the experiments, such as the tasks and the incentive schemes, were common knowledge.

Table 1 gives a summary of all three experiments, including tasks, FDRS treatments, performance measure, supervisorś incentives, number of teams, employees, supervisors, and rounds, average age, percentage of male participants, average duration of the experiment and average payoffs of the supervisors and the employees.

4. Results

4.1. Distribution of the bonus pool

In the FDRS treatments, ratings were predefined to be 0, 15, and 30 euros, each with the likelihood of one third, in Experiments 1 and 2, and 0, 20, and 40 euros, each with the likelihood of one third, in Experiment 3. Fig. 1 shows the euros assigned to the employees by the supervisor in the non-forced rating system treatments in which the supervisors could decide how to distribute the bonus pool. Fig. 1 shows that supervisors' distributions were centralized, especially in

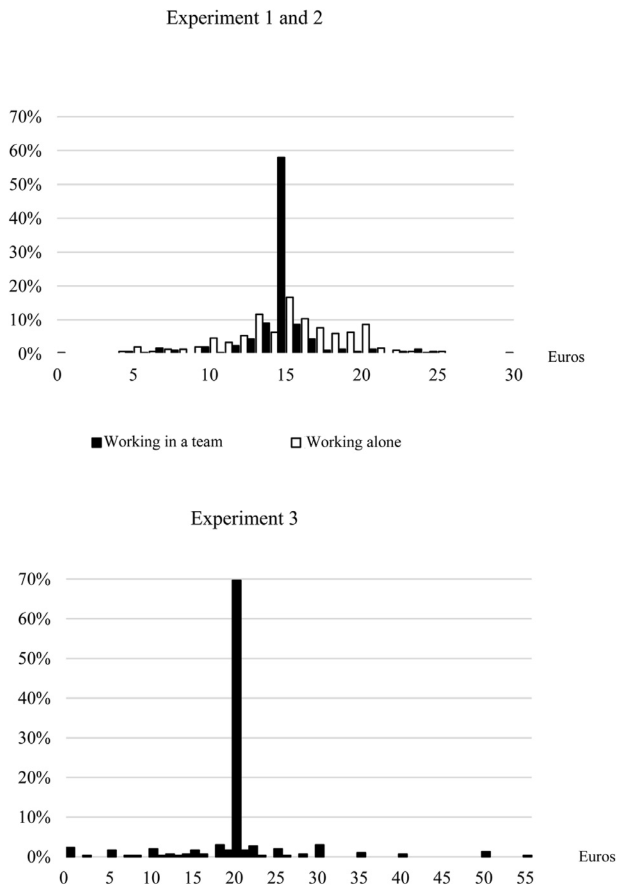

Fig. 1. Bonuses in euros assigned to the employees by the supervisor in the non-forced rating system treatments.

Experiment 2, in which individuals did the card sequencing task in a team, and in Experiment 3, in which the individuals could share relevant information through a chatroom that was invisible to the supervisor. In Experiment 2, 58% of the values assigned were exactly 15 euros, which is one-third of the 45 euros the supervisor could distribute among the three employees. In Experiment 3, 70% of the values assigned were exactly 20 euros, which is one-third of the 60 euros the supervisor could distribute among the three employees. In Experiment 1, where individuals did the card sequencing task alone, about 90% of the distribution was within 5 euros of the mean of 15 euros. This is in line with previous findings on the centrality bias (Prendergast, 1999).

The coefficient of variation of the employees' bonuses is significantly higher in the FDRS treatments than in the non-forced distribution treatments in all three experiments (see Table 2). The coefficient of variation of the employees' bonuses was lowest in the non-forced rating system in Experiment 2 (card sequencing task when working in a team), in which only team performance measures were available, followed by Experiment 3 (prediction task), in which employees could share relevant information through a chatroom that was invisible to the supervisor, and finally Experiment 1 (card sequencing task when working alone), in which the supervisor had access to individual performance measures, i.e. the time an employee needed to complete the card sequencing task.

4.2. Outcome variables and treatment effects of an FDRS

As subjects were randomly assigned to treatments and roles within the experiments, treatment assignment is uncorrelated with potential confounders such as team composition variables. Thus, simple mean comparison tests are sufficient to identify treatment effects. Table 2 shows the means of the team performance, the coefficient of variation of the employees' bonuses, and the perceived distributive justice in both the non-forced rating system treatment and in the forced distribution rating system (FDRS) treatment. In Experiments 1 and 2, performance is measured by *the time needed to complete the card sequencing task in seconds*. In Experiment 3, we used the percentage deviation of the subjects' predictions from the true value, i.e. *the prediction error in percent* as a proxy for information sharing. The more information the employees exchanged and combined, the more accurate the predictions should be.

Table 2
Treatment effects of an FDRS.

	Non-forced rating system	FDRS	Statistical significance of mean differences (all observations)			Statistical significance of mean differences (one observation per team)		
	Means		n	t-statistics	p-value	n	t-statistics	p-value
Experiment 1 (Card sequencing task when working alone)								
Time needed to complete the task in seconds	313.9	273.9	600	**−3.82**	0.000	20	**−3.15**	**0.006**
Coefficient of variation of the employees' bonuses	0.27	1.00	200	**34.87**	**0.000**	20	**14.77**	**0.000**
Perceived distributive justice	3.62	3.28	58	−1.26	0.211	20	−1.59	0.130
Experiment 2 (Card sequencing task when working in a team)								
Time needed to complete the task in seconds	395.2	471.6	200	**3.38**	**0.001**	20	**2.21**	**0.040**
Coefficient of variation of the employees' bonuses	0.12	1.00	200	**45.81**	**0.000**	20	**16.18**	**0.000**
Perceived distributive justice	4.04	3.22	59	**−3.71**	**0.000**	20	**−3.91**	**0.001**
Experiment 3 (Prediction task)								
Prediction error in%	8.1	20.6	593	**6.69**	**0.000**	20	**2.90**	**0.010**
Number of words	62.9	40.5	600	**−7.84**	**0.000**	20	**−3.06**	**0.007**
Number of task-relevant statements	8.2	5.2	600	**−8.08**	**0.000**	20	**−3.24**	**0.005**
Number of informal statements	2.5	1.8	600	**−2.29**	**0.022**	20	−0.79	0.438
Coefficient of variation of the employees' bonuses	0.19	1.00	200	**22.24**	**0.000**	20	**8.95**	**0.000**
Perceived distributive justice	4.04	3.22	60	**3.71**	**0.000**	20	**3.06**	**0.007**

Notes: We had only 58 observations rather than 60 observations of perceived distributive justice in Experiment 1 and 59 observations rather than 60 observations of perceived distributive justice in Experiment 2 due to missing observations in the questionnaire. In Experiment 3, we had in total 7 missing observations of the prediction error in%, as in seven cases, employees did not enter a prediction within the given maximal time frame of 10 min.

Due to outliers identified in a boxplot, we winsorized the upper 5% of the *time needed to complete the card sequencing task in seconds* and the *prediction error in %* (although our results do not change in a significant way if these variables were not winsorized). In Experiment 3, we also used measures that we retrieved from the employees' chatroom conversations to better understand the mechanisms and processes potentially driving the prediction error. We measured the number of words written in the chatroom as a general measure of communication intensity. To receive more specific measures of information transfer, two independent raters analyzed the content of the chatroom conversations. They coded the number of task-relevant statements (Cohen's kappa = 0.85) and the number of informal (task-irrelevant) statements (Cohen's kappa = 0.85). If the classifications differed between raters, they discussed the cases to reach a joint conclusion.

Table 2 also shows the number of observations n, t-statistics, and p-values of mean comparison tests when using all observations available (at the individual-round level in Experiments 1 and 3, the team-round level in Experiment 2, and, for perceived justice, the individual-level in Experiments 1, 2, and 3) and when using team averages and only one observation per team. The analyses using all observations available may exaggerate statistical significance as they are based on observations that are not entirely independent. Individual behavior may be influenced by behavior and responses in previous rounds and by peer effects within the team, even in the working alone conditions, as individuals performed the task in the same room at the same time (concerning peer effects, see also Falk and Ichino, 2006). A more conservative test can be obtained by calculating the team averages over all ten rounds and using each team as a unit of observation (see also Hamman et al., 2010). As each team is randomly assigned to only one treatment, and as there is no interaction between teams (in Experiments 1 and 2 a session included only one team), each team-level observation is perfectly independent.

As a manipulation test, Table 2 shows the means of the coefficient of variation of the employees' bonuses in the non-forced rating system. To verify our argumentation based on justice judgment theory (Leventhal, 1976), Table 2 also shows the means of perceived distributive justice in the non-forced rating system and the FDRS. We assessed employees' perceptions of distributive justice through the questionnaire at the end of each of the three experiments and before the employees learned which round was chosen for payment. We used the 4-item scale constructed by Colquitt (2001). The employees specified their attitude on a 5-point Likert scale ranging from 1 "strongly disagree" to 5 "strongly agree." Illustrative items are "To what extent does your rating reflect the effort you have put into your work?" and "To what extent is your rating appropriate for the work you have completed?" Internal consistency was $\alpha = 0.89$ in Experiment 1, $\alpha = 0.92$ in Experiment 2, and $\alpha = 0.92$ in Experiment 3.

In Experiment 1, where individuals did the card sequencing task alone, they needed an average of 313.9 s in the non-forced rating system treatment and only 273.9 s in the FDRS treatment to complete the card sequences.[4] Mean comparison tests show that the difference is highly statistically significant, both when using all observations and when using team averages and only one observation per team. Thus, an FDRS significantly increases the speed when working alone, which supports Hypothesis 1.

Result 1. *An FDRS significantly increases performance when working alone compared to a non-forced rating system.*

In Experiment 2, the team needed an average of 395.2 s in the non-forced rating system treatment and 471.6 s in the FDRS treatment to complete the card sequences. Mean comparison tests show that the difference is statistically significant at the 1% level when using all observations available at the team-round level and at the 5% level when using team averages and only one observation per team as a conservative test. Thus, an FDRS significantly decreases the speed when working in a team, which supports Hypothesis 2.

Result 2. *An FDRS significantly decreases performance when working in a team compared to a non-forced rating system.*

The average prediction error in Experiment 3 was 8.1% in the non-forced rating system and 20.6% in the FDRS condition. Mean comparison tests are statistically significant at the 1% significance level, both when using all observations and when using team averages and only one observation per team. Thus, the FDRS significantly reduced prediction accuracy.

We also find that individuals wrote an average of 62.9 words in the chatroom in the non-forced rating system and 40.5 words in the FDRS condition. The mean difference is statistically significant at the 1% level, both when using all observations and when using team averages and only one observation per team. The same applies to the number of task-relevant statements, which is significantly lower in the FDRS treatment than in the non-forced rating system treatment. When analyzing the number of informal statements, we find that they are lower in number in the FDRS treatment but that the mean difference is not statistically significant when using team averages and only one observation per team as a conservative test. Thus, under an FDRS, employees specifically decide not to share task-relevant knowledge, while informal statements are not affected. We thus find support for Hypothesis 3, that an FDRS decreases employees' knowledge sharing behavior, especially the communication intensity and the self-initiated sharing of task-relevant knowledge.

Result 3. *An FDRS significantly decreases within-team knowledge sharing.*

[4] It is important to note that the increased speed in the FDRS treatment is not achieved at the expense of lower quality in terms of a higher number of incorrect sequences. An FDRS does not influence the frequency of incorrect sequences either when working alone or when working in a team.

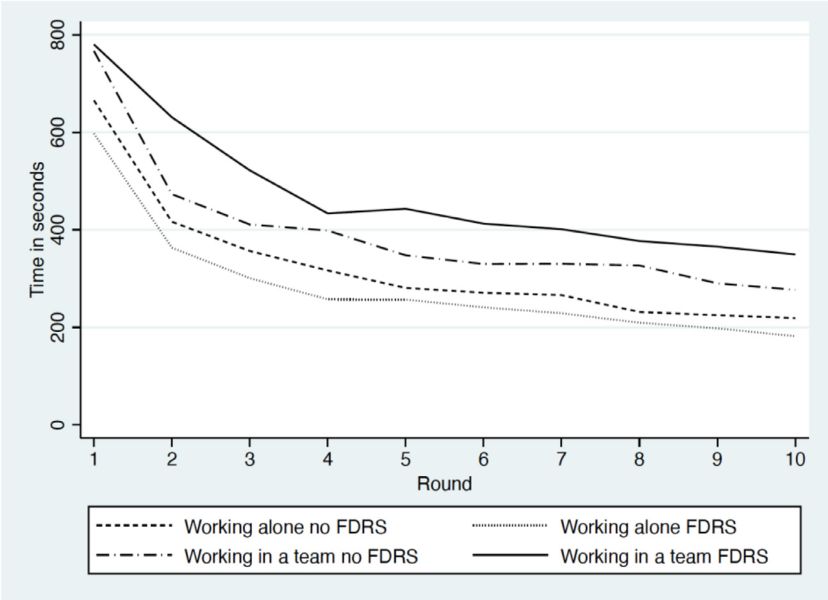

Fig. 2. Development of the time to complete the card sequencing task in Experiment 1 and 2.

When analyzing the treatment effects on perceived individual justice, we find that an FDRS significantly decreases perceived individual justice in Experiment 2 and 3, when working in a team and when sharing knowledge within a team, but not in Experiment 1, in which individuals performed the card sequencing task alone.

Result 4. *An FDRS significantly reduces perceived distributive justice when working in a team and when sharing knowledge within a team, but not when working alone.*

4.3. Evolution of teamwork over time

In the previous sections, we have pooled all 10 rounds and analyzed mean effects. This section looks at temporal developments. Fig. 2 illustrates the temporal development of the number of seconds needed to complete the card sequencing task, in both Experiment 1 and in Experiment 2 and for both the non-forced rating system treatment and FDRS treatment. As one would expect, individuals and teams become much faster over time as they become accustomed to the task. Interestingly, however, the learning curve is the same whether one works in a team or alone, and it is unaffected by the presence or absence of an FDRS. In all ten rounds, the order remains the same: The speed is highest when working alone in an FDRS, second-highest when working alone in a non-forced rating system, third-highest when working in a team in a non-forced rating system, and lowest when working in a team in an FDRS.

Result 5. *While strong learning effects can be observed in the card sequencing task, the treatment effects of an FDRS on the time needed to complete the card sequencing task are very similar in all ten rounds.*

Fig. 3 illustrates the development of the coefficient of variation of the employees' bonuses. While in the FDRS treatments, the coefficient of variation is fixed at 100% percent, Fig. 3 shows that the bonus disparities in the non-forced rating system are always lower when working in a team than when working alone, and that bonus disparities in the non-forced rating system tend to slightly decrease over time, both when working in a team and when working alone.

Fig. 4 illustrates the temporal development of the prediction error in Experiment 3. It shows that there is no clear time trend in the prediction error and that the prediction error is higher in the FDRS treatment than in the non-forced rating system in all rounds except in round 3. Figs. 5 and 6 illustrate the development of the number of words and the development of the number of task-relevant statements. Teams consistently communicate less in the FDRS treatment than in the non-forced rating system treatment. The number of words and the number of task-relevant statements are substantially lower in all ten rounds. Fig. 7 illustrates the temporal development of the coefficient of variation of the employees' bonuses. It shows that bonus disparities are significantly lower in the non-forced rating system in all ten rounds and that bonus disparities remain fairly constant over time in the non-forced rating system.

8. Discussion

The output of a team depends on the interrelated contributions of its members (Alchian and Demsetz, 1972). Our results show that in such high collaboration settings, introducing an FDRS reduces team productivity, inhibits knowledge sharing

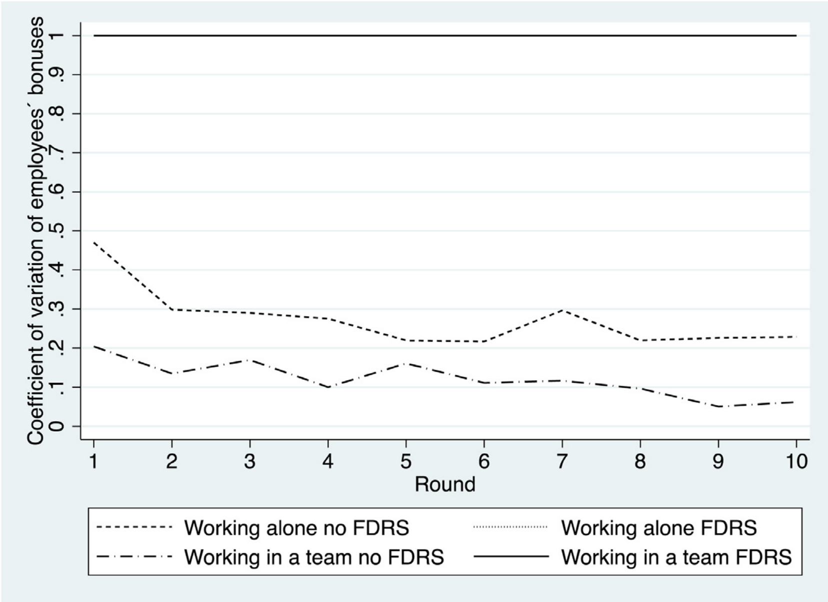

Fig. 3. Development of the coefficient of variation of the employees' bonuses in Experiments 1 and 2.

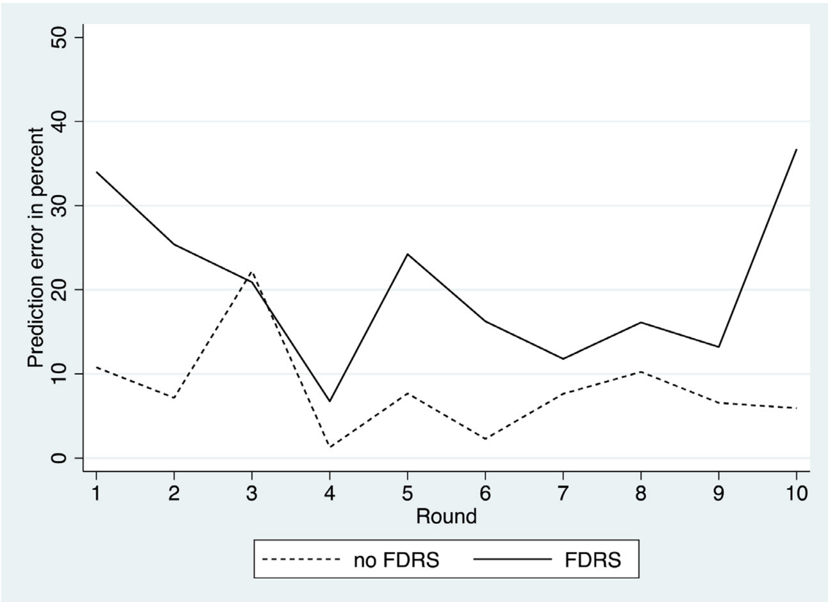

Fig. 4. Development of prediction error in percent in Experiment 3.

among team members, and decreases perceived distributive justice. In that regard, our findings substantiate a key tenet of justice judgment theory, that in collaborative settings, an equal distribution of rewards is the dominant principle of distributive justice (Leventhal, 1976), which is impossible under an FDRS. We find economically and statistically significant detrimental effects of an FDRS on team collaboration based on the evidence of two experiments, a face-to-face experiment, and a computerized experiment, with different tasks and different samples, attesting to the robustness of our results.

The negative effects of an FDRS in teams are not restricted to overt sabotage, as in the experimental study of Berger et al. (2013). Quite the contrary: In Experiment 2, when doing the card sequencing task in a team, overt sabotage was very unlikely due to the face-to-face interactions within the team, and it was not observed by the experimenter, even in the FDRS treatments. Furthermore, the supervisor watched the employees executing the card sequencing task and would have been able to sanction counterproductive behavior. We thus show that FDRSs are detrimental to positive team processes and have negative performance effects even when overt sabotage is unlikely to occur. This is of particular importance, as

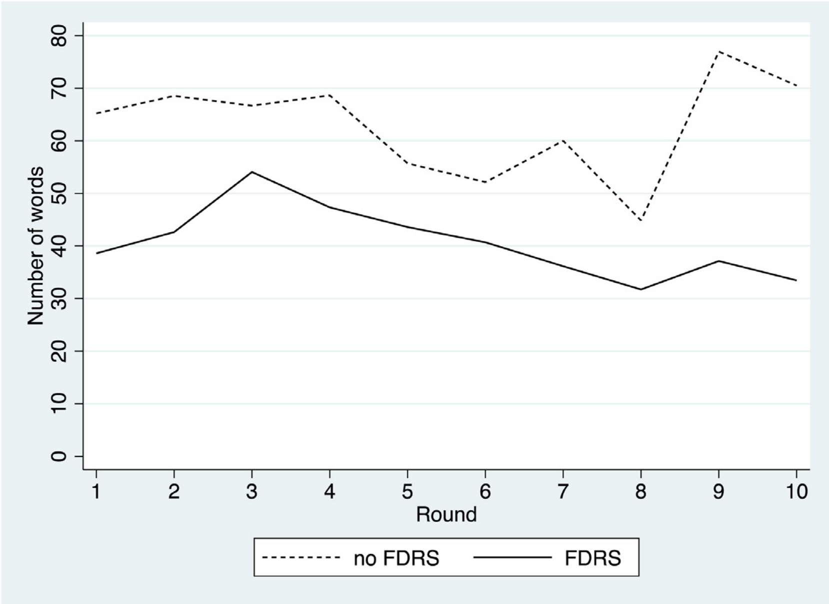

Fig. 5. Development of the number of words in Experiment 3.

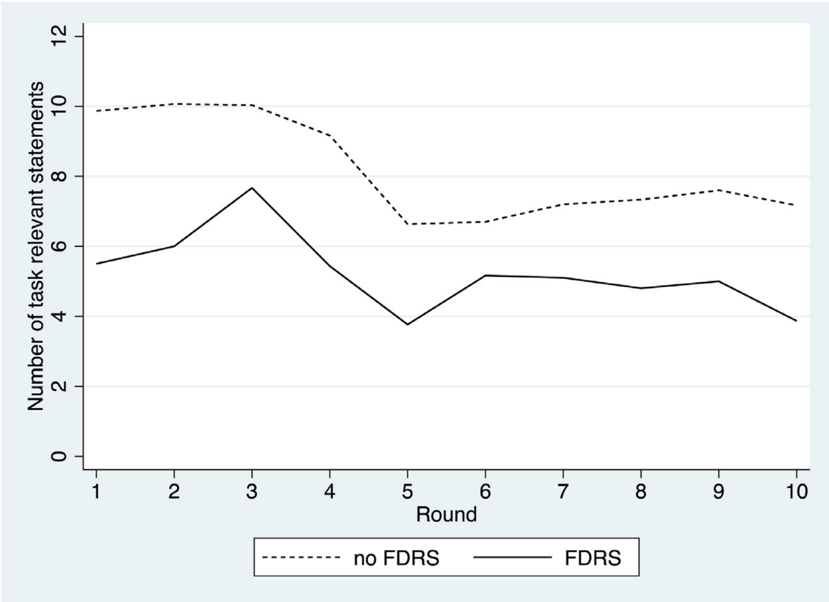

Fig. 6. Development of the number of task-relevant statements in Experiment 3.

overt sabotage is an action that many employees might not be willing to take, while decreasing personal effort and reducing communication and knowledge sharing with colleagues are more frequent under FDRSs and can have similar detrimental effects on the team outcome. Our findings suggest that the negative influence of an FDRS on team collaboration takes effect rather implicitly by reducing the team members' personal efforts and disincentivizing them to share relevant information.

We find that when working in a team, the perceived distributive justice is significantly lower in the FDRS treatments compared to the non-forced rating system treatments. According to the justice judgment theory, the equality rule, i.e. that all team members receive the same reward, is the dominant principle of distributive justice in collaborative settings (Deutsch, 1985; Leventhal, 1976). In the FDRS treatment, the equality rule cannot be applied, which creates feelings of injustice. Furthermore, our findings suggest that perceived injustice may have important consequences, such as decreased effort and performance, and that justice is, therefore, a construct of critical relevance in the context of performance appraisal and rewards.

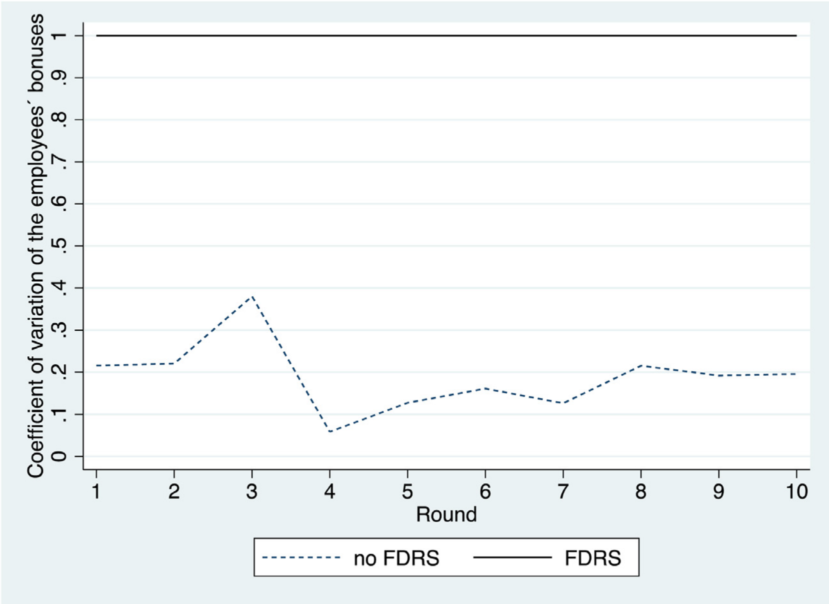

Fig. 7. Development of the coefficient of variation of the employees' bonuses in Experiment 3.

This study yields important practical implications for managers who seek to design effective incentive systems. The findings of this study indicate that FDRSs should only be implemented when individuals work alone. In team settings, managers should refrain from introducing FDRSs, so as not to decrease intra-team knowledge sharing, team performance, and perceived distributive justice. Perceived injustice may have negative effects beyond the ones we have considered here, such as increased turnover intentions (Colquitt et al., 2001). Given the increasing competition for skilled workers, perceived injustice can be particularly damaging to an organization's ability to attract and retain productive employees, and, hence, to its long-term success.

This study has some limitations concerning its scope and methods and opens avenues for future research. The external validity of our experimental results is yet to be examined. We drew on a student sample, which can be a threat to the validity of the results. We believe, however, that students are appropriate research subjects in our case, as (a) they are familiar with the processes important for our experiment, such as performance appraisal, being graded or rated, and teamwork; (b) the tasks were not specifically directed at professionals of any field, but could be solved by a broad variety of subjects with different backgrounds; and (c) most students have work experience due to internships, student jobs, and the like, and are, therefore, comparable to employees.

Although we implemented ten rounds of the task in our experimental design to consider how performance and knowledge sharing change over time, the relevant time periods in real-life situations are much longer. In most cases, employees invest months of effort between two performance appraisals, which makes the performance ratings more momentous than the ratings in our experiment. Additionally, rewards may be much higher in real-life situations and may have important implications for employees' careers. These two factors might reinforce the effects, both positive and negative, of an FDRS.

In many cases, the team that is evaluated with an FDRS comprises more than three people. Consequently, more than one person can reach top performer status, or medium performer status, or low performer status – situations that open possibilities for complex interactions between employees. For example, two employees might work together and support each other to gain advantages over the rest of the team. Moreover, findings from tournament literature provide evidence that the proportion of winners in a tournament influences team performance (Harbring and Irlenbusch, 2003; Knauer et al., 2017). Our experiment investigates teams of three under an FDRS that predetermines the proportion of winners to be one-third, which is a rather common proportion of top performers (Knauer et al., 2017). Effects may differ according to different proportions of ratings and depend on the structure of the team that is evaluated. Future research should therefore investigate the effects of an FDRS concerning different team sizes and proportions of rating categories to verify the generalizability of our findings.

In the FDRS treatments, the supervisors were forced to pay one employee zero bonus and the employee rated top double the bonus of the employee rated in the middle. Even though such a high bonus disparity is not uncommon in competitive team structures (see, e.g., Berger et al., 2013; Bloom, 1999; Scullen et al., 2005), we encourage future research to test the generalizability of our results using smaller spreads. Ehrenberg and Bognanno (1990a,b) show that prize spreads are positively related to individual performance in professional golf tournaments. Becker and Huselid (1992) confirm that prize spreads increase individual performance, drawing on panel data from professional auto racing. Harbring and Irlenbusch (2004) show that higher tournament prize spreads increase not only individual productivity but also sabotage behavior within teams. These studies suggest that decreasing the bonus spreads in the FDRS treatments is likely to also decrease the magnitudes of both the positive effect of an FDRS on individual performance and the negative effect of an FDRS on team collaboration. However, as long as the bonus spreads in the FDRS treatments are higher than the bonus spreads in the non-forced rating system treatments, the latter being small due to the centrality bias, the FDRS effects can still expected to be significant.

We show that the effect of an FDRS on performance depends on the type of work. When team collaboration is not required, and objective individual performance measures are available, an FDRS increases performance. In collaborative contexts, however, an FDRS significantly decreases performance and knowledge transfer. As predicted by justice judgment theory (Leventhal, 1976), we also find that an FDRS decreases perceived distributive justice in collaborative contexts, which may lead to withdrawals or resignations beyond the negative effects we have documented on team performance and knowledge sharing. We hope that our study inspires future research to identify more appropriate incentive schemes in interdependent team contexts.

Declaration of Competing Interest

None.

Acknowledgements

We would like to thank the editor Daniel Houser, the two referees, and the audiences at academic meetings for their very helpful comments. A special thanks goes to Matej Černe and Ola Kvaløy for their excellent remarks and suggestions on the manuscript. Finally, we would also like to thank our student assistants for their support in conducting the experiments.

Appendix I. Instructions of the first and second experiment

General instructions for participants
Note: This is an English translation of the German instructions of the four treatments:
Working alone no FDRS, Working alone FDRS, Working in a team no FDRS, Working in a team FDRS
The differences between the treatments are underlined here but not in the original instructions

We warmly welcome you to this economics experiment.

Your actions and possibly the actions of other participants will affect your payout. Therefore, it is important for you to thoroughly read the following instructions. If you have any questions, please direct them to us **before** the experiment begins.

During the experiment, we will speak of points instead of euros. Your income will therefore initially be calculated in points. The number of points you score during the experiment will be translated into euros at the end of the experiment, whereby the following rule applies:

1 point = 1 euro

At the end of today's experiment, you will receive the number of points you were rewarded in one randomly chosen round paid out in cash in euros. On the following pages, we explain the exact procedure of the experiment.

Experiment setup

In every round of this experiment, you will be in a group of four participants. The group will remain the same for the entire experiment. At the beginning of the experiment, the computer will randomly determine whether you are Participant A, B, C, or D. You will keep the same role during the entire experiment. Participant A is the supervisor. Participants B, C, and D are employees.

The experiment will consist of 10 rounds. We will now explain the procedure of one round. All 10 rounds will proceed in the same way.

Task

In each round, the following task must be solved as quickly as possible. This task will be the same for all 10 rounds. The task requires putting cards in sequences following certain rules.

Cards

[*Working alone treatments*: After reading the instructions, each of the employees, B, C, and D, will receive an identical pile of cards.] [*Working in a team treatments*: After reading the instructions, the employees, B, C, and D, will receive three identical piles of cards as a team]. Each of the three piles will contain 15 cards with a blue back, 15 cards with a red back and 15 cards with a green back. On the front side of the cards you will see that the piles only contain the cards 9, 10, jack, queen, and king, and the suits spades, hearts. and clubs (thus, there are no diamonds and no cards lower than 9 or higher than king). Each card appears exactly once per back side color. Hence, there is, for example, exactly one queen of spades with a red back.

Aim: Sequences

The aim is to put all cards into sequences as quickly as possible. Each of the sequences has to meet the following requirements:

1. The sequence has to start with a 9 followed by 10, jack, queen, and king in this exact order.
2. The suits have to alternate within each sequence in the following order: spades, clubs, hearts. It does not matter which suit a sequence starts with.
3. The colors of the backs of the cards must alternate in the following order: blue, green, red. It does not matter which color a sequence starts with.

Examples of correct sequences:

Front view Back view

[*Working in a team treatments only*: **Important: You start working on one pile as a team. Only when all cards of this pile are put into sequences may you start working on the next pile.**]

Procedure in one round

One round proceeds as follows.

1. Employees B, C, and D solve the described task as quickly as possible. The supervisor, Participant A, is present but does not participate in the task. The employees may talk to each other during the task. They are not allowed to talk in between rounds, though. The employees and the supervisor are not allowed to talk to each other during the entire experiment.
2. [*Individual treatments*: The experimenter measures the time that each employee needs to correctly put the cards into the 9 sequences, to put an elastic band around each sequence and to put the sequences in an envelope that is provided.] [*Team treatments*: The experimenter measures the time that the employees B, C, and D need as a team to correctly put the cards into the 27 sequences, to put an elastic band around each sequence and to put the sequences in the attached envelope.]

3. **Points for employees B, C and D**
   The supervisor subsequently assesses the respective performances of Employees B, C, and D and assigns points on this basis. In each round, the supervisor must distribute 45 points among Employees B, C, and D. [*FDRS treatments only:* **One employee must receive 30 points, another 15 points and yet another 0 points**]. The points are noted by the supervisor on the sheet "points assignment". The experimenter then informs each participant about the points he/she earned in the current round.
   In addition to the points assigned by the supervisor, each employee receives a fixed amount of 12 points. For every incorrect sequence of the respective round, i.e. for every sequence that does not meet the requirements, 0.5 points are deducted.

Points for the supervisor (participant A)

The supervisor receives a fixed amount of 12 points. He can receive additional points depending on the employees' speed in solving the task. [*Working alone treatments*: For this calculation, the average time of Employees B, C, and D is computed and rounded down to the next whole minute (i.e. 12:45 min becomes 12 min).] [*Working in a team treatments*: The time needed is rounded down to the next whole minute (i.e. 12:45 min becomes 12 min).] For every minute that this time is less than 20 min, the supervisor receives one point. [*Working alone treatments:* The supervisor loses 0.5 points, however, for every averaged incorrect sequence in the respective round. The average of incorrect sequences in one round is computed, i.e. the average number of sequences that do not meet the criteria, and is rounded down to the next whole sequence (i.e. 4.33 sequences become 4 sequences).] [*Working in a team treatments*: The supervisor loses 0.5 points for every incorrect sequence in the respective round, i.e. for every sequence that does not meet the criteria.]

After completion of all ten rounds, the points from one randomly selected round will be used for the payout. For example, if round three was selected, the participants would receive money according to the points they had earned in round three.

Amazon Voucher

The team with the **quickest** [*Working in a team treatments only*: **average**] **time in one round** with correct sequences will receive an Amazon voucher worth €10 per group member (total value therefore: €40). The winners will receive the voucher via e-mail.

Appendix II. Instructions of the third experiment

General instructions for participants
Note: This is an English translation of the German instructions of the two treatments:
no FDRS, FDRS
The differences between the treatments are underlined here but not in the original instructions

We warmly welcome you to this economics experiment.

Your own actions and possibly the actions of other participants will affect your payout. Therefore, it is important for you to thoroughly read the following instructions. If you have any questions, please direct them to us **before** the experiment begins.

During the experiment, we will speak of points instead of euros. Your income will therefore initially be calculated in points. The number of points you score during the experiment will be translated into euros at the end of the experiment, whereby the following rule applies:

1 point = 1 euro

At the end of today's experiment, you will receive the number of points you were rewarded in one randomly chosen round paid out in cash in euros. On the following pages, we explain the exact procedure of the experiment.

The use of mobile phones, tablets, or similar devices during the experiment will lead to exclusion from the experiment. Please make sure that these devices are switched off and stowed in a bag. This is essential for the success of the experiment.

Experiment setup

In every round of this experiment, you will be in a group of four participants. The group will remain the same for the entire experiment. At the beginning of the experiment, the computer will randomly determine whether you are Participant A, B, C, or D. You will keep the same role during the entire experiment. Participant A is the supervisor. Participants B, C, and D are employees. Communication with the other team members will be exclusively virtual. You will not learn which participant in the room has what role, meaning that you will remain anonymous throughout the whole experiment.

The experiment will consist of 10 rounds. We will now explain the procedure of one round. All 10 rounds will proceed in the same way.

During the whole experiment, it is forbidden to talk. Conversations lead to exclusion from the experiment. Employees B, C, and D, however, have the possibility to write messages into a chatroom. Each message can be read by the other employees of the same group. Supervisor A cannot read these messages nor write any messages.

Task

Before the start of the actual experiment, Employees B, C, and D are asked to each write a short message in the chatroom to make sure that every employee can send and receive messages. Afterwards, the first round will start. One round proceeds as follows.

1. The supervisor (Participant A) and the employees (Participants B, C, and D) see the same task on their screens. In this task, each employee has to make a prediction as precisely as possible (e.g. How high is the Eiffel Tower?). Additionally, some background information will be provided on the screen.
2. Employee B, C and D have at maximum 10 min to make a prediction and click on "OK". If an employee does not make a prediction within 10 min, his predicted value will be set to "zero" automatically. The supervisor A cannot make a prediction.
3. As soon as all employees have made a prediction (at the latest after 10 min), the correct solution and the employees' predictions will be shown to the supervisor and the employees.
4. **Points for employees B, C and D**
The supervisor subsequently assesses the respective performances of employees B, C and D and assigns points on this basis. In each round, the supervisor must distribute 60 points among the employees B, C and D. [*FDRS only:* **One employee must receive 40 points, another 20 points and yet another 0 points**]. The supervisor types the points for each employee into the computer. Each employee subsequently learns about the points he/she has earned in the current round.
In addition to the points assigned by the supervisor, each employee receives a fixed amount of 15 points.

Points for the supervisor (participant A)

The supervisor receives a fixed amount of 15 points. He can receive additional points depending on the precision of employees' predictions. Therefore, the average of the absolute deviations of employees' predictions is calculated. For every percentage point that this average is less than 25%, the supervisor receives 1 euro. Accordingly, the supervisor receives e.g.

20 points.............................if the average deviation is 5%
15 points.............................if the average deviation is 10%
5 points..............................if the average deviation is 20%
0 points..............................if the average deviation is >24%

If the true value e.g. is 100 and the average absolute deviation of the predictions is 20, the percentage deviation from the true value is 20%. In this case the supervisor would receive additional $5 \times 1 = 5$ points.

During all ten rounds, the employees can communicate via messages in the chatroom. After completion of all ten rounds, the points from one randomly selected round will be used for the payout. For example, if round three was selected, the participants would receive money according to the points they had earned in round three.

References

Adams, J.S., 1963. Toward an understanding of inequity. J. Abnorm. Psychol. 67, 422–436.
Angelovski, A., Brandts, J., Sola, C., 2016. Hiring and escalating bias in subjective performance evaluations: a laboratory experiment. J. Econ. Behav. Organ. 121, 114–129.
Alchian, A.A., Demsetz, H., 1972. Production, information costs, and economic organization. Am. Econ. Rev. 62, 777–795.
Bachrach, D.G., Powell, B.C., Collins, B.J., Richey, R.G., 2006. Effects of task interdependence on the relationship between helping behavior and group performance. J. Appl. Psychol. 91, 1396–1405.
Bartol, K.M., Srivastava, A., 2002. Encouraging knowledge sharing: the role of organizational reward systems. J. Leadersh. Organ. Stud. 9, 64–76.
Baumeister, R.F., Leary, M.R., 1995. The need to belong: desire for interpersonal attachments as a fundamental human motivation. Psychol. Bull. 117, 497–529.
Becker, B.E., Huselid, M.A., 1992. The incentive effects of tournament compensation systems. Admin. Sci. Q. 37, 336–350.
Beersma, B., Hollenbeck, J.R., Humphrey, S.E., Moon, H., Conlon, D.E., Ilgen, D.R., 2003. Cooperation, competition, and team performance: toward a contingency approach. Acad. Manag. J. 46, 572–590.
Berger, J., Harbring, C., Sliwka, D., 2013. Performance appraisals and the impact of forced distribution – an experimental investigation. Manag. Sci. 59, 54–68.
Bierly, P.E., Stark, E.M., Kessler, E.H., 2009. The moderating effects of virtuality on the antecedents and outcome of NPD team trust. J. Prod. Innov. Manag. 26, 551–565.
Bloom, M., 1999. The performance effects of pay dispersion on individuals and organizations. Acad. Manag. J. 42, 25–40.
Blume, B.D., Baldwin, T.T., Rubin, R.S., 2009. Reactions to different types of forced distribution performance evaluation systems. J. Bus. Psychol. 24, 77–91.
Cabrera, A., Cabrera, E.F., 2002. Knowledge-sharing dilemmas. Organ. Stud. 23 (5), 687–710.
Chambers, C.R., Baker, W.E., 2020. Robust systems of cooperation in the presence of rankings: how displaying prosocial contributions can offset the disruptive effects of performance rankings. Organ. Sci. published online ahead of print.
Colquitt, J.A., 2001. On the dimensionality of organizational justice: a construct validation of a measure. J. Appl. Psychol. 86 (3), 386–400.
Colquitt, J.A., Conlon, D.E., Wesson, M.J., Porter, C.O., Ng, K.Y., 2001. Justice at the millennium: a meta-analytic review of 25 years of organizational justice research. J. Appl. Psychol. 86 (3), 425–445.
Danilov, A., Harbring, C., Irlenbusch, B., 2019. Helping under a combination of team and tournament incentives. J. Econ. Behav. Organ. 162, 120–135.
Deutsch, M., 1985. Distributive Justice: a Social-Psychological Perspective. Yale University Press, New Haven.
Ehrenberg, R.G., Bognanno, M.L., 1990a. The incentive effects of tournaments revisited: evidence from the European PGA tour. Ind. Lab. Relat. Rev. 43, 74–88.
Ehrenberg, R.G., Bognanno, M.L., 1990b. Do tournaments have incentive effects? J. Polit. Econ. 98, 1307–1324.
Falk, A., Fischbacher, U., 2006. A theory of reciprocity. Games Econ. Behav. 54 (2) 293–135.

Falk, A., Ichino, A., 2006. Clean evidence on peer effects. J. Lab. Econ. 24 (1), 39–57.
Fehr, E., Herz, H., Wilkening, T., 2013. The lure of authority: motivation and incentive effects of power. Am. Econ. Rev. 103, 1325–1359.
Ferrin, D.L., Dirks, K.T., 2003. The use of rewards to increase and decrease trust: mediating processes and differential effects. Organ. Sci. 14, 18–31.
Fey, C.F., Furu, P., 2008. Top management incentive compensation and knowledge sharing in multinational corporations. Strateg. Manag. J. 29, 1301–1323.
Fischbacher, U., 2007. z-Tree: zurich toolbox for ready-made economic experiments. Exp. Econ. 10, 171–178.
Franck, E., Nüesch, S., 2011. The effect of wage dispersion on team outcome and the way team outcome is produced. Appl. Econ. 43, 3037–3049.
Garcia, S.M., Tor, A., 2007. Rankings, standards, and competition: task vs. scale comparisons. Organ. Behav. Hum. Decis. Processes 102 (1), 95–108.
Grant, R.M., 1996. Toward a knowledge-based theory of the firm. Strateg. Manag. J. 17, 109–122.
Greenberg, J., 1987. A taxonomy of organizational justice theories. Acad. Manag. Rev. 12, 9–22.
Greiner, B., 2015. Subject pool recruitment procedures: organizing experiments with ORSEE. J. Econ. Sci. Assoc. 1, 114–125.
Grote, R.C., 2005. Forced Ranking: Making Performance Management Work. Harvard Business School Press, Boston, MA.
Guralnik, O., Rozmarin, E., So, A., 2004. Forced distribution: is it right for you? Hum. Resour. Dev. Q. 15, 339–345.
Hamman, J.R., Lowenstein, G., Weber, R.A., 2010. Self-interest through delegation. An additional rationale for the principal-agent relationship. Am. Econ. Rev. 100, 1826–1846.
Harbring, C., Irlenbusch, B., 2003. An experimental study on tournament design. Lab. Econ. 10, 443–464.
Harbring, C., Irlenbusch, B., 2004. Anreize zu produktiven und destruktiven Anstrengungen durch relative Entlohnung. Zeitschrift für betriebswirtschaftliche Forschung 56, 546–576.
Harbring, C., Irlenbusch, B., 2008. How many winners are good to have? On tournaments with sabotage. J. Econ. Behav. Organ. 65, 682–702.
Harbring, C., Irlenbusch, B., 2011. Sabotage in tournaments: evidence from a laboratory experiment. Manag. Sci. 57, 611–627.
Katz-Navon, T.Y., Erez, M., 2016. When collective- and self-efficacy affect team performance. Small Group Res. 36, 437–465.
Knauer, T., Sommer, F., Wöhrmann, A., 2017. Tournament winner proportion and its effect on effort: an investigation of the underlying psychological mechanisms. Eur. Account. Rev. 26, 681–702.
Kozlowski, S.W., 2017. Enhancing the effectiveness of work groups and teams: a reflection. Perspect. Psychol. Sci. 13, 1–8.
Kozlowski, S.W.J., Bell, B.S., 2013. Work groups and teams in organizations: review update. In: Schmitt, N., Highhose, S. (Eds.), Handbook of Psychology, Vol.12. Industrial and Organizational Psychology. Wiley, Hoboken, NJ 412–469.
Kozlowski, S.W.J., Ilgen, D.R., 2006. Enhancing the effectiveness of work groups and teams. Psychol. Sci. Public Interest 7, 77–124.
Lazear, E.P., 1989. Pay equality and industrial politics. J. Polit. Econ. 97, 561–580.
Lazear, E.P., Rosen, S., 1981. Rank-order tournaments as optimum labor contracts. J. Polit. Econ. 89, 841–864.
Leszczyc, P.T.L.P., Rothkopf, M.H., 2010. Charitable motives and bidding in charity auctions. Manag. Sci. 56, 399–413.
Leventhal, G.S., 1976. Fairness in social relationships. In: Thibaut, J.W., Spence, J.T., Carson, R.C. (Eds.), Contemporary Topics in Social Psychology. General Learning Press, Morristown, N. J. (pp. 211–239).
Levine, D.I., 1991. Cohesiveness, productivity, and wage dispersion. J. Econ. Behav. Organ. 15, 237–255.
Machegiani, L., Reggiani, T., Rizzolli, M., 2016. Loss averse agents and lenient supervisors in performance appraisals. J. Econ. Behav. Organ. 131, 183–197.
Mesmer-Magnus, J.R., Dechurch, L.A., 2009. Information sharing and team performance: a meta-analysis. J. Appl. Psychol. 94, 535–546.
Moon, S.H., Scullen, S.E., Latham, G.P., 2016. Precarious curve ahead: the effects of forced distribution rating systems on job performance. Hum. Resour. Manag. Rev. 26, 166–179.
Mulligan, J.R., Bull Schaefer, R.A., 2011. A new hope for rank and yank. J. Leadersh. Organ. Stud. 18, 385–396.
Murphy, K.R., Cleveland, J.N., 1995. Understanding Performance Appraisal: Social, Organizational, and Goal-Based Perspectives. Sage Publications, Thousand Oaks, CA.
Nelson, A.J., 2016. How to share "a really good secret": managing sharing/secrecy tensions around scientific knowledge disclosure. Organ. Sci. 27, 265–285.
Nonaka, I., 1994. A dynamic theory of organizational knowledge creation. Organ. Sci. 5, 14–37.
Pfeffer, J., Langton, N., 1993. The effect of wage dispersion on satisfaction, productivity, and working collaboratively: evidence from college and university faculty. Admin. Sci. Q. 38, 382–407.
Prendergast, C., 1999. The provision of incentives in firms. J. Econ. Lit. 37, 7–63.
Riege, A., 2005. Three-dozen knowledge-sharing barriers managers must consider. J. Knowl. Manag. 9, 18–35.
Salas, E., Sims, D.E., Burke, C.S., 2016. Is there a "Big Five" in teamwork? Small Group Res. 36, 555–599.
Schleicher, D.J., Bull, R.A., Green, S.G., 2008. Rater reactions to forced distribution rating systems. J. Manag. 35, 899–927.
Scullen, S.E., Bergey, P.K., Aiman-Smith, L., 2005. Forced distribution rating systems and the improvement of workforce potential: a baseline simulation. Pers. Psychol. 58, 1–32.
Shaw, J.D., Gupta, N., Delery, J.E., 2002. Pay dispersion and workforce performance: moderating effects of incentives and interdependence. Strateg. Manag. J. 23 (6), 491–512.
Siegel, P.A., Hambrick, D.C., 2005. Pay disparities within top management groups: evidence of harmful effects on performance of high-technology firms. Organ. Sci. 16, 259–274.
Stewart, S.M., Gruys, M.L., Storm, M., 2010. Forced distribution performance evaluation systems: advantages, disadvantages and keys to implementation. J. Manag. Organ. 16, 168–179.
Strom, D.L., Sears, K.L., Kelly, K.M., 2014. Work engagement: the roles of organizational justice and leadership style in predicting engagement among employees. J. Leadersh. Organ. Stud. 21, 71–82.
Van der Vegt, G., van de Vliert, E., 2002. Intragroup interdependence and effectiveness. J. Manag. Psychol. 17, 50–67.
Van Knippenberg, D., Mell, J.N., 2016. Past, present, and potential future of team diversity research: from compositional diversity to emergent diversity. Organ. Behav. Hum. Decis. Processes 136, 135–145.
Welch, J., Byrne, J.A., 2003. Jack: Straight From the Gut. Warner Books, New York.
KPMG Consulting. 2000. Knowledge management research report 2000 (KPMG Consulting Reports).