Department of Strategy and Innovation
Working Paper No. 04/2020

How calibration committees can mitigate performance evaluation bias:

An analysis of implicit incentives

Isabella Grabner
Judith Künneke
Frank Moers

January 2020

WIRTSCHAFTS UNIVERSITÄT WIEN VIENNA UNIVERSITY OF ECONOMICS AND BUSINESS

How calibration committees can mitigate performance evaluation bias: An analysis of implicit incentives

Isabella Grabner[a,c]
Isabella.Grabner@wu.ac.at

Judith Künneke[b]
j.kunneke@maastrichtuniversity.nl

Frank Moers[b]
f.moers@maastrichtuniversity.nl

[a] Vienna University of Economics and Business
[b] Maastricht University School of Business and Economics
[c] Research Centre for Education and the Labour Market (ROA)

October 2019

This paper has benefited from comments and suggestions by Eddy Cardinaels (editor), two anonymous reviewers, Jasmijn Bol, Chris Ittner, Ken Merchant, Paolo Perego, two anonymous reviewers, and seminar participants at Frankfurt School of Finance and Management, Maastricht University, National Chengchi University, Tilburg University, University of Hong Kong, WU Vienna, the 9th EIASM Conference on New Directions in Management Accounting, the 2015 EAA Annual Conference, the 2015 GMARS, the 2015 AAA Annual Meeting, the 2016 MAS Meeting, the 2016 Dutch Accounting Research Conference, and the Emerging Management Accounting Scholars Symposium 2016.

How calibration committees can mitigate performance evaluation bias: An analysis of implicit incentives

ABSTRACT: While prior research on performance evaluation bias has mainly focused on the determinants and consequences of rating errors, we investigate how a firm can provide implicit incentives to supervisors to mitigate these errors via its calibration committee. We empirically examine the extent to which a calibration committee incorporates supervisors' evaluation behavior with respect to their subordinates in the performance evaluation outcomes, i.e., performance ratings and promotion decisions, for these supervisors. In our study, we distinguish between lack of skills and opportunism as two important facets of evaluation behavior, which we expect the calibration committee to address differently. Using panel data of a professional service firm, we show that supervisors' opportunistic behavior to strategically inflate subordinates' performance ratings is disciplined through a decrease in the supervisors' own performance rating, while the supervisors' skills to provide less compressed and thus more informative performance ratings is rewarded through a higher likelihood of promotion.

Keywords: subjective performance evaluation, performance evaluation bias, calibration committees, promotions, supervisory skills

I. INTRODUCTION

An increasingly used management control innovation aimed at mitigating performance evaluation bias is to allocate decision rights over performance ratings to so-called calibration committees (e.g., Demeré, Sedatole, and Woods 2019). These committees, typically comprised of higher-level managers, review and potentially adjust a subordinate's initial performance rating provided by the direct supervisor. Demeré et al. (2019) provide the first evidence on the role of calibration committees in subjective performance evaluation systems and examine the nature of calibration committee adjustments to supervisors' reported ratings of their subordinates. We extend this stream of research by investigating whether the calibration committee attaches consequences to the way supervisors evaluate and rate their subordinates. In particular, we examine whether a firm's calibration committee provides implicit incentives to supervisors to mitigate rating errors by incorporating supervisors' evaluation behavior with respect to their subordinates in their own performance evaluation outcomes, i.e., performance ratings and promotion decisions.

A relatively recent survey finds that more than 50% of firms set up calibration committees to validate the performance and potential of each employee against the relevant peer group (Hastings 2012). Accordingly, the primary stated purposes of calibration committees are to mitigate inter-rater differences, increase the information content of performance ratings, ensure consistency in the application of performance standards, attenuate the biases that supervisors introduce, but also to help identify supervisors who induce such rating errors (Demeré et al. 2019). During the calibration sessions, higher-level managers discuss the subordinates' performance based on information from a variety of stakeholders including the supervisors, who all can provide different insights about the subordinates. If considered necessary, the calibration committee adjusts the initial ratings that have been recommended by the supervisors. In this way, the calibration

committee tries to correct error-prone ratings provided by the supervisors, but at the same time tries to identify, to the best of its ability, the supervisors that cause biased performance ratings.

Now the calibration committee basically has two options to use this information: either it only adjusts the initial ratings without any further consequences for the supervisors, or it attaches consequences to these adjustments and thereby provides implicit incentives to supervisors to prepare more accurate, and thus more useful performance ratings of their subordinates. Given the costly consequences of performance evaluation bias for the firm in terms of incorrect personnel decisions and inadequate resource allocation, we expect that, over and above adjusting the initial ratings, the calibration committee will make use of this information for incentivizing supervisors to become better at their task, thereby decreasing the cost of performance evaluation bias.

In our study, we focus on the lack of supervisory skills and opportunism as two important facets of evaluation behavior that the calibration committee aims to address. We develop theory that links each of these facets to specific biases. More specifically, the lack of relevant supervisory skills is expected to be reflected in more compressed performance ratings, while opportunism is expected to be reflected in systematically higher-than-warranted ratings. Firms therefore face the dilemma that, while the input of supervisors is crucial when evaluating subordinates, the performance ratings might be biased by the supervisors. We expect that, in addition to correcting any bias by adjusting the ratings recommended by the supervisors, the calibration committee tries to mitigate opportunistic bias by providing disincentives for opportunistic behavior, while it aims to mitigate skill-driven bias by providing incentives for the investment in supervisory skills. We argue that the calibration committee uses the annual performance ratings and promotions to address opportunistic and skill-driven bias, and we describe the relative importance of these two mechanisms in solving these different types of incentive issues in our hypotheses development.

To analyze the decisions of the calibration committee, we conduct a field study at an international professional service provider. The typical calibration committee in professional service firms consists of higher-level managers as well as the direct supervisors and their peers, and has decision rights over final performance ratings and associated monetary compensation, but is also entrusted with promotion and termination decisions. Using proprietary archival data of 952 individual supervisors for the time period of 2010-2012, we investigate the incremental effect of different measures of supervisors' evaluation behavior on their own performance ratings and promotion opportunities, over and above other signals of their past and current performance. Our results indicate that, when the calibration committee observes signals of opportunistic rating behavior, the supervisor is disciplined with a decrease in her performance rating, which has both compensation and career consequences. We further find that signals relating to the supervisor's skill to discriminate among subordinate performance levels are positively associated with the likelihood of being promoted, which is consistent with the argument that promotion mechanisms are effective for inducing skill acquisition (Grabner and Moers 2019; Kahn and Huberman 1988; Prendergast 1993). Taken together, the firm encourages supervisors via its calibration committee to invest in their long-term skills and to refrain from activities that provide short-term rewards at the cost of the firm. Follow-up interviews with partners at our research site as well as other members of the professional service provider's network confirm our inferences.

Our study makes several contributions to the incentive and performance evaluation literature. While the use of calibration committees has become an important trend in business practice (Risher 2011, 2014), academic research on the role of calibration committees in decreasing the cost of subjectivity is still scarce. The few studies addressing calibration committees focus on the impact of the presence of calibration committees on supervisor evaluation behavior (Arshad, Cardinaels, and Dierynck 2017), the process through which calibration committees aim to mitigate bias (Lillis,

Malina, and Mundy 2017), and the nature as well as determinants of rating adjustments through the calibration committee (Bol, Aguiar, and Lill 2019; Demeré et al. 2019). While these concurrent studies focus on the calibration process, we provide insights into the consequences thereof. We contribute to this emerging stream of research by investigating the consequences the calibration committee attaches to the way supervisors evaluate and rate their subordinates, in an attempt to incentivize more accurate, and thus less biased rating behavior.

Relatedly, very little attention has been paid to the mechanisms that firms use to reduce rating errors (recent exceptions are Bol, Kramer, and Maas 2016, Demeré et al. 2019, and Lillis et al. 2017). Our results indicate that a firm's calibration committee can provide implicit incentives to supervisors to mitigate rating errors by incorporating supervisors' evaluation behavior with respect to their subordinates in their own performance evaluation outcomes. In particular, supervisors are rewarded when they sufficiently differentiate among subordinate performance levels and signal their supervisory skills, while they are disciplined when expected to have engaged in opportunistic rating behavior, which includes providing inflated performance ratings. Despite the suggested motivational effects of inflated ratings (Bol 2011), our results indicate that the calibration committee perceives higher ratings than warranted as a distortion of the employee performance assessment, and incentivizes supervisors to reduce such errors. Further, our study contributes to the growing literature on how to incentivize the acquisition of human capital in general (Grabner and Moers 2019), and in particular how to encourage supervisors to effectively pass on their knowledge and experience to their subordinates and guide them in their professional and personal development. Given the increasing importance of human capital for firm performance in today's growing knowledge society, the results of our study give an indication of how firms can incentivize supervisors to effectively use, or further develop, their supervisory skills.

II. THEORY DEVELOPMENT AND HYPOTHESES

Costs of performance evaluation bias

The evaluation of subordinates' performance relies to a large extent on a supervisor's subjective assessment. The major advantage of relying on such subjective judgment is that supervisors are expected to have superior information about their subordinates that is not captured by objective performance indicators (Gibbs, Merchant, Van der Stede, and Vargus 2004; Prendergast 1999). However, as soon as subjectivity enters the evaluation process, the complexity of the appraisal process as well as personal limitations provide sufficient opportunity for errors in performance evaluations, both intentional and unintentional ones. That is, independent of the cause of those rating errors (biases), supervisors potentially provide inaccurate information and hence distorted signals about subordinate performance, despite their possession of superior information.

This distortion inhibits the firm's decision-making in several ways. First, management and the human resource department receive inaccurate signals about the true performance of the workforce, and hence cannot efficiently (re-)allocate workers to particular jobs and tasks. Furthermore, while such rating errors in general imply both too high as well as too low ratings, supervisors lean towards upward-biasing their subjective ratings (Bol 2008, 2011; Moers 2005). Since those ratings are often tied to some type of compensation, the upward-biased ratings result in unjustified increased personnel costs, and potential budgeting distortions for future periods. On top, the respective subordinates receive a misrepresentation of their performance. Therefore, overstating a subordinate's performance actually disguises the necessity for further development, and harms the subordinate's potential for performance improvement. Even more problematic, if supervisors who tend towards such rating errors are promoted to higher ranks in the organization, these already substantial costs are amplified because after promotion they typically supervise a larger number of subordinates who also have more wide-ranging decision rights.

The role of the calibration committee in discovering and mitigating bias

To minimize the occurrence of errors and bias, and thus mitigate the costs thereof, firms often install calibration committees (hereafter CC) to review and potentially adjust the initial performance ratings provided by the supervisors (Demeré et al. 2019). During the calibration sessions, higher-level management deliberates with each supervisor on the performance of the supervisor's subordinates and their initial performance ratings. At the same time, such a session provides the CC with the opportunity to scrutinize the supervisor's performance evaluation and rating behavior, and to assess her potential contribution to the aforementioned costs of rating errors. First of all, these sessions require the supervisor to thoroughly prepare in terms of gathering and integrating various information related to the performance of each subordinate, and a clear justification for the initial performance rating for each subordinate (Arshad et al. 2017). Yet, the proper preparation of information and her argumentation proficiency is only a subset of the information that the CC can use to evaluate the supervisor. Due to the close personal interaction during the calibration sessions and the use of questioning techniques, verbal and nonverbal cues are helpful in detecting whether an individual supervisor engages in impression management tactics, such as presenting themselves or their subordinates in an exaggerated positive way (Roulin, Bangerter, and Levashina 2015). Furthermore, sessions with multiple supervisors allow a comparison of how different supervisors act and bring forward arguments about their subordinates' performance, and thus increase the chance that the CC can infer information about each supervisor's rating and evaluation behavior. That is, while detecting biases remains an inherently difficult task and hence detection is far from guaranteed, the direct interaction with the supervisors enables the CC to at least partly monitor the supervisors' rating and evaluation behavior.

Therefore, whenever the CC is convinced to observe signals of inaccurate rating behavior, it adjusts the ratings that have been recommended by the supervisors. The adjustments to the initial

performance rating provided by the supervisors signal to the supervisors that the CC considers their ratings as inaccurate. Thus, the adjustments provide immediate feedback to the supervisors that their evaluation behavior needs to be improved.

Given the considerable costs of inadequate performance evaluations, simply adjusting the initial ratings without any further actions by the CC may, however, not be sufficient for mitigating bias and hence increasing the quality of the supervisors' evaluation behavior. Instead, we argue that the CC will incorporate its assessment of supervisors' evaluation behavior when later evaluating the supervisors' own performance and promotion. Given the CC's insights from the previously described sessions with the supervisors, it has information at hand that can be used to attach consequences, both in a negative as well as a positive sense, to the supervisors' evaluation behavior, and thereby provide implicit incentives to supervisors to prepare accurate performance ratings of their subordinates.

We distinguish two important facets of evaluation behavior that the CC wants to attach consequences to: opportunistic behavior and supervisory skills. Regarding the former, the primary purpose of supervision is to guide and support subordinates in their work-related decisions and career. As such, subordinate performance reflects to some extent the actual supervision efforts provided for which the supervisor is responsible, and thus held accountable for in her own performance evaluation.[1] This implies that higher performing subordinates will likely positively reflect on the supervisor's own performance. Consequently, if her payoff is linked to the measurement of subordinate performance, it creates a purely economic motivation to manipulate this performance indicator (Ilgen, Mitchell, and Fredrickson 1981; Merchant and Van der Stede

[1] A field study conducted by Dineen, Lewicki, and Tomlinson (2006) indicates that by providing guidance to subordinates, supervisors are able to induce greater levels of organizational citizenship and less deviant behavior, which in turn enhances performance (Podsakoff, Ahearne, and MacKenzie 1997). Along these lines, Bol, Estep, Moers, and Peecher (2018) find that supervisors high on tacit knowledge are better in developing subordinate competencies.

2017).[2] Even without an explicit financial reward, the supervisor's behavior is predominantly shaped by the expectations created by her superiors (Pfeffer and Salancik 1975), thus increasing the likelihood of impression management if expectations cannot be met.[3] Hence, when a supervisor is asked to subjectively assess subordinate performance, the supervisor might opportunistically upward-bias (some of) her subordinates' performance ratings to feign her productive efforts as supervisor. In other words, with the prospect in mind that the CC will not uncover this opportunistic behavior, supervisors have incentives to take actions that distort performance evaluations in their favor, thereby creating the aforementioned costs of rating errors for the firm.[4]

Besides this opportunistic motive to induce bias, supervisors might also lack the necessary skills to conduct a proper performance review and provide accurate ratings to their subordinates. Prior research argues that supervisors experience substantial discomfort when evaluating others and often have difficulties in confronting subordinates with negative performance feedback and ratings (e.g., Bol 2011; Moers 2005; Villanova and Bernardin 1989). That is, even when the supervisor has gathered superior information about the subordinate, she may refrain from sharing this information with the subordinate if it contains negative feedback and a low performance rating, as it leads to a potential dispute with the subordinate. The supervisor is motivated to avoid such a confrontation so as to avoid the associated high personal costs.[5] It logically follows that only for a

[2] Consistent with this expectation, Rosaz and Villeval (2012) provide evidence that supervisors selfishly misreport subordinate performance to present themselves in a better light if they expect a certain payoff attached to this presentation.

[3] Impression management tactics are known to be present at various levels in firms, ranging from the subordinate ingratiating the supervisor to receive better performance evaluations, selecting highly paid peers to increase CEO compensation, to strategically choosing favorable information for earnings benchmarks (Faulkender and Yang 2010; Gordon 1996; Schrand and Walther 2000).

[4] We label supervisors' behavior that is purely geared towards the increase of their performance rating to positively affect their financial payoff as *opportunistic*. That is, they decide to engage in impression management tactics to increase the chances of benefitting financially in the short-term instead of achieving long-term goals and rewards by actually increasing subordinate performance in a steady manner through adequate supervision.

[5] The costs of providing negative feedback relate to the unpleasant nature of giving such feedback, as well as the damage that such feedback can cause to the supervisor-subordinate relationship and subordinate motivation (Larson 1984, 1989).

subordinate who is below some performance threshold, the supervisor will provide a higher rating than warranted, i.e., lenient ratings, to avoid such psychological costs. This implies that the lower part of the rating scale is inflated, which decreases the variance in subordinates' initial performance ratings and hence causes compression. In essence, supervisors who lack the skills to provide accurate performance feedback and ratings are more likely to exhibit performance evaluation bias in terms of compressed ratings.

However, some supervisors are more able than others to overcome the disutility associated with the consequences of providing such feedback (Harris 1994; Murphy and Cleveland 1991; Westermann et al. 2015). Such supervisors have not only internalized the firm's objective to obtain accurate ratings, but also recognize the importance to foster the development of their subordinates in line with their responsibility as supervisor. As such, supervisors who are able and willing to discriminate among their subordinate performance levels and make use of the entire rating scale signal high supervisory skills. In line with that argumentation, Künneke (2017) provides empirical evidence that supervisors who exhibit a larger variance in their initial subordinate ratings contribute to a larger extent to their subordinates' performance improvement than their non-discriminating counterparts.

Despite both opportunism and lack of supervisory skills causing performance evaluation bias, they are two fundamentally different problems. That is, while the former is rather effort-related (both in the sense of not sufficiently providing the necessary effort to provide proper supervision as well as the ill-motivated effort to engage in impression management), the latter is more skill-related. These differences are important for the CC to take into account when designing mechanisms to mitigate bias. Given that annual performance ratings and promotions are the most common incentive mechanisms in firms, we investigate the relative importance of these two mechanisms in addressing opportunism and insufficient skill development.

How the calibration committee mitigates bias

Mitigating opportunistic rating behavior (H1a and H1b). Given that it is close to impossible for the CC to take away the economic incentives that trigger opportunism, it has to rely on dis-incentivizing supervisors' opportunistic behavior to mitigate biased subordinate information. Therefore, we argue that the CC attaches direct consequences for the supervisors to this opportunistic behavior.[6] In particular, we expect that the CC decreases the supervisor's own annual performance rating to send a clear and especially timely signal to the supervisor to direct her efforts in the next performance period towards the supervision and development of her subordinates instead of manipulating subordinate performance indicators. Since firms predominantly link annual compensation to performance ratings (e.g., Merchant and Van der Stede 2017), a decrease in these ratings directly impacts the supervisor through lower compensation in a timely manner. In addition, performance ratings typically play a role in promotion decisions and opportunistic behavior thus further indirectly jeopardizes the supervisor's long-term career opportunities.

Given the inherent difficulty to detect opportunism, the CC needs to resort to identifying signals that make it more likely that opportunism is the core root of the observed evaluation behavior. To draw conclusions about the likelihood that opportunistic behavior has occurred, the CC will combine signals of opportunistic behavior with further information and situations that the CC members experienced with the respective supervisor or her subordinates throughout the last year.[7] Subsequently incorporating the identified signals of opportunistic behavior in the

[6] One might argue that one option the CC has to mitigate opportunistic bias is to simply take away the incentive for opportunism by committing to not incorporate (average) subordinate performance in the evaluation of supervisors. While the CC can *ex-ante* commit to not take into account subordinate performance in evaluating the supervisor, this commitment is *ex-post* not efficient. The reason why the commitment is *ex-post* not efficient is that, if supervisors believe the commitment, they have no incentive to opportunistically bias the ratings proposed by them, which would make these ratings highly useful in evaluating the supervisor. Because it is efficient to use the ratings *ex-post*, the commitment is not credible.

[7] Given the incentive to inflate ratings and the inherent difficulty of detecting bias, some but not all supervisors will inflate ratings as the probability of detection is not one. This leads to a setting where many ratings are accurate, some are biased and will be corrected during the calibration sessions, but some still remain biased as not all biases are uncovered by the CC.

supervisor's own annual performance rating is an effective mechanism to discipline undesirable behavior as it corresponds to the supervisor's myopic attempt to minimize supervision efforts, and maximize personal short-term benefits at the expense of the effectiveness of the firm's annual performance reviews. As a result, when concluding that the supervisor exhibits opportunistic evaluation behavior, we expect that the CC mitigates opportunistic bias by directly disciplining the supervisor with a decrease in her performance rating, which in turn creates implicit incentives to refrain from such behavior. Hypothesis 1a summarizes this expectation.

H1a: Signals of opportunistic rating behavior have a negative impact on supervisors' performance ratings.

We further consider the effect of opportunistic behavior on the supervisor's promotion probability. The promotion rule for sorting purposes makes promotion more likely when employees have developed the necessary skills for the next job (Grabner and Moers 2013).[8] If opportunistic behavior is considered as a signal of an employee skill or type that the firm does not value, signals of opportunistic behavior will decrease a supervisor's probability of promotion, and thus also dis-incentivize such behavior.

Regarding this disciplining effect, however, promotions might not be sufficiently effective. That is, promotion opportunities are typically not able to provide persistent effort incentives as the available opportunities vary between years (e.g., Baker, Jensen, and Murphy 1988), and therefore are less suited to dis-incentivize undesired behaviors in a timely and consistent manner. Given that opportunistic behavior needs to be disciplined as soon as it is detected to prevent the next performance period to be impacted by any ill-motivated effort, performance ratings seem more suitable to discipline opportunistic behavior than promotion opportunities.[9] We thus expect that

[8] The term *promotion rule* captures the implicit expectation upon which promotion decisions are based.
[9] To be complete, signals of opportunistic behavior do affect the likelihood of promotions *indirectly*. Given that signals of opportunistic behavior are expected to affect performance ratings (H1a) and performance ratings typically impact

promotion opportunities are less important for disciplining opportunistic behavior than performance ratings. We summarize this expectation in the following hypothesis.

H1b: Signals of opportunistic rating behavior are less associated with supervisors' probability of promotion than with supervisors' performance ratings.

Incentivizing investments in the development of supervisory skills (H2a and H2b). The inability to provide accurate feedback and ratings, which is typically reflected in compressed performance ratings (e.g., Künneke 2017), can be mitigated by investing in supervisory skills. These investments are unobservable and costly, and thus need to be incentivized (Mohrman and Lawler 1983; Murphy 1992). Previous theoretical and empirical research shows that promotion opportunities are an effective mechanism to induce human capital acquisition, especially when the associated skills are more important for the next job (Grabner and Moers 2019; Prendergast 1993). The promotion rule for sorting purposes makes promotion more likely when the expected productivity of the acquired skills is higher in the next job (Grabner and Moers 2013). This promotion rule thus provides incentives to invest in skills that are (even more) relevant in the next position. Promotions are therefore an adequate tool for providing incentives to supervisors to invest in their supervisory skills (DeVaro and Gürtler 2015; Grabner and Moers 2019; Prendergast 1993). Supervisory skills not only include the willingness to differentiate among subordinate performance levels, but also the necessary capability to inform and support the development of subordinates when performing unsatisfactorily, and dealing with potential confrontations due to negative feedback.

What makes promotions especially adequate for incentivizing such supervisory skills is that these skills are even more crucial when the supervisor advances in the corporate hierarchy, since this typically requires the supervision of higher-level subordinates who hold more far-reaching

the likelihood of promotions, signals of opportunistic behavior affect the promotion decisions through their effect on performance ratings. We return to this indirect effect when discussing the empirical results.

responsibilities. Therefore, especially at higher levels, the firm wants to ensure to promote supervisors with the necessary skills to adequately evaluate and develop their subordinates, hence minimizing the costs associated with performance evaluation bias (Mumford, Campion, and Morgeson 2007). Thus, as soon as the firm observes signals of outstanding supervisory skills, it is in the firm's own interest to promote these supervisors to utilize the expected increase in the productivity of their supervisory skills. Hence, to ensure that supervisors have the appropriate skills that mitigate performance evaluation bias, we argue that, in making promotion decisions, the CC will directly take into account the supervisor's ability to sufficiently differentiate among subordinate performance levels.[10] This promotion rule provides supervisors with the implicit incentives to acquire the necessary skills (cf. Grabner and Moers 2019), and therefore we hypothesize:[11]

H2a: Signals of supervisors' skills to discriminate among subordinate performance levels increase their probability of promotion.

While the literature shows that promotion-related mechanisms are effective mechanisms for inducing skill acquisition (Grabner and Moers 2019; Kahn and Huberman 1988; Prendergast 1993), we additionally examine the impact of signals of supervisory skills on performance ratings. A supervisor is usually held accountable for the performance of her subordinates, and subordinate

[10] It is highly unlikely that all subordinates assigned to one supervisor will initially perform (equally) well and a high-skilled supervisor will reveal that in her proposed ratings (rating discrimination). While a high-skilled supervisor improves the development of her subordinates, the full performance improvement is, however, not instantaneous but develops gradually over time. In the case supervisors would only keep the same subordinates, then one can expect that for high-skilled supervisors, over time the performance improvements will cause less dispersed ratings. However, due to the typical high inflow and outflow of employees and hence subordinates in professional service firms, the composition of the subordinate team per supervisor is highly volatile. Thus, it can be expected that high-skilled supervisors have continuously more dispersed ratings compared to low-skilled supervisors.

[11] For completeness, there is heterogeneity in responses to the skill development incentives triggered by promotion opportunities. While for some supervisors the increased likelihood of promotion is economically relevant, for others this incentive will not be sufficient to invest in their supervisory skills, especially given the costs they have to incur. The latter group of people, however, can be expected to be naturally "sorted out" in an up-or-out setting, and eventually leave the firm.

performance is not only affected by the effort the supervisor spends on the subordinates, but also the interaction of this effort with her existing supervisory skills. As such, supervisory skills already indirectly affect the supervisor's own performance rating to the extent that such skills have been effectively used in supporting the development of her subordinates. Whether, over and above this indirect effect, skill development can be directly incentivized with an increase in the supervisor's annual performance rating depends on the annual rating's suitability for that purpose. In general, efficient incentives imply that certain behaviors can only be incentivized to the extent that these behaviors are productive. This specifically implies here that performance ratings can only efficiently incentivize skill development to the extent that supervisory skills are productive in the current job, i.e., to the extent that an investment in supervisory skills is beneficial to the supervisor's current performance. While supervisory skills are important at the current level, these skills are more important at the next level because of the supervisor's impact on potentially more, but also higher-level subordinates. That is, the marginal productivity of supervisory skills is lower at the current job level compared to the next job level. Given that behaviors can only be efficiently incentivized to the extent that these behaviors are productive, the incentives for developing supervisory skills via performance ratings is limited. While this does not necessarily imply that performance ratings are not useful for incentivizing the development of supervisory skills, it does imply that they are less useful than promotions since the latter mechanism focuses on the marginal productivity at the next position. We summarize this expectation in the following hypothesis:

H2b: Signals of supervisors' skills to discriminate among subordinate performance levels are less associated with supervisors' performance ratings than with supervisors' probability of promotion.

III. RESEARCH SETTING, SAMPLE, AND MEASURES

Description of our research site

Our research site is a member firm of an international professional service provider specialized in the fields of assurance, tax, strategy consulting, and corporate finance. The member firms are organized as separate legal entities in a network operating in over 150 countries worldwide, and we received our data from one of their largest member firms. The subsequent description of our research site is based on a set of interviews with important stakeholders in the firm, as well as internal documentation provided by the company.

Supervision process. Each newly hired employee is assigned to a supervisor who is responsible to help the subordinates accomplish their performance and development goals.[12] The supervisor is an experienced professional and typically at least two ranks higher in the hierarchy. Typically, in the beginning of the supervisor-subordinate relationship, the supervisor seeks information about the subordinate's general profile such as existing knowledge and skills, as well as work-related interests. Based on the employee's profile, the rank-specific general development plan for the next year is complemented by personalized goals that are developed together with the supervisor. Throughout the year, the supervisor arranges in coordination with the central resource planning office the regular staffing of the subordinate, i.e., the assignment to projects fitting the subordinate's profile. This assignment is largely driven by capacity constraints, which implies that subordinates are typically assigned to projects where there is a demand for people. For each project, the employee receives formal feedback about the personal and professional project performance. In particular, the project leader rates each member of the project team on a number of pre-

[12] The matching of the supervisor and subordinate is executed by the HRM department, and based on availability, industry interest and to some extent educational background, but independent of any other criteria. That is, the matching of supervisor and subordinate is to a large extent random, and hence is unlikely to be correlated with potential determinants of supervisor performance.

established dimensions, resulting in a project rating.[13] The project ratings serve as a first indicator helping the supervisor to determine the employee's performance level.

Annual performance evaluation process. At the end of the year, a formal performance review takes place, requiring several steps: prior to having an individual performance evaluation meeting with the subordinate, the supervisor gathers information on the subordinate's performance, including project ratings and informal feedback from project leaders. During the meeting with the subordinate, the subordinate's performance is assessed based on the achievement of the assigned and self-established goals, as well as the personal and professional development during the year. The supervisor subsumes all information with her personal impression about the subordinate's capabilities into an initial performance rating. While it is highly recommended that the supervisor formally prepares this initial rating and communicates it to the CC prior to the meeting with the committee, it is not mandatory. As soon as all official annual performance reviews between the supervisors and subordinates are completed, all supervisors and (HR) partners gather to finalize the performance evaluation process in CC meetings, typically organized per office. The CC starts with the assessment of non-management employees. Typically, the supervisors present each subordinate to substantiate their initial rating. Based on the subjective evaluation of the supervisor, the CC discusses the performance of the subordinate, also taking into account other available performance data such as information on overtime and chargeable hours, as well project ratings, some of which are typically provided by other members of the committee. Furthermore, the experience of other CC members who worked on one or more projects with the employee in question are shared, and any sort of performance inconsistencies are clarified.

[13] Note that the supervisor might be the project leader, but that is not necessarily the case given the process of assignment and because employees typically work on multiple projects with different project leaders. Our interview respondents indicated that in roughly half of the projects, the supervisor is not the project leader. Unfortunately, we cannot obtain information on which projects the supervisor also served as project leader.

After deliberation, the CC agrees on a final rating for each employee that reflects the employee's performance relative to his peers.[14,15] The final rating has direct consequences for salary increases and bonus allocation. Subsequent to the determination of the final ratings, the promotion candidates at each rank are compared and selected based on their current performance (final rating), but also with regard to competencies that are relevant in the next position. Once the non-management employees have been evaluated, their supervisors are evaluated next, and thus leave the meeting. This procedure ensures that the subordinates' final rating is always determined prior to the final rating of the respective supervisor. It further allows the remaining evaluators to take the supervisors' rating behavior into account when assessing the supervisors' own final rating and promotion. Special attention is paid to supervisory skills when making promotion decisions.

Sample, measures, and empirical models

As we analyze decisions on the final rating and promotions of supervisors, our samples are restricted to employees having supervisor responsibilities. The final sample consists of 1,796 supervisor-year observations, covering 952 unique supervisors in the period from 2010 to 2012.[16]

Our main dependent variables represent the supervisors' own final performance evaluation outcomes as determined by the CC. *RATING* reflects the final rating the supervisor receives from the CC and ranges from 1 to 5. In addition, we define a promotion as the advancement to the next hierarchical level, as defined by the internal job rating system. The indicator *PROM* equals 1 in the year the promotion decision takes place, 0 otherwise.

[14] The rating scale for all performance-related information including project, initial, and final rating ranges from 1 to 5.

[15] The firm's documented rating guideline states the use of a forced rating curve when evaluating employees. However, several interview partners confirm that, in practice, this guideline is not enforced, but rather used as an advisory distribution. Interview partners confirm the widely held belief that a forced curve can lead to unfairness as employees might not receive the rating they deserve.

[16] As is typical for professional service firms, our research site is organized in organizational sub-units based on global geographic regions. We obtained data from one of the largest regional organizational sub-units. Given that the performance evaluation process is standardized throughout the whole organization and even comparable to competitors, our results are representative of the entire company/industry.

To test H1a and H1b, we aim to capture two different aspects of rating behavior that signal to the CC that it is more likely that opportunism is the core root of the observed evaluation behavior. The first measure is an indicator that equals 1 when the CC's average adjustment of the initially proposed ratings of one supervisor (i.e., the average of all adjustments including upward, downward, and no adjustments made by the CC) is more negative than the average adjustments compared to all other supervisors per year, 0 otherwise (*CC_DOWNGRADE*).[17] As alternative, we use a refined categorization in which we classify each supervisor based on the CC's reaction to the supervisor's reported initial ratings. For each supervisor, the CC has four options, which we capture with a respective indicator variable: (1) no adjustments to any of the initial ratings proposed by the supervisor (used as base category), (2) at the same time upgrade some initial ratings and downgrade others (*RATING_ADJ*), (3) only downgrade one or more initial ratings (*RATING_DOWNGRADE*), and (4) only upgrade one or more initial rating (*RATING_UPGRADE*).

For the second measure, we define a variable intended to capture the strategic provision of information to the CC by the supervisor. By gathering and aggregating relevant performance-information, the supervisor forms her evaluation about subordinate performance, resulting in the creation of the initial rating for her subordinates. However, when a supervisor prefers to first reflect more thoroughly with the CC members because, for example, the supervisor is in doubt about certain information, the initial rating is not necessarily formally determined before the meeting. While this flexibility allows supervisors to first align with the CC members whenever it is needed, supervisors can also strategically exploit this opportunity and single out favorable information about well-performing subordinates, i.e., provide initial ratings only for high performers, and

[17] It is important to note that the average adjustment the CC makes, i.e., the difference between the final ratings and the initial ratings, is negative. This implies that the CC downgrades more than it upgrades, and thereby already corrects lenient ratings provided by the supervisors. *CC_DOWNGRADE* thus measures whether the CC downgrades a supervisor even further than the average adjustment.

disguise the underperformers. Such tactics of strategic information provision have been particularly found to be present in situations where the accountability of the supervisor is high, which is the case in our setting (Fandt and Ferris 1990). Therefore, our second measure captures supervisors‘ systematic, incomplete reporting behavior in favor of top performing subordinates, which potentially aims at opportunistically enhancing the committee's impression of the supervisor. *PRESENT_BEST* is an indicator variable that equals 1 if, prior to the CC meeting, the supervisor decides to formally create the initial ratings only for those subordinates who are among the best according to existing performance data, 0 otherwise. We consider a subordinate among the best performing subordinates based on performance information that is mostly provided by sources other than the supervisor, i.e., project ratings; if the subordinate belongs to the top quartile in terms of average project ratings within one service line, then the subordinate is labeled as among the best performing subordinates.

To test H2a and H2b, we aim to capture supervisory skills, i.e., the ability to handle the potential costs as result of the provision of honest performance feedback and consequently the discrimination among subordinate performance levels. *SUBORD_DISCR* reflects the discrimination among subordinate performance levels and, following Künneke (2017), is measured as the standard deviation of the initial ratings a supervisor provides to the subordinates in one year. In line with Künneke (2017), we argue that it is possible to derive valuable information about supervisory skills when (not) observing compression. When a supervisor decides to give accurate, and hence unbiased ratings, she signals that she is able to manage the potential costs, and hence is considered as having high supervisory skills. Given the general tendency towards compressed performance ratings, such supervisory skills are reflected in less compressed ratings. Next to the empirical validation of this measure for supervisory skills (Künneke 2017), the assumption that the variance in ratings in part reflects supervisory skills is consistent with the observation that firms

typically interpret compression as a signal of poor supervisory skills (e.g., Bingham and Beer 2012; Hall and Madigan 2000).

To examine the relative importance of the various signals of supervisors' rating behavior for performance ratings and promotions, we estimate the following two models using ordinary least squares and probit regressions, respectively:

$$RATING_{jt} = \beta_0 + \beta_1 CC\_DOWNGRADE_{jt} + \beta_2 PRESENT\_BEST_{jt} + \beta_3 SUBORD\_DISCR_{jt} + \beta_4 SUBORD\_RATING_{jt} + \beta_5 PROJ\_RATING_{jt} + \beta_6 EFF\_UTIL_{jt} + \beta_7 DIRECT\_MARGIN_{jt} + \beta_8 OVERTIME_{jt} + \beta_9 RATING_{jt-1} + \beta_{10} SPAN\_CONT_{jt} + \beta_{11} JOBTENURE_{jt} + \beta_{12} GENDER_j + Fixed\ Effects + \varepsilon_{jt}$$

$$PROMOTION_{jt} = \beta_0 + \beta_1 CC\_DOWNGRADE_{jt} + \beta_2 PRESENT\_BEST_{jt} + \beta_3 SUBORD\_DISCR_{jt} + \beta_4 RATING_{jt} + \beta_5 SUBORD\_RATING_{jt} + \beta_6 PROJ\_RATING_{jt} + \beta_7 EFF\_UTIL_{jt} + \beta_8 DIRECT\_MARGIN_{jt} + \beta_9 OVERTIME_{jt} + \beta_{10} RATING_{jt-1} + \beta_{11} SPAN\_CONT_{jt} + \beta_{12} JOBTENURE_{jt} + \beta_{13} GENDER_j + Fixed\ Effects + \varepsilon_{jt}$$

where *j* relates to the supervisor and *t* to the year. Next to our main independent variables, we capture a set of supervisor-specific characteristics that supposedly influence the performance evaluation outcomes as control variables. The premise underlying H1a is that the average performance of the subordinates is positively associated with the supervisor's performance, which provides the supervisor with incentives to inflate (some of) her subordinates' initial ratings. To test this premise, we include the average performance of the subordinates per supervisor (*SUBORD_RATING).* Further, we measure the supervisor's current performance in terms of average project ratings (*PROJ_RATING*), net engagement revenues (*DIRECT_MARGIN*), overtime measured as hours worked above the specified contract hours (*OVERTIME*), as well as past performance ($RATING_{t-1}$), all of which we expect to be positively related to the supervisor's performance ratings, and potentially promotions. We additionally include the current final rating

($RATING_t$) in the promotion model to account for the potential relevance of current performance in promotion decisions. We measure the supervisor's effective utilization (*EFF_UTIL*), that is, the time directly spent on ongoing engagements, as chargeable hours divided by available hours.[18] We measure the supervisor's span of control (*SPAN_CONT*) by the number of subordinates to account for the possibility that the CC takes a supervisor's varying contribution in terms of supervision time and effort into account when determining the performance outcomes of the supervisor. We further measure the supervisor's tenure in the current job (rank) in years (*JOBTENURE*). Job tenure can be positively related to performance evaluation outcomes due to increases in effective ability over time (Gibbons and Waldman 1999) or negatively due to talented employees not remaining long in the same rank as a result of the up-or-out system (Lazear 2004). We therefore have no directional prediction. Finally, *GENDER* is an indicator variable that equals 1 (0) if the supervisor is a male (female). We consider *GENDER* to be a relevant predictor of rating and promotion decision outcomes, but as prior research provides mixed results (Rosen and Jerdee 1974; Tsui and Gutek 1984), we make no directional statement. Further, we control for rank-fixed effects to control for differences in performance levels and promotion probabilities across the hierarchy. Finally, we control for year-fixed effects.

Descriptive statistics

Table 1 reports descriptive statistics for the sample to test our predictions. We observe a promotion rate of 22 percent, which is in line with the company's structured career development plan. The supervisor's final ratings cover the entire rating range from 1 to 5, with a median of 4. In 36 percent of the supervisor-years, the supervisor receives at least one project rating per year, with an average of 1.6 project ratings per year. The project evaluations received have a rating of 4.18 on average. The effective utilization is around 78%, which can be explained by the fact that as of manager rank

[18] *EFF_UTIL*, *DIRECT_MARGIN*, and *OVERTIME* are winsorized at the 1% level.

the time spent on other tasks than chargeable hours increases. Supervisors typically support 3 subordinates per year.

--------- Insert Table 1 ----------

Table 1 further shows that downgrades of initial performance ratings provided by the supervisors regarding their subordinates is way more common than upgrades, which is consistent with a tendency towards upward-bias. In 23% of the supervisor-years, at least one of the initial ratings provided by the supervisor is downgraded without any ratings being upgraded. In contrast, in only 4% of the supervisor-years, at least one of the initial ratings is upgraded without any ratings being downgraded. To dig deeper into the link between the supervisors' evaluation behavior and the calibration process, we provide detailed statistics on the initial performance ratings provided by the supervisors. The average initial performance rating at the subordinate-year level is 3.67. We divide the sample into subordinate-year observations where the initial rating provided by the supervisor was either downgraded, upgraded, or remained unchanged during the calibration process. Consistent with expectations, the average initial rating provided by the supervisors is significantly higher for those subordinates whose initial ratings were downgraded in the calibration process (4.05) as compared to an upgrade (3.45) or no change (3.60). We further conduct a similar analysis at the supervisor-year level. Supervisors for whom at least one initially provided performance rating of their subordinates was downgraded, while none was upgraded, show a higher average initial performance rating (3.86) as compared to the remaining supervisors (3.59). This analysis provides first evidence consistent with the CC lowering lenient performance ratings.

Table 2 reports the Pearson correlations between the independent variables. While none of the correlations cause multicollinearity concerns, a few correlations are noteworthy. First, *SUBORD_DISCR* is not correlated with sub-dimensions of performance in the current job such as *DIRECT_MARGIN* and *PROJ_RATING*, suggesting that our measure of supervisory skills indeed

captures a different underlying dimension. Further, *SUBORD_DISCR* is significantly negatively associated with effective utilization (*EFF_UTIL*). This negative association is in line with Westermann et al. (2015), who find that one of the factors that inhibit the provision of proper supervisory feedback is other competing demands on a supervisor's time. The negative correlation reflects these competing demands and suggests that the more supervisors are staffed on ongoing projects during the year, the less time they have to invest in their supervisory skills.

--------- Insert Table 2 ----------

IV. DATA ANALYSIS AND RESULTS

Consequences of opportunistic rating behavior (H1a and H1b)

Table 3 reports the results for H1a. As reported in Column I of Table 3, when the CC identifies the need to correct overly inflated ratings, i.e., substantially downgrades initial ratings that are suggested by the supervisor (*CC_DOWNGRADE*), the supervisor's own final rating is lowered as well ($p<0.01$, two-tailed). In addition, if a supervisor presents a more detailed assessment of only her best subordinates during the CC meeting (*PRESENT_BEST*), we observe an additional lowering of the supervisor's performance rating ($p<0.05$, two-tailed). This finding implies that opportunistic behavior, both in terms of directly inflating subordinates' ratings and impression management, is disciplined by the CC and comes at a direct cost for the supervisor. In sum, our findings indicate that the CC disciplines supervisors via lower performance ratings when they are assumed to provide opportunistically biased information, providing support for H1a. To provide insights into the economic magnitude, we express the economic significance in terms of the probability of receiving a particular performance rating and how this is affected by *CC_DOWNGRADE* and *PRESENT_BEST*. We find that the probability of receiving a rating of 3 or less increases by 13% when observing significant downgrades and increases by 37% when the

supervisor presents a more detailed assessment of only her best subordinates. This suggests that the effects of *CC_DOWNGRADE* and *PRESENT_BEST* are economically significant.[19]

Column II of Table 3 reports the results when using our alternative measure for *CC_DOWNGRADE*. We find that when the CC identifies the need to correct inflated ratings, i.e., downgrades one or more initial ratings that are suggested by the supervisor while upgrading none (*RATING_DOWNGRADE*), the supervisor's final rating is lowered ($p<0.05$, two-tailed). We do not find this effect for *RATING_UPGRADE* or *RATING_ADJ*. That is, if one or more ratings provided by the supervisor are (also) adjusted upwards, there is no additional effect on the supervisor's performance rating. Importantly, the insignificant coefficient of *RATING_UPGRADE* confirms our expectation that the CC does not attribute low ratings provided by the supervisor to any kind of strategic behavior as there is no economic incentive to decrease subordinate ratings on purpose. Most importantly, these results confirm that indeed opportunistic rating behavior is disciplined. Overall, our results provide strong support for H1a.

Table 4, Column I reports the results regarding H1b, which predicts that signals of opportunistic rating behavior are less associated with supervisors' probability of promotion than with supervisors' performance ratings. We find that signals of opportunistic behavior, captured by *CC_DOWNGRADE* and *PRESENT_BEST,* are not associated with the probability of promotion, which together with the findings in Table 3 is consistent with H1b. To formally test H1b, we compare the marginal effects of *CC_DOWNGRADE* and *PRESENT_BEST* between the probability of receiving a promotion and the probability of receiving a performance rating greater than 3. When

[19] The results also show that *SUBORD_RATING* has a highly significant and positive effect on the supervisor's final rating ($p<0.01$, two-tailed), after controlling for different dimensions of the supervisor's current performance (*PROJ_RATING*, *DIRECT_MARGIN*, *OVERTIME*) and overall past performance ($RATING_{t-1}$), all of which also positively affect the supervisor's final rating. This finding is consistent with the assumption that subordinate performance reflects to some extent the actual supervision provided, for which the firm rewards the supervisor, thereby providing the supervisor with incentives to inflate average subordinate performance.

comparing the marginal effects between the first model using *RATING* as dependent variable and the second model using *PROMOTION* as dependent variable, we find that the marginal effect of *CC_DOWNGRADE* is, at the margin, significantly different between the two models (p=0.119, two-tailed) and the marginal effect of *PRESENT_BEST* is significantly different between the two models (p<0.05, two-tailed). A test of the joint significance of these differences is also statistically significant (p<0.05, two-tailed). These results provide support for H1b. Note that the probability of promotion is positively associated with *RATING*, which confirms the expectation that opportunistic rating behavior has an indirect effect on promotion decisions. A formal test of this indirect effect (Sobel 1982) shows that the indirect effect is statistically significant for both *CC_DOWNGRADE* (p<0.01, two-tailed) and *PRESENT_BEST* (p<0.05, two-tailed).[20]

--------- Insert Table 3 ----------

Consequences of supervisory skills (H2a and H2b)

The results for H2a are reported in Table 4, Column I. We find that supervisors who sufficiently discriminate among performance levels of their subordinates *(SUBORD_DISCR)*, and thus are also less lenient, have a significantly higher probability of promotion (p<0.05, two-tailed). Note that such discrimination cannot be simply mimicked by a supervisor who does not have the skills to provide negative feedback because it requires using the entire rating scale, and especially a low rating implies costs and further consequences that such a supervisor tries to avoid. In terms of economic significance, the results imply that moving from the first quartile of *SUBORD_DISCR* to the third quartile of *SUBORD_DISCR*, i.e., comparing a low-skilled supervisor to a high-skilled supervisor, increases the probability of being promoted by approximately 28%.

[20] The significance levels are based on confidence intervals of bootstrap OLS estimates of the indirect effects, using random sampling with replacement and 5,000 bootstrap samples (Preacher and Hayes 2004).

In addition, we repeat this analysis within a sample of supervisors who are responsible for at least 3 subordinates, thus eliminating supervisors for which the lack of discrimination is a less clear signal of a lack of supervisory skills. The results reported in Column II of Table 4 confirm our expectation that rating discrimination is an important determinant of promotion decisions. The fact that the results are even stronger in this subsample is consistent with the assumption that the ability to identify the performance levels of subordinates should become more salient, the higher the number of subordinates a supervisor takes care of. To corroborate this assumption, we include an interaction term between *SUBORD_DISCR* and *SPAN_CONTR*. The results for this additional analysis reported in Column III of Table 4 show that the interaction is positive and significant ($p<0.01$, two-tailed). This suggests that the CC realizes that, with an increasing number of subordinates, the likelihood of having the same performance level decreases, and thus considers a differentiation as inevitable, which needs to be reflected in the provided initial ratings.[21] These findings support our conclusion that the firm indeed takes the ability to discriminate amongst subordinate performance levels into account when making promotion decisions. In sum, we provide strong support for H2a.

Finally, our results for H2b, as reported in Table 3, show that supervisory skills in the form of subordinate discrimination *(SUBORD_DISCR)* are not significantly associated with the supervisor's final rating in any of the two models. H2b predicts that supervisory skills are less associated with supervisors' performance ratings than with supervisors' probability of promotion. Our finding that supervisory skills in the form of subordinate discrimination are strongly associated with promotions while it is not associated with performance ratings is in line with H2b. To formally test H2b, we compare the marginal effect of *SUBORD_DISCR* between the probability of receiving

[21] This also confirms that, given the lower likelihood of all subordinates having equal performance, the supervisor's discrimination level is interpreted as an even stronger signal for supervisory skills, and hence receives more weight in the promotion decision.

a promotion and the probability of receiving a performance rating greater than 3. When comparing the marginal effect between the first model using *RATING* as dependent variable and the second model using *PROMOTION* as dependent variable, we find that the marginal effect of *SUBORD_DISCR* is significantly different between the two models ($p<0.10$, two-tailed). These results provide support for H2b and confirm a lesser role of performance ratings in incentivizing the development of supervisory skills as compared to promotions.

--------- Insert Table 4 ----------

Robustness tests

Alternative test of hypotheses. Given that the decisions regarding the rating and the promotion are taken during the same meeting, we also use an alternative estimation approach. In particular, we use seemingly unrelated regression (SUR) equations to reflect the simultaneity inherent in these decisions. To have different sets of explanatory variables for these equations, we solely focus on H1a and H2a and jointly estimate the following models:[22]

$$RATING_{jt} = \beta_0 + \beta_1 CC\_DOWNGRADE_{jt} + \beta_2 PRESENT\_BEST_{jt} + \beta_3 SUBORD\_DISCR_{jt} + \beta_4 PROJ\_RATING_{jt} + \beta_5 EFF\_UTIL_{jt} + \beta_6 DIRECT\_MARGIN_{jt} + \beta_7 OVERTIME_{jt} + \beta_8 RATING_{jt-1} + \beta_9 SPAN\_CONT_{jt} + \beta_{10} JOBTENURE_{jt} + \beta_{11} GENDER_j + Fixed\ Effects + \varepsilon_{jt}$$

$$PROMOTION_{jt} = \beta_0 + \beta_1 SUBORD\_DISCR_{jt} + \beta_2 SUBORD\_RATING_{jt} + \beta_3 RATING_{jt} + \beta_4 PROJ\_RATING_{jt} + \beta_5 EFF\_UTIL_{jt} + \beta_6 DIRECT\_MARGIN_{jt} + \beta_7 OVERTIME_{jt} + \beta_8 RATING_{jt-1} + \beta_9 SPAN\_CONT_{jt} + \beta_{10} JOBTENURE_{jt} + \beta_{11} GENDER_j + Fixed\ Effects + \varepsilon_{jt}$$

where *j* relates to the supervisor and *t* to the year. Table 5 reports the results of the SUR using *CC_DOWNGRADE*, while Table 6 reports the results using the alternative measures for

[22] Note that if the same set of explanatory variables are used, SUR is identical to equation-by-equation OLS. Such an analysis would therefore not provide any information over and above those reported in Table 3 and 4.

CC_DOWNGRADE. In both tables, Column I (II) relates to the supervisor's performance rating (promotion). These analyses confirm that our results for H1a and H2a are robust when controlling for the correlation between the error terms of the two equations.[23]

--------- Insert Table 5 and 6 ----------

Alternative measures for subordinate discrimination. We apply alternative measures for the extent to which the supervisor discriminates among subordinates and define the following additional measures to capture supervisory skills: *HIGH_DISCR* (indicator variable that equals 1 if a supervisor's discrimination is larger than the median discrimination of all supervisors within a year, 0 otherwise), *ADJ_DISCR* (supervisor's discrimination adjusted by the median discrimination of all supervisors within a year), *RANGE_SV* (the difference between a supervisor's highest and lowest initial rating provided to subordinates), *HIGH_RANGE* (indicator variable that equals 1 if a supervisor's initial rating range is larger than the median rating range of all supervisors within a year, 0 otherwise), and *BELOW3* (indicator variable that equals 1 if a supervisor provides an initial rating that is below 3, and 0 otherwise). In addition, we use a factor analysis to estimate the underlying latent construct, using these four variables plus *SUBORD_DISCR.* Principal component analysis reveals one factor with an eigenvalue greater than 1, explaining approximately 91% of the total variance. The analyses show that all alternative measures for supervisory skills have a positive and significant association with the promotion likelihood of a supervisor.[24] None of these alternative measures have an effect on the final rating.

[23] For completeness, if we switch the key explanatory variables across models, none of the variables related to opportunism or supervisory skills are significant, in line with our results in Table 3 and 4 related to H1b and H2b.
[24] *HIGH_DISCR*: β=0.332, $p<0.01$; *ADJ_DISCR*: β=0.435, $p<0.05$; *RANGE_SV*: β=0.277, $p<0.01$; *HIGH_RANGE*: β=0.671, $p<0.01$; *BELOW3*: β=0.613, $p<0.05$; and *FACTOR*: β=0.144, $p<0.01$.

V. DISCUSSION OF THE RESULTS

In line with our theory, the analysis of the firm's archival performance evaluation and personnel data confirms that supervisors' documented evaluation behavior is systematically related to their performance evaluation outcomes. We provide evidence that the CC considers skills such as the development of subordinates, which includes the provision of reliable ratings and corresponding feedback, as necessary to climb the corporate hierarchy. Further, supervisors are rewarded for the effort they exert in increasing their subordinates' performance. In contrast, supervisors who presumably deliver biased information and reveal evaluation behavior that is undesirable for the firm, are likely to face direct negative compensation consequences as well as indirect negative career consequences. To validate the extent to which these findings are recognized by the management of the company, we conducted interviews with partners involved in the performance evaluation process at our research site, as well as another member firm in the professional network.[25]

First, our interview partners stress that the CC emphasizes the accuracy of the initial rating as they rely to a great extent on the supervisor to form a final performance rating. Given the information asymmetry between the CC and lower-ranked employees, the information signals provided by the supervisor have to be as accurate and unbiased as possible to make informed decisions regarding the final rating as well as promotions. One partner notes that one of the main reasons for calibration sessions to exist is the fact that supervisors tend to overrate their subordinates, which has to be corrected. This not only confirms that biased ratings are a significant concern but also that the CC is used as a mechanism to identify and mitigate the biases that supervisors introduce.

[25] In particular, we conducted two interviews with the same partner at our research site, as well as interviews with two partners at another member firm in the professional network.

Second, the interviewees indicated that managers and senior managers are evaluated based on how much time they spend on supervision and whether they focus on developing the subordinates' skills or just gather a group around them to perform their work. In all this, the quality of feedback by the supervisor is crucial for good supervision and both interviewees clearly confirmed the importance of rating discrimination as a relevant signal of supervisory skills. As one partner put it when referring to the lack of discrimination as a signal of poor supervision: "*it is striking when a supervisor only has good subordinates…this is where subjectivity starts to play a negative role*." The interviewees also stated that the promotion criteria for (senior) managers focus on how well they perform on projects as well as how good they are as a supervisor, which confirms that rating discrimination as a signal of supervisory skills is considered in promotion decisions.

In addition, while the partners confirmed that signals of opportunistic behavior, such as impression management, are recognized by the CC, they suggest different strategies of how this is typically dealt with. In line with our predictions, the partner from our research site confirms that such behavior is immediately disciplined through the performance rating process. This disciplining seems to be effective, given that we observe that supervisors whose initial ratings were substantially downgraded, are approximately 50% less likely to be downgraded again in the following year. The partner mentions that while it is understandable that supervisors do not want to be associated with underperforming subordinates for their own sake, this behavior is unacceptable in terms of rating accuracy and fairness concerns. One of the partners in the member firm suggests a more implicit way of dealing with this. In particular, supervisors exhibiting this kind of behavior are confronted with it in a private conversation, either during a break or immediately after the CC meeting. This highlights the importance of addressing this behavior in a timely manner. Repeated offenders will also "feel" it directly in their ratings. The interviews not only clearly confirm that the CC is aware of the occurrence of opportunistic rating behavior, they

further confirm that – if detected – these signals of opportunistic behavior can impact the performance rating of the supervisor, resulting in the pattern we document. Overall, our interviews provide support for our theory that supervisors' evaluation behavior is monitored by the CC and taken into account in their performance ratings and promotion decisions.

VI. CONCLUSION

Research on subjective performance evaluations has identified the most important drawback of its use: the fact that these evaluations can be distorted by the evaluator, and therefore not always reflect the employee's "true" performance. Bias in performance evaluations is problematic because it comes at high direct and indirect cost for the firm, especially when these evaluations are used in compensation and personnel decisions (Moers 2005). Firms therefore face the problem that the information content of subordinates' performance evaluations might be misleading because supervisors lack the skills to provide adequate ratings or because they intentionally inflate them.

In this study, we show that the firm is aware of this problem, and tries to mitigate it by holding supervisors accountable for their evaluation behavior. In particular, CCs make use of the annual performance evaluation process, and its effect on subsequent compensation to dis-incentivize opportunistic rating behavior. In particular, the CC disciplines such behavior by downgrading the performance ratings of supervisors identified as providers of opportunistically biased information.

Further, consistent with the literature on human capital acquisition (e.g., Grabner and Moers 2019; Prendergast 1993) we find that promotion opportunities are used to incentivize investments in supervisory skills by making promotions more likely for supervisors that have acquired these skills. The relevance of this type of skill, in addition to being a driver of subordinate development (Künneke 2017), is highlighted in another study by Ahn, Hwang, and Kim (2010). They provide evidence that the perceived degree of discriminability of performance measures motivates employees to exert more effort and improve their performance, which is one of the main goals of

a firm's evaluation system. Although prior literature suggests that lenient ratings can have motivational effects on employee performance (Bol 2011), our findings indicate the importance of discriminability and the CC thus incentivizes supervisors to reduce rating errors. This supports the firm in identifying true performers and assigning appropriate rewards, and increases the perceived fairness of the evaluation system, which potentially outweighs the benefits of lenient performance feedback. Thus, our results reinforce the importance of evaluation skills as a promotion criterion to secure the achievement of these goals, while also showing that the effort related to the provision of accurate evaluations and the actions to improve employee performance are rewarded.

Our findings underscore the different roles of compensation-based and promotion-based incentives in achieving goal alignment, i.e., incentives to provide goal-directed effort versus incentives to invest in the development of productivity-enhancing skills. A natural question that arises is to what extent these incentives are indeed effective in changing supervisors' evaluation behavior over time. Given our short sample period, we can unfortunately not explicitly examine any learning issues, but we see this as a fruitful avenue for future research.

Finally, although we examine a specific set-up of a CC that is typical in professional service firms, our theory is not limited to this particular design or even having a CC. As long as higher-level managers are able to observe supervisors' evaluation behavior, either directly or indirectly, they can incorporate the observed evaluation behavior in their assessments of the supervisors. Thus, similar incentives can be created irrespective of whether, for example, the same CC also evaluates the supervisors' performance or has the right to make promotion decisions. Nevertheless, another fruitful avenue for future research is to examine whether different types of CCs in terms of composition and allocated decision rights address the incentive problem of biased performance evaluations in different ways.

REFERENCES

Ahn, T. S., I. Hwang, and M. I. Kim. 2010. The impact of performance measure discriminability on ratee incentives. *The Accounting Review* 85 (2): 389–417.

Arshad, F., E. Cardinaels, and B. Dierynck. 2017. *Calibration committees. Costly information acquisition and use for subjective performance evaluation*. Working Paper, Tilburg University. Available at SSRN 3021683.

Baker, G., M. Jensen, and K. Murphy. 1988. Compensation and incentives: Practice vs. theory. *Journal of Finance* 43 (3): 593–616.

Bingham, J., and M. Beer. 2012. *Performance management at Vitality Health Enterprises, Inc.* HBS Case 913-501, Boston, MA: Harvard Business School Publishing.

Bol, J. C. 2008. Subjectivity in compensation contracting. *Journal of Accounting Literature* 27: 1–24.

Bol, J. C. 2011. The determinants and performance effects of managers' performance evaluation biases. *The Accounting Review* 86 (5): 1549–1575.

Bol, J. C., A. Aguiar, and J. Lill. 2019. *Peer-level calibration of performance evaluation ratings: Are there winners or losers?* Working Paper, Tulane University. Available at SSRN 3161692.

Bol, J. C., C. Estep, F. Moers, and M. E. Peecher. 2018. The role of tacit knowledge in auditor expertise and human capital development. *Journal of Accounting Research* 56 (4): 1205–1252.

Bol, J. C., S. Kramer, and V. S. Maas. 2016. How control system design affects performance evaluation compression: The role of information accuracy and outcome transparency. *Accounting, Organizations and Society* 51: 64–73.

Demeré, B. W., K. L. Sedatole, and A. Woods. 2019. The role of calibration committees in subjective performance evaluation systems. *Management Science* 65 (4): 1562–1585.

DeVaro, J., and O. Gürtler. 2015. Strategic shirking in promotion tournaments. *The Journal of Law, Economics, and Organization* 32 (3): 620–651.

Dineen, B. R., R. J. Lewicki, and E. C. Tomlinson. 2006. Supervisory guidance and behavioral integrity: Relationships with employee citizenship and deviant behavior. *Journal of Applied Psychology* 91 (3): 622–635.

Fandt, P. M., and G. R. Ferris. 1990. The management of information and impressions: When employees behave opportunistically. *Organizational Behavior and Human Decision Processes* 45 (1): 140–158.

Faulkender, M., and J. Yang. 2010. Inside the black box: The role and composition of compensation peer groups. *Journal of Financial Economics* 96 (2): 257–270.

Ferris, G. R., and T. A. Judge. 1991. Personnel/ human resources management: A political influence perspective. *Journal of Management* 17 (2): 447–488.

Gibbons, R., and M. Waldman. 1999. A theory of wage and promotion dynamics inside firms. *The Quarterly Journal of Economics* 114 (4): 1321–1358.

Gibbs, M., K. A. Merchant, W. A. Van der Stede, and M. E. Vargus. 2004. Determinants and effects of subjectivity in incentives. *The Accounting Review* 79 (2): 409–436.

Gordon, R. A. 1996. Impact of ingratiation on judgments and evaluations: A meta-analytic investigation. *Journal of Personality and Social Psychology* 71 (1): 54–70.

Grabner, I., and F. Moers. 2013. Managers' choices of performance measures in promotion decisions: An analysis of alternative job assignments. *Journal of Accounting Research* 51 (5): 1187–1220.

Grabner, I., and F. Moers. 2019. *Promotions as implicit incentives for human capital acquisition*. Working paper, Maastricht University. Available at SSRN 2626366.

Hall, B., and C. Madigan. 2000. *Compensation and performance evaluation and Arrow Electronics*. HBS Case 800-290, Boston, MA: Harvard Business School Publishing.

Harris, M. M. 1994. Rater motivation in the performance appraisal context: A theoretical framework. *Journal of Management* 20 (4): 737–756.

Hastings, R. R. 2012. Most large companies calibrate performance, poll finds. *HR Magazine* 57 (2): 87.

Ilgen, D. R., T. R. Mitchell, and J. W. Fredrickson. 1981. Poor performers: Supervisors' and subordinates' responses. *Organizational Behavior and Human Performance* 27 (3): 386–410.

Kahn, C., and Huberman, G. (1988). Two-sided uncertainty and "up-or-out" contracts. *Journal of Labor Economics* 6(4): 423–444.

Künneke, J. 2017. Talent identification and development: A management control perspective. PhD Dissertation, Maastricht University.

Larson, J. R. 1984. The performance feedback process: A preliminary model. *Organizational Behavior and Human Performance* 33 (1): 42–76.

Larson, J. R. 1989. The dynamic interplay between employees' feedback-seeking strategies and supervisors' delivery of performance feedback. *Academy of Management Review* 14(3): 408–422.

Lazear, E. P. 2004. The Peter principle: A theory of decline. *Journal of Political Economy* 112 (S1): S141–S163.

Lillis, A. M., M. A. Malina, and J. Mundy. 2017. *Rendering subjectivity informative in performance measurement and reward systems: Field study insights*. Working Paper, University of Melbourne. Available at SSRN 2998471.

Merchant, K. A., and W. A. Van der Stede. 2017. *Management control systems: Performance measurement, evaluation and incentives*. Harlow, UK: Prentice Hall.

Moers, F. 2005. Discretion and bias in performance evaluation: The impact of diversity and subjectivity. *Accounting, Organizations and Society* 30 (1): 67–80.

Mohrman, A. M., and E. E. Lawler. 1983. Motivation and performance appraisal behavior. In *Performance measurement and theory*, edited by F. J. Landy, S. Zedeck, and J. Cleveland, 173–188. Hillsdale, NJ: Lawrence Erlbaum Associates.

Mumford, T. V., M. A. Campion, and F. P. Morgeson. 2007. The leadership skills strataplex: Leadership skill requirements across organizational levels. *The Leadership Quarterly* 18 (2): 154–166.

Murphy, K. R. 1992. Performance measurement and appraisal: Motivating managers to identify and reward performance. In *Performance measurement, evaluation, and incentives*, edited by W. J. Bruns, 37–62. Boston, MA: Harvard Business School.

Murphy, K. R., and J. N. Cleveland. 1991. *Performance appraisal: An organizational perspective*. Boston: Allyn and Bacon.

Pfeffer, J., and G. R. Salancik. 1975. Determinants of supervisory behavior: A role set analysis. *Human Relations* 28 (2): 139–154.

Podsakoff, P. M., M. Ahearne, and S. B. MacKenzie. 1997. Organizational citizenship behavior and the quantity and quality of work group performance. *Journal of Applied Psychology* 82 (2): 262–270.

Preacher, K. J. and A. F. Hayes (2004). SPSS and SAS procedures for estimating indirect effects in simple mediation models. *Behavior Research Methods, Instruments, and Computers* 36 (4): 717–831.

Prendergast, C. 1993. The role of promotion in inducing specific human capital acquisition. *The Quarterly Journal of Economics* 108 (2): 523–534.

Prendergast, C. 1999. The provision of incentives in firms. *Journal of Economic Literature* 37 (1): 7–63.

Risher, H. 2011. Getting performance management on track. *Compensation & Benefits Review* 43 (5): 273–281.

Risher, H. 2014. Reward management depends increasingly on procedural justice. *Compensation & Benefits Review* 46 (3): 135–138.

Rosaz, J., and M. C. Villeval. 2012. Lies and biased evaluation: A real-effort experiment. *Journal of Economic Behavior and Organization* 84 (2): 537–549.

Roulin, N., A. Bangerter, and J. Levashina. 2015. Honest and deceptive impression management in the employment interview: Can it be detected and how does it impact evaluations? *Personnel Psychology* 68 (2): 395–444.

Rosen, B., and T. H. Jerdee. 1974. Influence of sex role stereotypes on personnel decisions. *Journal of Applied Psychology* 59 (1): 9–14.

Schrand, C. M., and B. R. Walther. 2000. Strategic benchmarks in earnings announcements: The selective disclosure of prior-period earnings components. *The Accounting Review* 75 (2): 151–171.

Sobel, M. E. 1982. Asymptotic confidence intervals for indirect effects in structural equations models. In *Sociological methodology*, edited by S. Leinhart, 290–312. San Francisco: Jossey-Bass.

Tsui, A. S., and B. A. Gutek. 1984. A role set analysis of gender differences in performance, affective relationships, and career success of industrial middle managers. *Academy of Management Journal* 27 (3): 619–635.

Villanova, P., and H. J. Bernardin. 1989. Impression management in the context of performance appraisal. In *Impression management in the organization*, edited by R. A. Giacalone, and P. Rosenfeld, 299–313. Hillsdale, NJ: Lawrence Erlbaum Associates.

Westermann, K., J. Bedard, and C. Earley. 2015. Learning the "craft" of auditing: A dynamic view of auditors' on-the-job learning. *Contemporary Accounting Research* 32 (3): 864–896.

APPENDIX

Variable Descriptions

Variable	Description
BELOW3	indicator variable that equals 1 if a supervisor provides an initial rating that is below three, 0 otherwise.
CC_DOWNGRADE	indicator variable that equals 1 if the calibration committee downgrades the initial ratings of the supervisor more than the average adjustment across all supervisors per year, 0 otherwise.
DIRECT_MARGIN	the supervisor's net engagement revenues for one period.
EFF_UTIL	the supervisor's effective utilization measured as chargeable hours divided by available hours.
GENDER	indicator variable that equals 1 (0) if the supervisor is a male (female).
HIGH_DISCR	indicator variable that equals 1 if a supervisor's discrimination is larger than the median discrimination of all supervisors per year, 0 otherwise.
HIGH_RANGE	indicator variable that equals 1 if a supervisor's rating range is larger than the median rating range of all supervisors per year, 0 otherwise.
JOBTENURE	the supervisor's tenure in the current job (rank) in years.
OVERTIME	the supervisor's hours spent in addition to available hours according to contract.
PRESENT_BEST	indicator variable that equals 1 if the supervisor provides initial ratings only for the best-performing subordinates, 0 otherwise (best is defined as top 25% employees in terms of project ratings within one serviceline per year).
PROJ_RATING	the supervisor's average project rating received from projects completed during the last year.
PROM	indicator variable that equals 1 if the supervisor advances to the next rank, 0 otherwise.
RANGE_SV	the difference between a supervisor's highest and lowest initial rating provided to the subordinates.
RATING	the supervisor's own final rating provided by the calibration committee.
RATING_ADJ	indicator variable that equals 1 if in the same year, the calibration committee downgraded at least one initial rating and upgraded at least one initial rating suggested by the supervisor, 0 otherwise.
RATING_DOWNGRADE	indicator variable that equals 1 if at least one final subordinate rating is lower than the initial rating, i.e., the calibration committee lowered the initial rating suggested by the supervisor but did not upgrade any initial rating, 0 otherwise.
RATING_UPGRADE	indicator variable that equals 1 if at least one final subordinate rating is higher than the initial rating, i.e., the calibration committee increased the initial rating suggested by the supervisor but did not downgrade any initial rating, 0 otherwise.
SPAN_CONT	the number of subordinates per supervisor.
SUBORD_DISCR	reflects the supervisor's discrimination among subordinate performance levels measured as the standard deviation of the initial ratings.
SUBORD_RATING	the average final rating across all subordinates that are assigned to the same supervisor per year.

TABLE 1
Summary Statistics

	n	*Mean*	*Std. Dev.*	*Q1*	*Median*	*Q3*
PROM	1,796	0.22	0.42	0	0	0
RATING	1,796	3.76	0.71	3	4	4
CC_DOWNGRADE	1,122	0.36	0.48	0	0	1
RATING_DOWNGRADE	1,122	0.23	0.42	0.00	0.00	0.00
RATING_UPGRADE	1,122	0.04	0.20	0.00	0.00	0.00
RATING_ADJ	1,122	0.01	0.10	0.00	0.00	0.00
PRESENT_BEST	471	0.06	0.23	0	0	0
SUBORD_DISCR	808	0.43	0.35	0	0.55	0.71
SUBORD_RATING	1,745	3.41	0.45	3	3.33	3.67
PROJ_RATING	645	4.18	0.48	3.88	4.25	4.50
EFF_UTIL	1,796	77.83	24.36	67.0	80.94	92.42
DIRECT_MARGIN	1,796	383,120	293,665	149,632	328,807	564,172
OVERTIME	1,796	183.75	144.42	68.05	153.90	273.70
$RATING_{t-1}$	1,739	3.73	0.68	3	4	4
SPAN_CONT	1,796	3.07	2.19	2	3	4
JOBTENURE	1,796	5.33	4.84	3	4	6
GENDER	1,796	0.68	0.47	0	1	1

Data are for the time period 2010-2012 and the sample size relates to supervisor-years. See Appendix for variable descriptions.

TABLE 2
Correlation Matrix
Correlations Between Variables

		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1	*PROM*	1.00																
2	*RAT'ING*	0.28***	1.00															
3	*CC_DOWNGRADE*	-0.07	-0.07	1.00														
4	*RATING_DOWNGRADE*	0.08	0.08	-0.24**	1.00													
5	*RATING_UPGRADE*	-0.08	-0.09	0.96***	-0.24**	1.00												
6	*RATING_ADJ*	0.14	0.09	-0.14	-0.05	-0.20**	1.00											
7	*SUBORD_DISCR*	0.02	0.00	-0.04	-0.03	-0.04	-0.03	1.00										
8	*PRESENT_BEST*	0.07	0.02	-0.04	0.00	-0.03	-0.01	-0.02	1.00									
9	*SUBORD_RATING*	0.04	0.25**	-0.25***	0.21**	-0.26***	-0.03	0.10	0.03	1.00								
10	*PROJ_RATING*	0.24**	0.74***	0.08	0.02	0.06	0.08	-0.03	-0.08	0.14	1.00							
11	*EFF_UTIL*	0.16*	0.21**	0.05	-0.11	0.03	0.10	-0.16*	-0.03	0.05	0.20**	1.00						
12	*DIRECT_MARGIN*	0.18*	0.38***	0.15	-0.07	0.11	0.01	0.08	-0.03	0.21**	0.34***	0.40***	1.00					
13	*OVERTIME*	0.13	0.13	0.01	0.01	-0.03	-0.15	-0.01	-0.03	0.04	0.13	-0.02	0.13	1.00				
14	$RATING_{t-1}$	0.25**	0.73***	-0.03	0.03	-0.07	0.03	0.02	0.01	0.23**	0.61***	0.15*	0.34***	0.18*	1.00			
15	*SPAN_CONT*	-0.09	0.16*	0.12	0.05	0.18*	-0.04	0.01	0.16*	-0.08	0.15	-0.03	-0.07	-0.11	0.09	1.00		
16	*JOBTENURE*	-0.13	-0.22**	-0.07	-0.05	-0.06	0.07	-0.02	-0.08	-0.15	-0.18*	-0.05	-0.26***	-0.18*	-0.20**	0.09	1.00	
17	*GENDER*	0.08	0.13	0.10	0.15	0.08	0.05	-0.02	-0.08	0.01	0.16*	-0.05	0.07	0.05	0.15*	0.02	-0.04	1.00

Data are for the time period 2010-2012 and the sample relates to supervisor-years. See Appendix for variable descriptions.
*, **, *** is significant at 10 percent, 5 percent, and 1 percent, respectively (two-tailed).

TABLE 3
Effect of Supervisor's Evaluation Behavior on the Supervisor's own Final Rating
(OLS estimations)

	Pred. sign	*(I)*	*(II)*
CC_DOWNGRADE	−	-0.081***	
		(0.029)	
RATING_DOWNGRADE	−		-0.071**
			(0.030)
RATING_UPGRADE	?		0.0135
			(0.063)
RATING_ADJ	?		0.044
			(0.094)
PRESENT_BEST	−	-0.227**	-0.225**
		(0.089)	(0.089)
SUBORD_DISCR	+/0	-0.012	-0.012
		(0.051)	(0.051)
SUBORD_RATING	+	0.076***	0.077***
		(0.027)	(0.027)
PROJ_RATING	+	0.398***	0.397***
		(0.046)	(0.030)
EFF_UTIL	?	-0.0001	-0.0001
		(0.001)	(0.001)
DIRECT_MARGIN	+	0.0002***†	0.0002***
		(0.000)	(0.000)
OVERTIME	+	0.0005***	0.0005***
		(0.000)	(0.000)
$RATING_{t-1}$	+	0.582***	0.582***
		(0.024)	(0.024)
SPAN_CONT	?	0.014	0.015
		(0.010)	(0.010)
JOBTENURE	?	-0.013***	-0.013***
		(0.002)	(0.002)
GENDER	?	0.005	0.004
		(0.025)	(0.025)
Constant		0.031	0.033
		(0.123)	(0.123)
Rank-fixed effects		yes	yes
Year-fixed effects		yes	yes
Observations		1,796	1,796
Adjusted R^2		0.537	0.537

Data used in the estimation are for the time period 2010-2012 and the sample size relates to supervisor-years. Standard errors are in brackets and adjusted for clustering at the supervisor level. See Appendix for variable descriptions.
*, **, *** is significant at 10 percent, 5 percent, and 1 percent, respectively (two-tailed).
† multiplied by 1000.

TABLE 4
Effect of Supervisor's Evaluation Behavior on the Supervisor's Probability of Promotion
(Probit estimations)

	Pred. sign	*(I)*	*(II)*	*(III)*
CC_DOWNGRADE	−/0	0.007	0.072	-0.009
		(0.111)	(0.137)	(0.112)
PRESENT_BEST	−/0	0.272	0.171	0.316
		(0.373)	(0.487)	(0.370)
SUBORD_DISCR	+	0.426**	0.741***	0.307*
		(0.175)	(0.234)	(0.167)
*SUBORD_DISCR*SPAN_CONT*	+			0.144***
				(0.054)
RATING	+	0.675***	0.562***	0.691***
		(0.106)	(0.127)	(0.107)
SUBORD_RATING	?	0.080	0.273	0.074
		(0.097)	(0.166)	(0.096)
PROJ_RATING	?	-0.149	-0.232	-0.150
		(0.139)	(0.194)	(0.139)
EFF_UTIL	?	-0.007***	-0.009***	-0.007***
		(0.002)	(0.003)	(0.002)
DIRECT_MARGIN	?	0.0005***†	0.0004*	0.0005***†
		(0.000)	(0.000)	(0.000)
OVERTIME	?	0.001**	0.001	0.001**
		(0.000)	(0.000)	(0.000)
$RATING_{t-1}$	?	0.326***	0.395***	0.331***
		(0.081)	(0.115)	(0.081)
SPAN_CONT	?	-0.025		-0.068*
		(0.028)		(0.036)
JOBTENURE	?	-0.012	-0.0126	-0.011
		(0.014)	(0.012)	(0.014)
GENDER	?	0.368***	0.270**	0.379***
		(0.089)	(0.116)	(0.090)
Constant		-0.838**	0.090	0.231
		(0.383)	(0.320)	(0.097)
Rank-fixed effects		yes	yes	yes
Year-fixed effects		yes	yes	yes
Observations		1,796	954	1,796
Pseudo R^2		0.357	0.351	0.360

Data used in the estimation are for the time period 2010-2012 and the sample size relates to supervisor-years. Column I (III) reflects results for the whole sample for which data are available. Column II relates to a subsample of supervisors who have at least three subordinates. Standard errors are in brackets and adjusted for clustering at the supervisor level. See Appendix for variable descriptions.
*, **, *** is significant at 10 percent, 5 percent, and 1 percent, respectively (two-tailed).

TABLE 5
Effect of Supervisor's Evaluation Behavior on the Supervisor's own Final Ratings and Promotions (Seemingly Unrelated Regression estimation)

	Pred. sign	*RATING*	*PROM*
CC_DOWNGRADE	−	-0.077**	
		(0.031)	
PRESENT_BEST	−	-0.219**	
		(0.098)	
SUBORD_DISCR	+		0.086**
			(0.035)
SUBORD_RATING	+, ?	0.075***	0.017
		(0.028)	(0.019)
PROJ_RATING	+, ?	0.397***	0.033
		(0.043)	(0.031)
EFF_UTIL	?	-0.0001	-0.001***
		(0.001)	(0.000)
DIRECT_MARGIN	+, ?	0.0002***†	0.0001***†
		(0.000)	(0.000)
OVERTIME	+, ?	0.0005***	0.0002***
		(0.000)	(0.000)
$RATING_{t-1}$	+, ?	0.582***	0.135***
		(0.019)	(0.014)
SPAN_CONT	?	0.013**	-0.003
		(0.006)	(0.004)
JOBTENURE	?	-0.013***	-0.004**
		(0.003)	(0.002)
GENDER	?	0.005	0.086***
		(0.026)	(0.018)
Constant		0.028	0.332***
		(0.103)	(0.073)
Rank-fixed effects		yes	yes
Year-fixed effects		yes	yes
Observations		1,796	1,796
Adjusted R^2		0.537	0.307

Data used in the estimation are for the time period 2010-2012 and the sample size relates to supervisor-years. Standard errors are in brackets. See Appendix for variable descriptions.
*, **, *** is significant at 10 percent, 5 percent, and 1 percent, respectively (two-tailed).
† multiplied by 1000.

TABLE 6
Effect of Supervisor's Evaluation Behavior on the Supervisor's own Final Ratings and Promotions (Seemingly Unrelated Regression estimation)

	Pred. sign	*RATING*	*PROM*
RATING_DOWNGRADE	−	-0.072**	
		(0.032)	
RATING_UPGRADE	?	0.007	
		(0.059)	
RATING_ADJ	?	0.0112	
		(0.118)	
PRESENT_BEST	−	-0.217**	
		(0.099)	
SUBORD_DISCR	+		0.086**
			(0.035)
SUBORD_RATING	+, ?	0.076***	0.017
		(0.028)	(0.019)
PROJ_RATING	+, ?	0.397***	0.033
		(0.043)	(0.031)
EFF_UTIL	?	-0.0001	-0.001***
		(0.001)	(0.000)
DIRECT_MARGIN	+, ?	0.0002***†	0.0001***†
		(0.000)	(0.000)
OVERTIME	+, ?	0.0005***	0.0002***
		(0.000)	(0.000)
$RATING_{t-1}$	+, ?	0.581***	0.135***
		(0.019)	(0.014)
SPAN_CONT	?	0.014**	-0.003
		(0.006)	(0.004)
JOBTENURE	?	-0.013***	-0.004**
		(0.003)	(0.002)
GENDER	?	0.004	0.086***
		(0.026)	(0.018)
Constant		0.028	0.332***
		(0.103)	(0.073)
Rank-fixed effects		yes	yes
Year-fixed effects		yes	yes
Observations		1,796	1,796
Adjusted R^2		0.537	0.307

Data used in the estimation are for the time period 2010-2012 and the sample size relates to supervisor-years. Standard errors are in brackets. See Appendix for variable descriptions.
*, **, *** is significant at 10 percent, 5 percent, and 1 percent, respectively (two-tailed).
† multiplied by 1000.