Influence Activities and Bureaucratic Performance: Evidence from a Large-Scale Field Experiment in China

Alain de Janvry[*], Guojun He[•], Elisabeth Sadoulet[‡],

Shaoda Wang[†], Qiong Zhang[▪]

[Preliminary]

Subjective performance evaluation is widely used by firms and governments to provide work incentives. However, delegating evaluation power to senior leadership could induce influence activities: agents might devote much efforts to please their supervisors, rather than focusing on productive tasks that benefit their organizations. We conduct a large-scale randomized field experiment among Chinese local government employees and provide the first rigorous empirical evidence on the existence and implications of influence activities. We find that state employees are able to impose evaluator-specific influence to affect evaluation outcomes, and that this process could be partly observed by their co-workers. Furthermore, introducing uncertainty in the identity of the evaluator, which discourages evaluator-specific influence activities, can significantly improve the work performance of state employees.

Keywords: subjective evaluation, influence activities, civil servants, work performance, incentive

JEL: M12; D73; F63

[*] University of California, Berkeley. Email: alain@berkeley.edu.
[•] Hong Kong University of Science and Technology. Email: gjhe@me.com.
[‡] University of California, Berkeley. Email: esadoulet@berkeley.edu.
[†] University of Chicago. Email: shaoda@uchicago.edu. *Corresponding Author.*
[▪] Renmin University of China. Email: zhangqiong8@ruc.edu.cn.

I. Introduction

For a large share of jobs in a modern economy, objective performance measures are difficult to obtain, leading employers to rely heavily on supervisors' subjective evaluations to provide work incentives (Prendargast, 1999; Deb et al., 2016). This is particularly ubiquitous in the public sector, due to the inherent problems of measurability and multiplicity of the goals for most civil service jobs (Olken and Pande, 2013; Finan et al., 2015).

While subjective performance measures could potentially improve contractual power (Gibbons and Murphy, 1992; Baker et al., 1994), they also open the door to influence activities: employees can take actions to affect the evaluator's decision in their favor, which might be detrimental to the interests of the organization (Milgrom and Roberts, 1988; Milgrom, 1988). For the organization, influence activities thus create a tradeoff between taking advantage of the supervisor's local information and letting the supervisor manipulate decisions in his own interests, undermining the organization's ability to fully incentivize the agents to perform on their jobs. Despite being consistent with anecdotal examples and case studies, the long-standing theoretical discussion on the existence and implications of influence activities is yet to be backed up by rigorous empirical evidence (Oyer and Schaefer, 2011; Lazear and Oyer, 2012).

Empirically investigating influence activities is fundamentally challenging due to at least three reasons. First, behaviors such as buttering up supervisors or providing personal favors to them are difficult to observe by nature. Agents may try their best to hide such behaviors because they are usually deemed inappropriate. Second, even if these behaviors are observed, it is difficult to conclude that they are driven by intentions to improve evaluation outcomes (instead of simply being friendly), making us unable to exclusively classify them as influence activities. Third, even if the existence of influence activities is established, quantifying the implications of influence activities on work performance still requires exogenous variation in such behaviors across agents.

In this paper, we conduct a large-scale field experiment in two Chinese provinces, which aims to address these three challenges altogether, and provide the first rigorous empirical evidence on the existence and implications of influence activities in workplace. Our experiment focuses on China's "3+1 Supports" program, which hires more than 30,000 college graduates annually to work in rural township governments on two-year contracts.

These junior state employees are referred to as College Graduates Civil Servants (CGCSs) in this paper.

An important institutional feature of the Chinese governance system is its dual-leadership arrangement (Shirk, 1993): each government organization/subsidiary always has two leaders: a "party leader" (i.e. party secretaries at various levels) and an "administrative leader" (i.e. head in a village, mayor in a city).[1] As a result of this duality system, each CGCS responds to two supervisors, both assign job tasks and provide feedbacks on a regular basis. In the *status quo,* each CGCS is evaluated by one of his two co-supervisors every year, and the evaluation outcome in turn determines whether the CGCS can be promoted to a "tenured" position upon completing the two-year term, which is a highly sought-after outcome for most CGCSs due to the prestige of government jobs in China.

Exploiting this unique setting, we collaborate with two provincial governments in China to randomize two different types of performance evaluation schemes among their 3,785 CGCSs. In both schemes, we randomly select one of the two supervisors to be the "evaluator" (whose opinion matters for future promotion), but also collect the assessment from the other non-evaluating supervisor. The only difference is that, in the first "revealed" scheme, we announce the identity of the evaluator to the CGCS at the beginning of the evaluation cycle, so that throughout the year, the CGCS knows whose opinion is 100% responsible for his career development; while in the second "masked" scheme, we keep the identity of the evaluator secret until the end of the evaluation cycle, so that throughout the year, the CGCS perceives each supervisor to be having a 50% chance of determining his promotion.

We find that, in the revealed scheme, the randomly selected evaluating supervisor gives significantly more positive assessments of CGCS performance than his non-evaluating counterpart. In addition, when we ask the colleagues of the CGCS to speculate which of the two supervisors would give higher assessments to the CGCS, they are also more likely to (correctly) point to the randomly selected evaluator, of which they were not informed. These results are consistent with a scenario where the agent is able to impose evaluator-specific influence activities to improve evaluation outcomes, and such influence activities could be at least partially observed by his other co-workers.

[1] The two supervisors have large overlap in their responsibilities, introducing *de facto* checks and balances. See Li (2018) for more detailed discussions on the institutional details of the duality system.

In the masked scheme, since the CGCS no longer knows who the evaluating supervisor is, the asymmetry in supervisor assessments disappears. Furthermore, the masked scheme improves CGCS performance according to a series of different measures, such as average colleague assessments, average colleague assessments, likelihood of being recommended for tenure, and monthly bonuses determined by simple objective performance indicators. The effect of the masked scheme is stronger when the CGCS is more risk averse, when the two supervisors have more aligned preferences for performance, and when the two supervisors have smaller information asymmetry. These results can be rationalized in a simple conceptual framework, where introducing uncertainty in the identity makes supervisor-specific influences riskier and costlier, therefore incentivizing the agent to reallocate efforts from evaluator-specific influence activities toward productive dimensions that would always be appreciated by both supervisors.

We conduct a battery of additional analyses to rule out alternative interpretations of our findings, such as supervisor behavioral change due to being selected to evaluate, agents responding differently to the evaluator's job assignments, evaluating supervisor/masked supervisors getting better information on CGCS performance, CGCSs trying to influence everyone under masked scheme, etc. We also suggest various ways to interpret the magnitude of our treatment effects.

This paper speaks to three strands of literature. First and foremost, it provides the first rigorous empirical test for the existence and implications of influence activities in workplace. As pointed out by Lazear and Oyer (2012), while a large theoretical literature has studied how agents try to extend influence activities in the workplace (e.g. Milgrom and Roberts, 1988; Milgrom, 1988; Meyer et al., 1992; Schaefer, 1998, Alonso et al., 2008; Powell, 2015), there lacks quality empirical evidence, aside from anecdotes and case studies, to verify these arguments. Our paper fills in this gap by providing field experimental evidence on influence activities among Chinese local state employees, as well as quantifying the causal impacts of alleviating influence activities on job performance. More broadly, subjective performance evaluation is ubiquitous across private and public sectors, and has been investigated extensively by a large body of theoretical works (Gibbons and Murphy, 1992; Baker et al., 1994; Prendergast and Topel, 1996; MacLeod, 2003). However, empirical evidence on the effectiveness and limitations of subjective evaluation is still largely missing, with only a handful of exceptions (Chevalier and Ellison, 1999; Hayes and Schaefer, 2000). Our paper contributes

to this literature by showing how influence activities could undermine the power of subjective performance evaluation.

Second, this paper adds to a growing empirical literature on the personnel economics of the state, specifically on incentivizing state employees (Finan et al., 2016). Most of the existing works on this topic focus on the role of financial incentives,[2] with only a few exceptions studying non-pecuniary incentives, such as transfers and postings (Banerjee et al., 2012), and intrinsic motivation (Ashraf at al., 2014). Our paper adds to this line of work by exploiting the (implicit) career incentive involved in performance evaluations, which is a prevalent form of incentive scheme in public sectors due to the often compressed wage structures, but has rarely been studied in the literature. In addition, we show that holding the "final reward" fixed, slightly revising the performance evaluation practice can lead to substantial improvement in job performance of state employees, indicating a highly cost-effective way to strengthen state effectiveness.

Third, our paper relates to the research agenda on Chinese political meritocracy. Since Li and Zhou (2005), a large number of empirical studies have tried to investigate how different performance indicators, such as GDP growth rate (Guo, 2009), environmental standard (He et al., 2018), and implementation of the one-child policy (Serrato et al., 2018), affect the incentives and behaviors of provincial and prefectural officials in China. However, the existing evidence has been focusing almost exclusively on relatively high-level government officials, leaving the vast majority of local bureaucrats under-researched, whose incentives and constraints could potentially be very different from their high-level leaders.[3] Our paper intends to unravel the black box of the incentive schemes for grassroots bureaucrats in China, who are the building blocks of state capacity and play key roles in policy implementation and public service delivery. More broadly, this paper also adds to an emerging literature on bureaucratic performance in developing countries (Bertrand et al., 2019; He and Wang, 2017).

The remainder of this paper is organized as follows. In Section II, we introduce the institutional background, design, and implementation of our field experiment. In Section III, we layout a simple conceptual framework to help rationalize the experimental design, and generate some testable hypotheses to guide the empirical analysis. In Section IV, we present

[2] See Finan et al. (2016) for a summary.

[3] For instance, a key distinction is that, the job tasks for the low-level bureaucrats are much more difficult to be quantified by objective measures such as GDP growth and environmental quality, so most of the grassroots bureaucrats are only rewarded based on subjective evaluations by their supervisors.

experimental results to test the main predictions of our conceptual framework. In Section V, we discuss the potential alternative interpretations of our findings. Section VI concludes.

II. Background and the Experiment

A. Institutional Background

Since the early 2000s, the Chinese government has launched several large-scale human capital transfer programs, which altogether hired more than one million college graduates to work in rural areas, in the hope that their modern human capital and independence from local interest groups could help improve state effectiveness at the local levels. For example, in the College Graduate Village Officials (CGVOs) program, new college graduates were hired as village officials on a contractual basis, and it is found that they were able to significantly improve policy implementation and reduce leakage in poverty subsidies (He and Wang, 2017).

In this paper, we focus on one of such human capital transfer programs, known as the "3+1 Supports" initiative, which was launched in 2006 by the Ministry of Human Resources and Social Security.[4] Several types of civil service positions are offered to the college graduates through this initiative, including township administrative and clerical officers who work on poverty alleviation, agricultural support, water resource management, and social security provision; teachers in township primary schools; and nurses in township clinics. By the end of 2018, more than 350,000 college graduates had been hired as grassroots civil servants through this program. We refer to these individuals as "College Graduate Civil Servants" (CGCSs) in this paper.

The CGCSs are recruited from colleges on a yearly basis. In May, before the job market season for new college graduates, the provincial governments announce the vacancies on their websites and invite college students to apply. In most provinces, the procedure of CGCS recruitment is similar to that of recruiting formal (tenured) state employees: the applicants first take a comprehensive written exam, which is similar to the Administrative Aptitude Test and Essay Writing Test used in the National Civil Service Exam. The high-score applicants are

[4] In Chinese, the initiative corresponds to the "*San Zhi Yi Fu* (三支一扶)" program. Six other major ministries and departments co-sponsored the program, including the Ministry of Education, the Ministry of Finance, the Ministry of Agriculture, the National Health Commission, the State Council Leading Group Office of Poverty Alleviation and Development, and the Communist Youth League Central Committee.

then interviewed, and the top-ranked candidates based on combined scores will be recruited. Some provinces may forego the tests and interviews and screen the applicants simply based on their application materials, where priorities are often given to (Chinese Communist) Party members, those who have local *Hukou* (household registration status), those who hold more advanced and relevant degrees, and those who graduated with honors.

The admission of CGCSs is highly competitive. For examples, Shandong province had around 1,500 positions in 2017, but attracted over 31,000 applicants (admission rate < 5%). Similarly, in Guangxi province, the government planned to hire 800 CGCSs in 2017, but the total number of applicants exceeded 13,600 (less than 6% of applicants were enrolled). In most provinces, the admission rate for the "3+1 Supports" program is consistently below 10%. Such intense competition ensures the quality of the selected CGCSs.

A CGCS is similar to a formal entry-level township civil servant in many ways. For instance, for CGCSs in clerical positions, like most formal rural civil servants, their job tasks tend to be a combination of routine paper works, interactions with villagers, and case-based assignments from supervisors, making job performance difficult to quantify and compare across individuals. For CGCSs in more specialized positions like township clinic nurses or primary school teachers, the job tasks are also similar to their colleagues who are formal state employees. While certain dimensions of these jobs are better defined than clerical jobs, the grassroots nature of these positions still make objective performance evaluation largely infeasible, e.g., student score cannot be used to incentivize teachers due to the lack of unified written exams at the primary school level. As a result, in the *status quo*, the evaluation of a CGCS relies solely on the evaluating supervisor's subjective assessment, which is also the norm for the vast majority of civil service jobs across the world.

The only major difference between a CGCS position and a formal civil servant position is that the former is contract-based, while the latter is tenured.[5] The majority of the CGCSs are eager to be promoted to tenured positions upon finishing their two-year terms, which would only be approved by the government subject to satisfactory supervisor evaluation. As a result, aspiring CGCSs have exceptionally strong incentives to impress their supervisors and improve the evaluation outcomes. While such incentives could encourage higher work efforts, they

[5] In Chinese, it corresponds to the "*Bianzhi* (编制)" status.

might also cause influence activities in ways that are inconsistent with the government's interest.[6]

A CGCS is supervised by two leaders: a party leader and an administrative leader. This is due to the unique dual-leadership governance structure in China: in principal, the administrative leader is in charge of the daily operation of the government entity, while the party leader oversees the process. These two leaders have the same official ranking, but the party leader is normally perceived to have an edge in authority. At the grassroots level such as a township, which is the lowest layer of formal bureaucracy, the division of labor between the two leaders often become less clear, and there tends to be more overlap in their day to day operations. This duality arrangement provides *de facto* checks and balances in China's local governance system (Li, 2018), and is adopted by different levels of administrative units (ranging from the central government to the small village committees). It is also enforced in public institutions such as schools, hospitals, and state-owned firms, as long as there are more than three Communist Party members among the employees.

Under the current evaluation scheme, when a CGCS is first assigned to a township by the provincial Department of Human Resources, he is told explicitly that he needs to work under the supervision of both leaders in his unit, but the Department of Human Resources has designated one of the two leaders as the "evaluator," who is responsible for evaluating CGCS performance at the end of the year. The CGCS therefore knows whose opinion matters for his career development from the beginning of his appointment.

B. Experimental Design

In this section, we explain the experimental design, and lay out the intuitions for the main testable hypotheses. A formal rationalization of the experiment will be presented with a conceptual framework in Section III.

The experimental design is illustrated in Figure 1. In collaboration with two provincial governments in China, we randomize two subjective performance evaluation schemes, a "revealed scheme" vs. a "masked scheme," across all their 3785 CGCSs.

For every CGCS in our sample, one of his two supervisors is randomly selected to be the evaluator, meaning that this supervisor's assessment will be given a 100% weight in the

[6] Simple examples of such influence activities include picking up supervisor's kid from school, making coffee/tea for supervisor, doing chores for supervisor, etc.

evaluation outcome, which in turn affects the future career development of the CGCS. For the other randomized non-evaluator, we also collect his assessment of CGCS performance, but this assessment is given 0% weight in determining the final evaluation outcome, and will therefore not affect the career development of the CGCS. In both schemes, we never directly inform a supervisor whether or not he is chosen as the evaluating supervisor, nor do we inform the colleagues of the CGCS.

Two thirds of the CGCSs are assigned to the "revealed" scheme. In this scheme, we inform (and only inform) the CGCS about the identity of his evaluating supervisor, at the beginning of the evaluation cycle. This essentially mimics the *status quo* of CGCS performance evaluation, where the agent is informed about the evaluating supervisor *ex ante*. The key difference is that, in the *status quo*, the evaluator is endogenously chosen from the two supervisors, typically through an opaque process combining supervisor opinions, division of labor between supervisors, and other idiosyncratic factors. In our "revealed" scheme, by randomly selecting the evaluator, the endogeneity in evaluator selection is eliminated.

We exploit the revealed scheme to study whether "knowing evaluator identity" indeed generates "evaluator-specific influence activities." Since the evaluator is randomly selected from the two supervisors, in the absence of any influence activities, we should see both supervisors on average give similar assessments of CGCS performance. Therefore, if the evaluator ends up giving systematically more positive assessments than the non-evaluator, it can be inferred that the CGCS was able impose evaluator-specific influence activities to improve his evaluation outcomes. At the end of the evaluation cycle, we also asked the colleagues to speculate "who among the two supervisors would be more positive about CGCS performance." While we never informed the colleagues which supervisor was randomized as the evaluator, to the extent that colleagues can at least partially observe the evaluator-specific influence activities imposed by the CGCS, they will be systematically more likely to (correctly) speculate that the evaluator will give more positive assessments than the non-evaluator.

The remaining one third of the CGCSs are assigned to the "masked" scheme. In this scheme, while we still randomize one of the two supervisors as the evaluator, we do not inform the CGCS about the identity of the evaluator until the end of the evaluation cycle. Therefore, from the CGCS's perspective, each supervisor will have a 50% chance to matter for his evaluation outcome. Compared the revealed scheme, the masked scheme introduces uncertainty in evaluator identity, making "supervisor-specific influence activities" riskier and

less beneficial. As a result, risk-averse CGCSs will have incentives to reallocate more efforts toward the "common productive dimensions" that would certainly be appreciated by both supervisors, resulting in better overall work performance.

Exploiting the randomization of CGCSs into "revealed" vs. "masked" schemes, we can test whether "introducing uncertainty in evaluator identity" indeed improves CGCS performance. The main performance indicator we use is the assessments given by other colleagues. We define "colleagues" as co-workers in the same office as the CGCS that are not hired through the "3+1 Supports" program. We consider the colleagues' assessments as the ideal performance measure in this context for three reasons. First, the colleagues were randomly chosen from the same office where the CGCSs worked. They worked closely with CGCSs and can thus best observe the CGCSs' performances. Second, there is no obvious conflict of interest between the CGCSs and their colleagues. Unlike the CGCSs who work in the office only under 2-year contracts, most colleagues already have tenure and have worked in the office for many years. As a result, the CGCSs and their senior colleagues do not directly complete with each other for career advancement. Finally, the CGCSs do not have strong incentive to influence their colleagues for evaluation reasons throughout the study. At the baseline, the provincial governments explicitly told the CGCSs that only the evaluating supervisor's opinions would be used for evaluating their final performance. Most CGCSs in fact did not even expect that we would survey their colleagues until the surveyors were sent to their workplaces at the end line.

To complement colleague assessments, we also measure CGCS performance in three other ways. First, we collect performance assessments given by both the evaluating supervisor, and the non-evaluating supervisor, and analyze how they change when masking evaluator identity. Second, to make sure that the performance results are not driven by "cheap talk," we ask the supervisors and colleagues to make a recommendation to the provincial government, on whether the CGCS should be promoted to a permanent position in this current work unit. Third, we try to benchmark performance objectively using the actual salaries received by the CGCSs. While it is difficult to measure performance objectively due to the multi-dimensional and vaguely-defined job tasks for most CGCSs, for some specific positions, a modest amount of "monthly bonus" is linked to some key objective performance indicators.[7] Therefore, we

[7] For example, CGCSs who serve as nurses receive bonuses based on the amount of night shifts they take.

can compare the actual salaries received by CGCSs between the two schemes, and recover (based on the bonus pay algorithms) the differences in objective performance.

C. Implementation

Our experiment is conducted in collaboration with two of the largest provinces in China, with a combined population of more than 170 million. Province A is coastal and more developed, while Province B is inland with lower average income. Province A recruits the CGCSs using an automatic scoring system, while Province B recruits through traditional written tests and interviews. Our sample covers all the 3,785 CGCSs employed by these two provinces in China as of September 2017 (cohorts of 2016 and 2017). Our research team was appointed by the two provincial Human Resources Departments as the "third-party evaluator" for their "3+1 Supports" program, to help launch new performance evaluation schemes, which the provincial governments officially informed all the CGCSs multiple times. This high-level endorsement helped ensure that the vast majority of the CGCSs were well aware of the high stakes involved in the evaluation outcomes under our newly launched evaluation schemes.

The baseline survey was carried out in September 2017, one month after the 2017 cohort CGCSs finished job training and were assigned to their positions. Every CGCS was then randomized into one of the two evaluation schemes. The randomization was conducted at the work unit level instead of the individual level.[8] Different CGCSs who worked in the same unit (i.e., an organization branch led by the same set of supervisors) were assigned to the same scheme. This is per the request of our government partner to ensure that the evaluation results of CGCSs working in the same unit can be fairly compared to each other. In this setting, because 83.9% of the work units only had one CGCS allocated, randomizing at the work unit level did not hurt our statistical power in any substantial way, as compared to randomizing at the individual level.

Based on the result of the randomization, in September 2017, we informed each CGCS about the evaluation scheme that he was assigned to. Specifically, if a CGCS was randomized into the revealed scheme, we notified him that "*among your two supervisors A and B, we randomly selected supervisor A to be your evaluator, whose opinion will be collected by the end of this evaluation cycle and provided to the provincial Human Resources Department for their review.*" If a CGCS was randomized

[8] In Chinese, a work unit corresponds to a "*Gong Zuo Dan Wei* (工作单位)."

into the masked scheme, we notified him that "*among your two supervisors A and B, we will randomly select one of them to be your evaluator. The randomization will be realized at the end of this evaluation cycle, at which time the evaluator's opinion will be collected and provided to the provincial Human Resources Department for their review."* The exact contents of these individualized notification letters are translated in the Appendix.

To ensure the credibility of our intervention, the two provincial governments also sent formal notifications with official stamps to every CGCS. The government notifications emphasized the importance of this "third-party" performance evaluation, and confirmed the evaluation schemes that we sent to the CGCSs. We reminded the CGCSs about their evaluation schemes in January 2018.

The end-line survey was carried out in June 2018, and consisted of three parts: colleague assessment, supervisor assessment, and self-assessment. When the surveyors visited the office where a CGCS worked, if there were less than 5 colleagues in the office, all of them were invited to fill in the colleague questionnaire; if there were more than 5 colleagues, the surveyor randomly sampled 5 of them to fill in the colleague questionnaire, using a random number generator.[9] To protect the privacy of the colleagues and encourage truth-telling, colleague questionnaires were strictly anonymous, and CGCSs were not allowed to communicate with the colleagues during the entire process. The CGCS survey was also conducted on-site, but independently from the colleague survey to avoid interference. Supervisor assessment was completed online, with an individual-specific link for each supervisor, listing all the CGCSs in his unit.

In the colleague and supervisor surveys, we collected information on the main characteristics of the colleague/supervisor, their interactions and familiarity with the CGCS, the job tasks of the CGCS, and their assessments of the CGCS along various dimensions. Specifically, we asked for an overall assessment of CGCS performance, as well as a "revealed preference" measure asking each colleague/supervisor whether he recommends that the CGCS be promoted to a permanent civil servant in the current work unit.

The end-line CGCS survey followed a similar structure by asking about interactions with supervisors/colleagues and self-assessments along multiple dimensions. In addition, we also

[9] If a colleague was not in the office when the surveyor visited, his contact information was collected and he would be surveyed over the phone in the following day. To ensure data accuracy, the leader of the surveying team randomly called some of the surveyed colleagues in the following days to verify the sampling procedure and the answers collected.

asked a series of questions related to future career plans and satisfaction with the "3+1 Supports" program.

D. Balance and Attrition Tests

To ensure that the randomization was well executed, we conduct a battery of balance tests: in the revealed scheme, supervisor characteristics should be balanced between the evaluating and non-evaluating supervisors, which supervisor is selected to evaluate should also be orthogonal to both CGCS and colleague characteristics; between the two evaluation schemes, CGCS characteristics, supervisor characteristics, as well as colleague characteristics should all be balanced. As shown in Appendix Tables A1-A3, these balance tests all go through, suggesting that our randomization was well executed.

Between the baseline and the end-line surveys, we lost 918 CGCSs (24.3%) in the sample. The main cause for attrition is that some CGCSs or their supervisors were re-assigned to different job posts during our study period (14.9%). For example, a CGCS could be relocated from one township to another because of changes in government priorities. The supervisors could also be promoted, retired, or rotated to other institutions during this period. Such job changes would break the supervisor-subordinate relationship defined by our intervention, and thus invalidate the experimental design. Besides, some CGCSs passed the formal civil service exams or got admitted by graduate schools, and decided to quit their jobs during our experiment (7.4%). To test whether our experiment would suffer from any attrition bias, we estimate the relationship between attrition status and the treatment status in Appendix Table A4. We find that our treatment does not lead to selective attrition.

III. Conceptual Framework

In this section, we present a simple conceptual framework to rationalize the experiment, and derive a series to propositions which will in turn guide the empirical analysis.

Assume an CGCS i works on a job with a productive dimension x_i, which can be observed by co-workers and supervisors, but cannot be verified quantitatively. The organization therefore relies on a subjective performance evaluation scheme, where the reward of the agent depends on the assessments given by his supervisors.

To mimic our "duality" empirical setting, we assume that there are two supervisors, $j \in \{a, b\}$. In addition to working on the productive dimension of the job, CGCS i also has the option to exert (supervisor-specific) influence activities u_{ij} to please supervisor j, in order to improve his assessment score:

$$Y_{ij} = x_i + u_{ij}$$

The agent maximizes his expected utility, which is a function of both evaluation outcome and work efforts, subject to a time constraint:

$$\max_{x_i, u_{ij}} \sum_{j \in \{a,b\}} s_{ij} \cdot V(Y_{ij},\ L_i)$$

s.t.

$$f(x_i) + \sum_{j \in \{a,b\}} g(u_{ij}) + L_i = T$$

where s_{ij} is the probability of each supervisor j's assessment being used to determine individual i's reward in the performance evaluation scheme ($\sum_{j \in \{a,b\}} S_{ij} = 1$); L_i is leisure; $V(Y, L)$ is the utility function of the CGCS, which is increasing and concave in both evaluation score and leisure. In the time constraint, $f(x)$ is a convex and increasing function measuring the time needed to obtain performance x; $g(u)$ is a convex and increasing function measuring the time needed to exert influence activities u; T is the total time budget for an individual.

Other colleagues also observe CGCS performance, but their opinions are not included in the performance evaluation scheme, so the agent is not incentivized to adjust his efforts to improve colleague assessments. Colleagues therefore receive no influence activities from the CGCS, and base their assessments solely on the productive dimension:

$$Y_{ic} = x_i$$

To mimic our experiment, suppose that one of the two supervisors is randomly chosen to evaluate CGCS performance, and the other supervisor's opinion bears no weight in the evaluation. When we inform the CGCS about the identity of the evaluator (*revealed scheme*), the CGCS knows exactly whose opinion matters for his career development: $s_a = 1, s_b = 0$ or $s_a = 0, s_b = 1$. When we do not inform the CGCS about the identity of the evaluator until

the end of the evaluation cycle (*masked scheme*), the CGCS perceives each supervisor as equally likely to determine his career development: $s_a{}' = s_b{}' = \frac{1}{2}$. Solving the model under both the revealed scheme and the masked scheme, we can derive several propositions to guide the empirical analysis.

Proposition 1: *Under the revealed scheme, the agent extends evaluator-specific influence activities, and the evaluating supervisor would give higher assessment than the non-evaluating supervisor.*

Without loss of generality, assume: $s_a = 1, s_b = 0$. Solving the model, we can show that:

$$u_{ia} > u_{ib} = 0$$

which leads to:

$$Y_{ia} > Y_{ib}$$

The intuition is that, when the agent knows the identity of the evaluator, he would exert more evaluator-specific efforts ($u_{ia} > u_{ib}$) that lead to more positive assessment from the evaluating supervisor ($Y_{ia} > Y_{ib}$).

Proposition 1.1: *Under the masked scheme, the agent spends equal efforts influencing each supervisor. Evaluator-specific influence is lower compared to the case where evaluator identity is known.*

In this masked scheme, $s_a{}' = s_b{}' = \frac{1}{2}$. Since both supervisors are equally likely to become the evaluator, the agent should spend equal efforts trying to influence both. Since $s_a = 1 > s_a{}' = \frac{1}{2}$, the marginal return to influence activities is lower, which leads to reduced levels of influence activities toward the evaluator ($u'_{ia} = u'_{ib} < u_{ia}$).

Proposition 2: *Switching from the revealed scheme to the masked scheme would improve CGCS performance. Specifically, both average colleague assessment and average leader assessment would improve.*

In the masked scheme, since the identity of the evaluator is unknown, supervisor-specific influence activities become riskier and less beneficial. As a result, the agent would reallocate efforts from supervisor-specific influence activities (u_{ij}) toward the common productive tasks that would be appreciated by both supervisors (x_i). Therefore, compared to the revealed scheme, we can show:

$$x'_i > x_i$$

and:

$$Y'_{ia} + Y'_{ib} > Y_{ia} + Y_{ib}$$
$$Y'_{ic} > Y_{ic}$$

The intuition is that, since the identity of the evaluator is unknown, the marginal cost of improving expected evaluation outcome through influence activities doubles, while the marginal cost of improving evaluation outcome through the productive dimension remains unchanged. Therefore, the agent re-optimizes by switching from exerting influence activities towards being productive for the organization.

Proposition 2.1: *If the evaluator is more positive under the masked scheme than the revealed scheme, the CGCS must enjoy less leisure and lower utility under the masked scheme than the revealed scheme.*

Formally, given that $(x_i, u_{ia}) \in \underset{x_i, u_{ia}}{\operatorname{argmax}} V(Y_{ia}, L_i)$, we know that $V(Y'_{ia}, L'_i) < V(Y_{ia}, L_i)$. If $Y'_{ia} > Y_{ia}$, it must be that $L'_i < L_i$. Since $Y'_{ia} = Y'_{ib}$, we also have $0.5 \cdot V(Y'_{ia}, L'_i) + 0.5 \cdot V(Y'_{ib}, L'_i) < V(Y_{ia}, L_i)$.

The intuition is that, in the revealed scheme, the agent only has incentives to please the evaluating leader. If the evaluating leader is happier in the masked scheme, then by revealed preference, the agent must have lower leisure and utility, otherwise his inputs under the revealed scheme would be strictly dominated by the inputs under the masked scheme.

Proposition 2.2: *More risk-averse CGCSs are more strongly incentivized when the identity of the evaluating supervisor is unknown.*

When the agent is more risk averse (i.e. $V(Y, L)''$ being pointwise more concave than $V(Y, L)'$), the reallocation from uncertain dimension (u_{ij}) to certain dimensions (x_i and L_i) would be more salient. So we can show:

$$(Y''_{ia} + Y''_{ib}) - (Y_{ia} + Y_{ib}) > (Y'_{ia} + Y'_{ib}) - (Y_{ia} + Y_{ib})$$

Proposition 2.3: *When at least one of the supervisors has imperfect information about job performance, masking evaluator identity is less effective as the supervisor information gap enlarges.*

Without loss of generality, assume:

$$Y_{ia} = x_i + u_{ia};\ \ Y_{ib} = \alpha \cdot x_i + u_{ib}$$

where $\alpha < 1$. Solving the model, it is straightforward to show that:

$$\frac{\partial (Y'_{ic} - Y_{ic})}{\partial \alpha} > 0$$

The intuition is that, when $\alpha \to 0$, x_i becomes more and more specific to supervisor a, therefore similar to the role of u_{ia}. As a result, the effect of "reallocating towards commonly appreciated dimension" is attenuated.

Proposition 2.4: *When job tasks are multi-dimensional, masking evaluator identity is more effective when the two supervisors' subjective weights for different dimensions of performance are more consistent.*

Without loss of generality, we extend the previous setting by allowing for two dimensions of performance: x_{i1}, x_{i2}, for which a supervisor imposes weights w_{j1} and w_{j2} respectively ($w_{j1} + w_{j2} = 1$). So the evaluation function becomes:

$$Y_{ij} = w_{j1} \cdot x_{i1} + w_{j2} \cdot x_{i2} + u_{ij}$$

and the time constraint becomes:

$$f(x_{i1}) + f(x_{i2}) + \sum_{j \in \{a,b\}} g(u_{ij}) + L_i = T$$

Solving the model, we prove that:

$$\frac{\partial [(Y'_{ia} + Y'_{ib}) - (Y_{ia} + Y_{ib})]}{\partial |w_{a1} - w_{b1}|} < 0$$

and:

$$\frac{\partial (Y'_{ic} - Y_{ic})}{\partial |w_{a1} - w_{b1}|} < 0$$

The intuition is that when both supervisors have more consistent subjective weights, it is more desirable to reallocate efforts from supervisor-specific influence activities (u_{ij}) to productive dimensions that would be appreciated by both supervisors (x_{i1}, x_{i2}), leading to better CGCS performance.

IV. Main Results

In this section, we present a series of experimental results to verify the main propositions of our conceptual framework. Specifically, we show that: (1) when the identity of the evaluator is revealed to the agent, the assessment given by the evaluating supervisor is substantially higher than that given by the non-evaluating supervisor, and this asymmetry in supervisor assessments is correctly predicted by the colleagues of the agent, confirming *proposition 1*; (2) when the identity of the evaluator is masked, the asymmetry in supervisor assessments disappears, colleagues also no longer speculate that the randomized evaluator will give more positive assessments, confirming *proposition 1.1*; (3) masking evaluator identity leads to significant improvements in CGCS work performance according to various measures, confirming *proposition 2*; (4) if masking evaluator identity improves evaluator assessment, it decreases the leisure and utility of the CGCS, confirming *proposition 2.1*; (5) the masked scheme is more effective when the CGCS is more risk-averse, when the two supervisors have similar opinions on what constitutes good performance, and when there is little information asymmetry between the two supervisors, confirming *propositions 2.2-2.4*.

These results are highly compatible with the interpretation of the agent undertaking influence activities (that target the evaluator and at the expense of job performance) in the revealed scheme, and reallocating efforts from influence activities toward productive dimensions in the masked scheme, as formalized in our conceptual framework. The combination of these findings could hardly be reconciled with alternative interpretations, and we will further confront each of those remaining confounding mechanisms in Section V.

A. Proposition 1: Asymmetry in Supervisor Assessments under the Revealed Scheme

In Table 1, the outcome variable is "Supervisor 1's assessment score minus Supervisor 2's score," measuring the extra positiveness of Supervisor 1 than Supervisor 2 towards the same CGCS.[10] The explanatory variable is a dummy variable indicating whether Supervisor 1 was (randomly) chosen to be the evaluating supervisor. We focus on the revealed scheme, in which every CGCS is informed about the identity of his randomized evaluator at the beginning of the evaluation cycle.

[10] "Supervisor 1" and "Supervisor 2" are random labels we give to each CGCS's two supervisors.

In Column (1), we find that in the revealed scheme, if a supervisor was chosen as the evaluator at the baseline, he indeed gave more positive assessment at the endline. In Column (2), we include a rich set of control variables in the regression, and the estimated coefficient remains highly stable, confirming that the treatment of "Supervisor 1 Evaluating" was indeed randomly assigned. This asymmetry in supervisor assessments is consistent with the agent imposing evaluator-specific influence activities to improve evaluation outcomes.

If this asymmetry is indeed driven by influence activities, to the extent that such behaviors can at least be partially observed by other co-workers in the same office, we should expect that in the revealed scheme, colleagues could update their priors on which of the two supervisors would be more positive about CGCS performance. In other words, when colleagues receive some signals of u_{ij}, even without knowing who was chosen as the evaluator, they should still be more likely to correctly speculate that the evaluating supervisor would be more positive.

We test this prediction in Columns (3)–(4), where the outcome of interest is a dummy variable indicating "whether a colleague thinks supervisor 1 would be more positive." We see that when supervisor 1 is randomly selected as the evaluator, colleagues are more likely to think he is going to give more positive assessments. Combined with the results in the first two columns, colleagues indeed correctly predict the direction of supervisor assessment asymmetry.

B. Proposition 1.1: No Asymmetry under the Masked Scheme

If the findings in Table 1 are indeed caused by evaluator-specific influence activities as we interpreted, then these results should only exist when the agent knows who the "target" is. Specifically, under the masked scheme, when the CGCS no longer knows the identity of the evaluator, there should no longer be any asymmetry in leader assessments either.

Therefore, as a placebo test, in Table 2, we focus on the masked scheme, where the identity of the randomly chosen evaluator is not announced until the end of the evaluation cycle. As we can see in the first two columns, being selected as the evaluator no longer has any impact on assessment positiveness.

In Columns (3)–(4), we find that when the identity of the evaluating supervisor is unknown to the agent, colleagues are no longer able to identify the evaluator as the more positive supervisor.

Combined together, these results are consistent with a scenario where agents impose evaluator-specific influence activities when and only when knowing who the evaluator is.

C. Proposition 2: Improved Performance under Masked Scheme

We exploit the random assignment of CGCSs between the two evaluation schemes to estimate the causal effect of masking evaluator identity on job performance.

First, in Table 3, we examine colleague assessments, which we consider as the ideal measure of CGCS performance, for reasons explained in Section III. Specifically, the dependent variable is average colleague assessment of the CGCS's performance, which is framed relative to other civil servants working in the same work unit. The assessment score ranges from 1 to 7, representing different categories from "worse than all the colleagues" to "better than all the colleagues" in the questionnaire. Relatedly, we have an outcome variable indicating whether the colleagues think the CGCS's performance ranks in the Top 10% of the organization.

The results in Columns (1)–(2) show that masking the identity of the evaluator significantly improved colleague assessments of CGCS performance. In Columns (3)–(4), we use the "Top 10%" as the outcome variable, all the results remain similar: CGCSs in the second treatment arm are more likely to be recognized as top performers by their colleagues.

These results are consistent with our theoretical prediction that masking evaluator identity would reduce influence activities, thus improving job performance. To better interpret the economic significance of our "masked scheme" intervention, we compare the magnitude of the "masking effect" to other CGCS characteristics that strongly predict colleague assessment score. In Appendix Table X, we report the correlations between subjective assessments and a rich set of CGCSs' characteristics. As we can see, education and CCP membership have the strongest predictive power in explaining colleague assessments: graduating from a 4-year college instead of a 3-year community college is correlated with a 0.15-point increase in colleague assessment, and being a party member is associated with a 0.17-point increase in colleague assessment. The "masking" effect (0.22) is therefore substantially larger than the effect of "upgrading" 3-year college graduates to 4-year college graduates, or replacing non-party members with party-members. Given the substantial edge in prestige and ability associated with 4-year colleges and party member status, the treatment effect does seem to be economically significant.

We corroborate the colleague assessments with supervisor assessments. Table 4 reports the effect of masking evaluator identity on supervisor assessments. In Columns (1), the outcome variable is the mean assessment of the two supervisors. We find that "masking the identity of the evaluator" significantly improves average supervisor assessment. To understand how the CGCSs manage to achieve overall higher assessments, we look at the treatment effects on evaluator supervisor and non-evaluating supervisor separately. Columns (2) and (3) indicate that both the evaluator and the non-evaluator become more positive about CGCS performance under the masked scheme, and the improvement is more salient for the non-evaluator. The difference between evaluating and non-evaluating supervisors' treatment effects motivate Column (4): deviation between supervisor assessments is smaller under the masked scheme.

In addition to the subjective assessments given by colleagues and supervisors, we also explore other more "revealed-preference" and "objective" performance measures to further support our findings. The results are presented in Table 5.

In the first column, as a "revealed-preference" measure, we directly ask the colleagues whether they recommend to the provincial government that the CGCS be promoted to a tenured position in this office after finishing his 2-year term. Using this as an alternative outcome, we find that masking evaluator identity makes more colleagues think the CGCS deserves tenured, again suggesting an improvement in performance.

We also asked each CGCS to report their total monthly remuneration, including basic wages and performance bonuses (if there is any), which we in turn verified using administrative information provided by the provincial governments. The basic wage is set by upper-level government and should be exactly the same for all the CGCSs, conditional on the county of residence, enrollment year, and position. In addition to the basic wage, each working unit has some discretion over a modest amount of performance bonus to reward the best performing employees. In Columns (2) to (3) of Table 5, we observe that the CGCSs in the masked scheme earn 2% (roughly 49 RMB per month, 0.3 S.D.) higher salaries than those in the revealed scheme. Since the basic salary for CGCSs is fixed (matched to the entry-level permanent civil servant wage), this 2% gap completely reflects the difference in performance bonus.

During our field interviews, we were informed that mainly CGCSs who work as nurses in township clinics enjoy performance bonuses, because these clinics have a "business" feature and can keep some profits to reward the most hard-working staffs. For nurses, the most

important factor determining their wage differentials is the number of night shifts taken each month: every additional night shift is rewarded by about 20 RMB (about $3). In Column (4), when we restrict the sample to nurses, we find the coefficient indeed gets substantially larger (116 RMB). The compensation differential between the "revealed" group and "masked" group is therefore equivalent to 6 more night shifts every month. This result suggests that the performance improvement caused by the masked scheme is indeed substantial when measured objectively.

Taken together, by exploring various measures, we find highly consistent evidence suggesting that CGCS performance improved significantly under the masked scheme.

D. Proposition 2.1: Rationalizing Results on Supervisor Assessments

In Table 4, the result for the non-evaluating supervisor's assessment is intuitive: in the revealed scheme, the CGCS does not have incentives to please the non-evaluator; while in the masked scheme, the CGCS has incentives to please both leaders since he does not know who will end up being selected as the evaluator. Therefore, the non-evaluator is expected to be more positive about the CGCS in the masked scheme.

The improvement in the evaluating supervisor's assessment, however, is less intuitive: the evaluating supervisor's assessment score is also improved, despite no longer receiving influence activities catered to his private utility. As discussed in our conceptual framework, an improvement in evaluator assessment under the masked scheme has two implications: (1) the CGCS works much harder under the masked scheme; (2) the CGCS has lower utility under the masked scheme due to reduced leisure. We test these predictions in Table 6.

To measure "reduced leisure," in our survey, besides asking the colleagues to give an overall assessment of the CGCS's performance, we also asked them to specifically tell us whether they thought the CGCS worked harder and overtime relative to other colleagues. In Panel A of Table 6, we find that masking the evaluator's identity indeed makes the CGCS work harder and overtime (as observed by their colleagues), suggesting that the CGCS allocates more efforts to the productive dimensions. This is also corroborated by the self-assessments of the CGCSs.

To measure "lower utility," in our survey, we asked the CGCS whether they would be willing to find a job in the private sector, and if so, what the acceptable wage was. In Panel B of Table 6, we find that masking the identity of the evaluator makes the CGCS more willing to accept

a job in the private sector, and the "reserve wage" is reduced by about 5%. Our interpretation is that because the CGCSs had to work harder under uncertainty, consequently more willing to quit and accept a low-pay job. These results help rationalize our previous finding that both the evaluating and non-evaluating supervisors' assessments improved.

E. Proposition 2.2: Masking is more Effective for Risk-Averse CGCSs

Masking Evaluator identity introduces uncertainty in evaluator identity, which makes supervisor-specific influence activity a riskier investment. Therefore, *Proposition 2.2* of our conceptual framework suggests that if a CGCS is more risk-averse, when we mask evaluator identity, he should have a more salient performance bump under the masked scheme.

We confront this proposition in the first column of Table 7. We elicit the level of risk aversion of each CGCS, and interact this measure with the dummy for masked scheme.[11] As we can see, the "masking effect" is indeed significantly stronger for more risk-averse CGCSs, consistent with the theoretical prediction.

F. Proposition 2.3: Masking is more Effective when the Supervisors Have More Aligned Preferences

Another prediction of the model is that when the preferences of the two supervisors are highly aligned, masking evaluator identity should have the most powerful effect, because the "economy of scale" for reallocating efforts toward the "common productive dimensions" increases with the alignment of supervisor preferences.

In our survey, we separately elicited each supervisor's subjective ranking of the importance of each performance dimension. In Column (2) of Table 7, we find that the treatment effect is larger when the two supervisors give the exact same ranking of importance, confirming the prediction. This result suggests that the treatment effect of masking evaluator identity is not driven by "CGCS responding to more diversified supervisor preferences," because if that is the case, we should see a weaker treatment effect when the supervisors have more aligned preferences, not stronger.

[11] Risk attitude is elicited through a hypothetical coin-flipping game. An individual taking a certain 400 Yuan over a 50% chance to win 2000 Yuan is defined as highly risk averse. Our results are robust to alternative cutoffs for this definition.

G. Proposition 2.4: Masking is more Effective when the Supervisors Have More Symmetric Information

Our model also predicts that when the information gap of the two supervisors gets larger, the "productive dimension" effectively becomes more supervisor-specific. In other words, larger information asymmetry between the two supervisors weakens the treatment effect of masking evaluator identity.

This proposition is confirmed by both Column (3), where we use "difference in two supervisors' self-reported information on CGCS performance" to measure information asymmetry, and Column (4), where we use "two supervisors' difference in work assignment frequency" to proxy for information asymmetry. In both regressions, we find that when the two supervisors have more asymmetric information, masking evaluator identity leads to a smaller improvement in performance.

H. Additional Analyses for Heterogeneity

In the Appendix, we further conduct a rich set of heterogeneity analyses along other dimensions. In Table A6, we find that the treatment effect is similar for different cohorts. In Table A7, we explore whether the treatment effect is slightly larger in the poor province, which may reflect the conventional wisdom that favoritism and influence activities are more prevalent in less developed areas. In Table A8, we report the heterogeneity with respect to the CGCSs' personal characteristics. We find that female CGCSs respond slightly more strongly to the treatment than their male counterparts. Table A9 investigates impact heterogeneity regarding to CGCSs' motivations. We see that the treatment effect is larger among those who are motivated to stay in the post (close to family) and those who have stronger career incentives (i.e. wanting to purse graduate degree or become formal civil servants). Table A9 further reports the results for CGCSs' attitudes. The treatment effect is smaller for those who reported that they "are satisfied with their work performance," "always try the best to finish the job tasks," and "live a very fast pace of life" at the baseline. One likely explanation is that those CGCSs could be the hard-working ones at the first place and would find it difficult to further improve their work performances.

V. Alternative Interpretations

Our model suggests that under the revealed scheme, an agent imposes evaluator-specific influence activities; and under the masked scheme, the agent no longer knows who is evaluating, so would re-optimize his efforts to work harder on the productive dimensions that would be appreciated by both supervisors. In Section IV, we documented a series of evidence supporting our theoretical propositions. In this section, we discuss the main confounders for our empirical results, and test their validity using our data.

A. Alternative Interpretations to Proposition 1

Our interpretation of the findings in Table 1 is based on *Proposition 1*: in the revealed scheme, the subordinate is able to perform evaluator-specific influence activities, and such behaviors can be observed by their co-workers. There are two potential confounding explanations.

1. Behavioral Differences between the Evaluating and Non-Evaluating Supervisors

The first alternative explanation is that in the revealed scheme, the evaluating and non-evaluating supervisors may act differently, simply because "revealing the identity of the evaluator" might directly affect the evaluator's behaviors. For example, if the evaluating supervisor feels more responsible after knowing that he is the evaluator, he may give the CGCS more positive assessments to avoid potential psychological burdens. This concern is alleviated by our experimental design, because the supervisors were not informed by the research team about their roles in the evaluation. But it is still possible that the CGCS might have delivered this information to his evaluator.

To investigate this possibility, in our endline survey, we directly asked the supervisors whether they were aware of their responsibility in evaluating the CGCS. It turns out that the majority of them (more than 65%) did not know whether they were chosen as evaluators until the end of the evaluation cycle. In Panel A of Table 8, we re-estimate the specification in Table 1 separately for two subsamples: the subsample in which the supervisors did not know their (evaluator) roles, and the subsample in which the supervisors knew their evaluator roles (likely through the CGCS). We find that the "supervisor asymmetry" is almost identical in the two subsamples, suggesting that our results are not driven by supervisor behavioral change due to being the evaluator.

Moreover, in Panel B of Table 8, we directly examine the existence of behavioral differences between the evaluating and non-evaluating supervisors. We focus on three outcomes: (1) the likelihood of a supervisor not responding to the survey; (2) whether one supervisor writes more words in describing the CGCS's job tasks than the other supervisor; and (3) whether one supervisor assigns more job tasks to the CGCS than the other. Our hypothesis is that, if the evaluating supervisor indeed paid more attention to the CGCS, we should observe the evaluating supervisor to be more likely to answer the survey, write more words in their assessment, and even assign more tasks to the CGCS. Again, the data does not support this interpretation, as evidenced by the regression results in Panel B.

2. Higher Information Quality for Evaluating-Supervisor

Another confounding story is that even without behavioral changes, the evaluating supervisor would receive more information regarding CGCS performance from various sources: the CGCS, the colleagues, and the other (non-evaluating) supervisor might try to send signals to help him better evaluate. This improvement in information quality might improve supervisor assessment and thus create the supervisor asymmetry that we observed in Table 1.

To examine this interpretation, in our end-line supervisor survey, we asked each supervisor how frequently did the subordinated CGCS, the colleagues of the CGCS, or the other supervisor discuss with him about CGCS performance. We are interested in whether the evaluating supervisor would receive additional information than the non-evaluating supervisor from these three sources. In panel C of Table 8, we find that "Supervisor 1 Evaluating" does not make supervisor 1 obtain extra information from any of these sources, as compared to supervisor 2. Therefore, the asymmetry in supervisor assessments under the revealed scheme cannot be explained by the difference in information quality between the two supervisors.

B. Alternative Interpretations to Proposition 2

Our interpretation of the "improved supervisor and colleague assessments" under the masked scheme is based on *Proposition 2*: masking evaluator identity makes "supervisor-specific influence activities" riskier and less beneficial, which incentivizes the CGCSs to work harder on the "common productive dimensions" appreciated by both supervisors, resulting in better work performance. There are three potential confounding explanations.

1. CGCS Imposing More Influence Activities to Both Supervisors

A potential alternative interpretation is that under the masked scheme, the CGCS does not work harder on the productive dimensions. Instead, he simply imposes more influence activities to both supervisors, which is why we see an improved assessments from both supervisors. However, this interpretation is inconsistent with a series of our empirical results.

First, it is inconsistent with the fact that colleague assessments improved substantially under the masked scheme. As explained in Section II, the CGCS has no incentives to influence his colleagues: every CGCS is clearly informed that only his evaluating supervisor's opinion will be taken into account by the provincial government, and colleague assessments will never enter into their promotion functions. Therefore, if the CGCS is simply extending more influence activities toward both supervisors instead of working harder, there should not be a significant improvement in colleague assessments.

Second, if the CGCS is imposing more influence activities instead of working harder, we also should not observe significant performance improvements under the masked scheme. As discussed in Section III, CGCSs under the masked scheme receive substantially higher performance bonuses, which are directly linked to objective performance indicators. This again supports our interpretation over the competing hypothesis.

Third, the alternative interpretation is also at odds with the heterogeneity results presented in Table 7: if the CGCS is indeed trying to impose more influence instead of working more on the productive dimensions, then whether the two supervisors have aligned preferences along the productive dimensions, and whether they have comparable information\discretion along the productive dimensions, should not affect the impacts of the masked scheme. The heterogeneity results regarding characteristics of the productive dimensions are therefore consistent with our theoretical predictions, and inconsistent with the "increased influences" hypothesis.

Besides the above arguments, we also conduct additional placebo tests to further rule out the alternative interpretation of "increased influences." Suppose that the CGCSs, for whatever reason, are indeed trying to influence their colleagues, and they do so to a larger extent in the masked scheme.

In the first placebo test, we hypothesize that under the masked scheme, the "additional influence activities toward colleagues" will result in more interactions between the CGCSs and their colleagues, especially in non-professional occasions. However, in Panel A of Table 9, we

find that according to the colleagues, "masking" did not make the CGCSs communicate more frequently with them (Columns (1) and (2)), nor did it make them become more familiar with the CGCS's work or life (Columns (3) and (4)). These results are inconsistent with the alternative interpretation.

In the second placebo test, we assume that it is easier for the CGCS to influence someone from the same hometown. Under this assumption, if the CGCSs are indeed influencing colleagues more in the masked scheme, we should observe "same hometown colleagues" being more positive about CGCS performance under the masked scheme than the revealed scheme. As shown in Appendix Table A11, this is not the case: the "same hometown colleague" coefficient is almost identical for the two evaluation schemes, again suggesting that the CGCSs did not impose additional influence activities toward colleagues under the masked scheme.

2. Higher Information Quality in Masked Scheme

Another possibility is that the supervisors in the masked scheme tend to get better information on CGCS performance, which might explain the improved supervisor assessments.

To address this concern, in Panel B of Table 9, we examine whether the supervisors get additional information on CGCS performance under the masked scheme, either from colleagues or the CGCS himself. We find that, for both the evaluating supervisor and the non-evaluating supervisor, being in the masked scheme does not increase the frequency of CGCSs and other colleagues reporting to them regarding CGCS performance. This suggests that the improved supervisor assessments in the masked scheme cannot be explained by changes in information quality.

VI. Conclusion

Subjective evaluations are widely used in both public and private sectors, especially in contexts where job tasks are inherently multi-dimensional and vaguely defined, making it impossible to obtain sharp measures of employee efforts and performances. A key limitation of subjective evaluation is that it may distort employees' incentives and make them more likely to cater to the personal preferences of the evaluator, rather than focusing on productive tasks that benefit the whole organization. However, rigorous empirical evidence on the existence and implications of influence activities is rare.

To shed light on these issues, we conduct a large-scale field experiment, where we randomize two subjective performance evaluation schemes among 3,785 junior state employees in China. In the first "revealed scheme," we randomly choose one supervisor (from two) as the performance evaluator, and inform the subordinate about the identity of the evaluator *ex ante*. We find that in this scheme, the subordinates are indeed able to impose evaluator-specific influence activities to improve evaluation outcomes. Furthermore, other colleagues could correctly predict that the randomly selected evaluator would favor the subordinate (as compared to the non-evaluating supervisor), which indicates that the subordinate's supervisor-specific influence activities could be partly observed by their fellow co-workers.

In the second "masked scheme," the identity of the evaluator is not disclosed to the subordinate, which makes supervisor-specific influences less beneficial and riskier. Therefore, masking evaluator identity should encourage the subordinate to reallocate his efforts from influence activities toward the common productive dimensions that could be appreciated by both supervisors. We find that this intervention indeed improves the work performance of the subordinate, as measured average colleague assessment, average supervisor assessment, likelihood to be recommended for "tenure," as well as actual monthly bonuses received which are determined objective performance indicators.

In addition to providing the first rigorous empirical evidence on the existence and implications of influence activities, our findings also have important policy implications. We find that by randomizing the identity of the evaluating supervisor, which has minimal implementation cost, the government could significantly improve the job performance of its employees. Our findings have direct policy implications for the more than 50 million state employees in China, and might also apply to many other contexts where high-stakes rewards are linked to the subjective opinions of certain evaluators.

REFERENCES

Ashraf, Nava, Oriana Bandiera, and B. Kelsey Jack. "No margin, no mission? A field experiment on incentives for public service delivery." *Journal of Public Economics* 120 (2014): 1-17.

Baker, George, Robert Gibbons, and Kevin J. Murphy. "Subjective performance measures in optimal incentive contracts." *The Quarterly Journal of Economics* 109, no. 4 (1994): 1125-1156.

Banerjee, Abhijit V., Raghabendra Chattopadhyay, Esther Duflo, Daniel Keniston, and Nina Singh. "Can institutions be reformed from within? Evidence from a randomized experiment with the Rajasthan police." (2012).

Bertrand, Marianne, Robin Burgess, Arunish Chawla, and Guo Xu. "The glittering prizes: Career incentives and bureaucrat performance." *The Review of Economic Studies* (2018).

Chevalier, Judith, and Glenn Ellison. "Career concerns of mutual fund managers." *The Quarterly Journal of Economics*114, no. 2 (1999): 389-432.

Deb, Joyee, Jin Li, and Arijit Mukherjee. "Relational contracts with subjective peer evaluations." *The RAND Journal of Economics* 47, no. 1 (2016): 3-28.

Finan, Frederico, Benjamin A. Olken, and Rohini Pande. *The personnel economics of the state*. No. w21825. National Bureau of Economic Research, 2015.

Gibbons, Robert, and Kevin J. Murphy. "Optimal incentive contracts in the presence of career concerns: Theory and evidence." *Journal of political Economy* 100, no. 3 (1992): 468-505.

Guo, Gang. "China's local political budget cycles." *American Journal of Political Science* 53, no. 3 (2009): 621-632.

Hayes, Rachel M., and Scott Schaefer. "Implicit contracts and the explanatory power of top executive compensation for future performance." *The RAND Journal of Economics* (2000): 273-293.

He, Guojun, Shaoda Wang, and Bing Zhang. "Environmental Regulation and Firm Productivity in China: Estimates from a Regression Discontinuity Design." (2018).

He, Guojun, and Shaoda Wang. "Do college graduates serving as village officials help rural China?" *American Economic Journal: Applied Economics* 9, no. 4 (2017): 186-215.

Hetts, John J., Michiko Sakuma, and Brett W. Pelham. "Two roads to positive regard: Implicit and explicit self-evaluation and culture." *Journal of Experimental Social Psychology* 35, no. 6 (1999): 512-559.

Lazear, Edward, and Paul Oyer. *The handbook of organizational economics*. Gibbons, Robert, and John Roberts, eds. Princeton University Press, 2012.

Li, Hongbin, and Li-An Zhou. "Political turnover and economic performance: the incentive role of personnel control in China." *Journal of public economics* 89, no. 9-10 (2005): 1743-1762.

MacLeod, W. Bentley. "Optimal contracting with subjective evaluation." *American Economic Review* 93, no. 1 (2003): 216-240.

Meyer, Margaret, Paul Milgrom, and John Roberts. "Organizational prospects, influence costs, and ownership changes." *Journal of Economics & Management Strategy* 1, no. 1 (1992): 9-35.

Milgrom, Paul R. "Employment contracts, influence activities, and efficient organization design." *Journal of political economy* 96, no. 1 (1988): 42-60.

Milgrom, Paul, and John Roberts. "An economic approach to influence activities in organizations." *American Journal of sociology* 94 (1988): S154-S179.

Olken, Benjamin A., Rohini Pande, and Raluca Dragusanu. "Governance review paper: J-PAL governance initiative." *Rep., Abdul Latif Jameel Poverty Action Lab, Cambridge, MA*(2011).

Oyer, P., and S. Schaefer. "Personnel Economics: Hiring and Incentives." *Handbook of Labor Economics*. (2011): 1769-1823.

Prendergast, Canice. "The provision of incentives in firms." *Journal of economic literature* 37, no. 1 (1999): 7-63.

Schaefer, Scott. "The Dependence of pay—Performance Sensitivity on the Size of the Firm." *Review of Economics and Statistics* 80, no. 3 (1998): 436-443.

Shirk, Susan L. *The political logic of economic reform in China*. Vol. 24. Univ of California Press, 1993.

Suárez Serrato, Juan Carlos, Xiao Yu Wang, and Shuang Zhang. *The limits of meritocracy: screening bureaucrats under imperfect verifiability*. No. w21963. National Bureau of Economic Research, 2016.

Figure1. Experimental Design

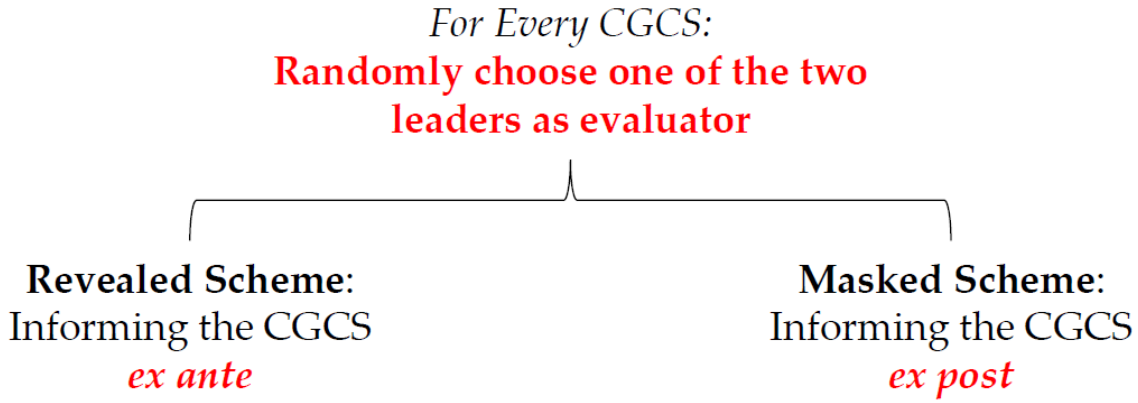


Table 1. Revealing Evaluator's Identity Causes Scoring Asymmetry

	Supervisor 1 Score Minus Supervisor 2 Score		Colleagues Speculate: Supervisor 1 more Positive	
	(1)	(2)	(3)	(4)
Supervisor 1 Evaluating (Revealed)	0.310***	0.310***	0.058***	0.057***
	(0.082)	(0.082)	(0.020)	(0.020)
Outcome Mean	-0.020	-0.020	0.564	0.564
Outcome S.D.	1.300	1.300	0.496	0.496
Controls	N	Y	N	Y
County FE	Y	Y	Y	Y
Type FE	Y	Y	Y	Y
Enrol Year FE	Y	Y	Y	Y
Obs.	1,301	1,301	5,582	5,582
R-Squared	0.160	0.162	0.108	0.110

Notes: Each column represents a separate OLS regression. Standard errors clustered at work unit level are reported below the coefficients. Control variables include CGCS and colleague characteristics listed in Appendix Tables A1-A2. * significant at 10% ** significant at 5% *** significant at 1%.

Table 2. Masking Evaluator's Identity Eliminates Scoring Asymmetry

	Supervisor 1 Score Minus Supervisor 2 Score		Colleagues Speculate: Supervisor 1 more Positive	
	(1)	(2)	(3)	(4)
Supervisor 1 Evaluating (Masked)	-0.003	0.014	-0.012	-0.010
	(0.113)	(0.113)	(0.018)	(0.018)
Outcome Mean	-0.005	-0.005	0.523	0.523
Outcome S.D.	1.217	1.217	0.500	0.500
Controls	N	Y	N	Y
County FE	Y	Y	Y	Y
Type FE	Y	Y	Y	Y
Enrol Year FE	Y	Y	Y	Y
Obs.	580	580	2,615	2,615
R-Squared	0.242	0.253	0.175	0.182

Notes: Each column represents a separate OLS regression. Standard errors clustered at work unit level are reported below the coefficients. Control variables include CGCS and colleague characteristics listed in Appendix Tables A1-A2. * significant at 10% ** significant at 5% *** significant at 1%.

Table 3. Treatment Effects on Colleague Assessments

	Performance (1-7)		Top 10%	
	(1)	(2)	(3)	(4)
Masking	0.220***	0.185***	0.081***	0.070***
	(0.033)	(0.027)	(0.012)	(0.011)
Colleague Controls	N	Y	N	Y
County FE	Y	Y	Y	Y
Type FE	Y	Y	Y	Y
Enrol Year FE	Y	Y	Y	Y
Obs.	9,256	9,167	9,256	9,167
R-Squared	0.130	0.371	0.084	0.267

Notes: Each column represents a separate OLS regression. Standard errors clustered at work unit level are reported below the coefficients. Control variables include colleagues' personal characteristics listed in Table A2. * significant at 10% ** significant at 5% *** significant at 1%.

Table 4. Treatment Effects on Supervisor Assessments

	Mean (Supervisor Assessment)	Evaluator Assessment	Non-Evaluator Assessment	Supervisors' Assessment Deviation
	(1)	(2)	(3)	(4)
Masking	0.151***	0.080*	0.214***	-0.085**
	(0.042)	(0.047)	(0.053)	(0.040)
Controls	Y	Y	Y	Y
County FE	Y	Y	Y	Y
Type FE	Y	Y	Y	Y
Enrol Year FE	Y	Y	Y	Y
Obs.	2,643	2,292	2,290	2,643
R-Squared	0.208	0.189	0.199	0.105

Notes: Each column represents a separate OLS regression. Control variables include CGCS characteristics listed in Appendix Table A1.Standard errors clustered at work unit level are reported below the coefficients. * significant at 10% ** significant at 5% *** significant at 1%.

Table 5. CGCSs Work Harder and Become More Willing to Leave

	(1)	(2)	(3)	(4)
Panel A. CGCSs Work Harder				
	Work Harder and Overtime		Self Assessment (1-7)	
Masking	0.023**	0.023**	0.081*	0.080*
	(0.012)	(0.012)	(0.048)	(0.048)
Obs.	9,349	9,349	2,771	2,771
R-Squared	0.491	0.491	0.117	0.125
Panel B. CGCSs' Welfare Revealed by Reservation Wage				
	Willingness to Accept Private Sector Job		Reservation Wage (log)	
Masking	0.028*	0.029*	-0.054*	-0.054*
	(0.016)	(0.016)	(0.029)	(0.029)
Obs.	2,737	2,737	2,738	2,738
R-Squared	0.233	0.234	0.235	0.235
Controls	N	Y	N	Y
County FE	Y	Y	Y	Y
Type FE	Y	Y	Y	Y
Enrol Year FE	Y	Y	Y	Y

Notes: Each column represents a separate OLS regression. Control variables include CGCS characteristics listed in Appendix Table A1. Standard errors clustered at work unit level are reported below the coefficients. * significant at 10% ** significant at 5% *** significant at 1%.

Table 6. Treatment Effects on "Revealed Preference" Measures

	Qualify for Tenure	Wage (log)	Wage	Wage (Medical Support)
	(1)	(2)	(3)	(4)
Masking	0.032***	0.020**	48.81**	115.54*
	(0.009)	(0.008)	(22.41)	(61.94)
County FE	Y	Y	Y	Y
Type FE	Y	Y	Y	Y
Enroll Year FE	Y	Y	Y	Y
Obs.	9,349	2,750	2,750	193
R-Square	0.099	0.665	0.64	0.74

Notes: Each column represents a separate OLS regression. Control variables include CGCSs' personal characteristics listed in Table 1. Standard errors clustered at work unit level are reported below the coefficients. * significant at 10% ** significant at 5% *** significant at 1%.

Table 7. Mechanism: Risk Aversion and Relative Importance of Supervisors

	Colleague Evaluation (1-7)			
	(1)	(2)	(3)	(4)
Risk Aversion	-0.044			
	(0.031)			
Masking*Risk Aversion	0.105**			
	(0.053)			
Supervisors' Weights Similarity		-0.088		
		(0.069)		
Masking*Supervisors' Weights Similarity		0.212*		
		(0.112)		
Supervisors' Info. Gap			0.195*	
			(0.101)	
Masking* Supervisors' Info. Gap			-0.365*	
			(0.189)	
Δ in Supervisors' Work Assign. Freq.				0.003***
				(0.001)
Masking*Δ in Work Assign. Freq.				-0.002***
				(0.001)
Masking	0.140***	0.173***	0.526***	0.208***
	(0.037)	(0.030)	(0.176)	(0.029)
Controls	Y	Y	Y	Y
County FE	Y	Y	Y	Y
Type FE	Y	Y	Y	Y
Enroll Year FE	Y	Y	Y	Y
Obs.	9,221	8,770	8,788	9,243
R-Squared	0.355	0.397	0.361	0.335

Notes: Each column represents a separate OLS regression. Control variables include colleagues' characteristics listed in Table A2. Standard errors clustered at work unit level are reported below the coefficients. * significant at 10% ** significant at 5% *** significant at 1%.

Table 8. Ruling Out Alternative Explanations to Influence Activities

	(1)	(2)	(3)
Panel A. Does Supervisor Evaluation Depend on their Awareness of their Roles?			
	Supervisor 1 Score Minus Supervisor 2 Score		
Supervisor 1 Eva. (*ex ante)*	0.310***	0.334***	0.320*
	(0.082)	(0.099)	(0.166)
Sample	Full Sample	Supervisor 1 Unaware of being the Evaluator	Supervisor 1 Aware of Being the Evaluator
Obs.	1,301	888	333
R-Squared	0.160	0.206	0.270
Panel B. Behavioral Changes of the Evaluating Supervisor?			
	Supervisor 1 Not Responding to the Survey	Sup.1 Writes More Words in Describing CGCS's Job	Sup. 1 Assigns More Tasks to the CGCS
Supervisor 1 Eva. (*ex ante)*	-0.010	0.649	0.236
	(0.019)	(0.431)	(0.181)
Obs.	1,910	1,910	1,910
R-Squared	0.144	0.147	0.144
Panel C. Does the Evaluating Supervisor Receive More Information?			
	Supervisor 1 Gets More Information from CGCS than Supervisor 2 Does	Supervisor 1 Gets More Information from Colleagues than Supervisor 2 Does	Supervisor 1 Gets More Information from Opposing Supervisor than Supervisor 2 Does
Supervisor 1 Eva. (*ex ante)*	-0.007	-0.009	0.022
	(0.016)	(0.018)	(0.020)
Obs.	1,910	1,910	1,910
R-Squared	0.121	0.143	0.158
County FE	Y	Y	Y
Type FE	Y	Y	Y
Enrol Year FE	Y	Y	Y

Notes: Each column represents a separate OLS regression. Standard errors clustered at work unit level are reported below the coefficients.* significant at 10% ** significant at 5% *** significant at 1%.

Table 9. Alternative Explanations to the Treatment Effects of Uncertainty

	(1)	(2)	(3)	(4)

Panel A. Do CGCGs Influence All Their Colleagues?

	Communication with Colleagues	Meeting with Colleagues	Colleagues Familiar with CGCS Work	Colleagues Familiar with CGCS Life
Masking	-0.008	0.013	0.020	0.066
	(0.013)	(0.020)	(0.034)	(0.059)
Obs.	9,272	9,349	9,252	9,244
R-Squared	0.055	0.066	0.066	0.083

Panel B. Masking the Identity of the Supervisor Does Not Lead to Information Difference

	Information Diff. between the Evaluating Supervisors in T1 and T2		Information Diff. between the Non-Evaluating Supervisors in T1 and T2	
Masking	0.022	-0.006	-0.004	-0.026
	(0.026)	(0.028)	(0.024)	(0.025)
Information from	CGCSs	Colleagues	CGCSs	Colleagues
Obs.	1,367	1,367	1,417	1,417
R-Squared	0.175	0.173	0.176	0.188
County FE	Y	Y	Y	Y
Type FE	Y	Y	Y	Y
Enrol Year FE	Y	Y	Y	Y

Notes: Each column represents a separate OLS regression. Standard errors clustered at work unit level are reported below the coefficients. * significant at 10% ** significant at 5% *** significant at 1%.

Appendix to "Influence Activities and Bureaucratic Performance: Evidence from a Large-Scale Field Experiment in China"

Table A1. Characteristics of CGCSs and Balance Checks

	Mean (Std. Dev.)	Difference between T1 and T2
	(1)	(2)
Age	25.01	0.04
	(1.56)	(0.06)
Gender	0.62	0.01
(=1 if Female)	(0.49)	(0.02)
Year of Enrollment	2016.6	-0.02
	(0.5)	(0.02)
Social Science Major	0.54	-0.02
(=1 if Yes)	(0.50)	(0.02)
4-Year College or Above	0.76	-0.00
(=1 if Yes)	(0.43)	(0.02)
STEM Students in High School	0.35	-0.01
(=1 if Yes)	(0.48)	(0.02)
Party Member	0.22	-0.00
(=1 if Yes)	(0.41)	(0.02)
Parent Completing High School	0.57	0.03*
(=1 if Yes)	(0.50)	(0.02)
Parent Completing College	0.29	-0.00
(=1 if Yes)	(0.45)	(0.02)
Work in Village	0.15	-0.01
(=1 if Yes)	(0.36)	(0.02)
CEE Score	483.30	5.72
(Points)	(73.43)	(3.55)
Risk Averse	0.47	-0.00
(=1 if Yes)	(0.50)	(0.02)
Obs.		2,839

Notes: Column (1) summarizes the mean and standard deviation of CGCS characteristics. Column (2) checks the covariate balances between the T1 and T2 groups. In T1 group, the identity of the randomly-chosen evaluator is disclosed to the CGCS ex ante, while in T2 group, the evaluator is chosen randomly ex-post. Standard errors clustered at work unit level are reported in the parentheses. Data are collected by the research team.

Table A2. Characteristics of CGCSs' Colleagues and Balance Checks

	Mean (Std. Dev.)	Difference between T1 and T2
	(1)	(2)
Colleague Age	34.50	-0.28
	(8.92)	(0.26)
Colleague Gender	0.57	-0.01
(=1 if Female)	(0.50)	(0.01)
Colleague Education	3.46	-0.02
	(0.71)	(0.02)
Colleague Tenured	0.74	0.00
	(0.44)	(0.01)
Meet Frequency with CGCS	4.75	0.01
Weekly	(0.72)	(0.02)
Know CGCS Well (Work)	9.28	0.02
(0-10)	(1.25)	(0.03)
Know CGCS Well (Life)	8.33	0.07
(0-10)	(2.03)	(0.06)
Colleague Self-Evaluation	4.46	0.05*
(1-7)	(1.21)	(0.03)
Obs.		9,349

Notes: Column (1) summarizes the mean and standard deviation of colleagues' characteristics. Column (2) checks the covariate balances between the T1 and T2 groups. In T1 group, the identity of the randomly-chosen evaluator is disclosed to the CGCS ex ante, while in T2 group, the evaluator is chosen randomly ex-post. Education is measured by a categorical variable (primary school =1, junior high =2, senior high=3, 3-year college =4, 4-year college =5, graduate school=6). Standard errors clustered at work unit level are reported in the parentheses. Data are collected by the research team.

Table A3. Characteristics of Supervisors and Balance Checks

	Mean (Std. Dev.)	Difference between T1 and T2
	(1)	(2)
Supervisor 1 Gender	0.17	-0.01
(=1 if Female)	(0.38)	(0.02)
Supervisor 1 Age	44.92	0.19
	(6.94)	(0.33)
Supervisor 1 Work Experience	7.0	0.03
(Years)	(3.4)	(0.16)
Supervisor 1 Education	4.72	-0.00
	(0.57)	(0.03)
Supervisor 1 Duty	0.54	-0.03
(=1 if Party, =2 if Admin)	(0.50)	(0.02)
Supervisor 2 Gender	0.27	-0.00
(=1 if Female)	(0.44)	(0.02)
Supervisor 2 Age	42.60	-0.59*
	(7.59)	(0.36)
Supervisor 2 Work Experience	6.89	0.02
(Years)	(3.46)	(0.17)
Supervisor 2 Education	4.64	0.02
	(0.61)	(0.03)
Supervisor 2 Duty	0.59	-0.02
(=1 if Party, =2 if Admin)	(0.49)	(0.02)
Obs.		2,249

Notes: Column (1) summarizes the mean and standard deviation of supervisors' characteristics. Column (2) checks the covariate balances between the T1 and T2 groups. In T1 group, the identity of the randomly-chosen evaluator is disclosed to the CGCS ex ante, while in T group, the evaluator is chosen randomly ex-post. Education is measured by a categorical variable (primary school =1, junior high =2, senior high=3, 3-year college =4, 4-year college =5, graduate school=6). Standard errors clustered at work unit level are reported in the parentheses. Data are collected by the research team.

Table A4. Test for Attrition Selection

	Attrition			
	(1)	(2)	(3)	(4)
Masking Evaluator's Identity	-0.015	-0.010	-0.010	-0.010
	(0.016)	(0.015)	(0.015)	(0.015)
County FE	N	Y	Y	Y
Type FE	N	N	Y	Y
Enroll Year FE	N	N	N	Y
Obs.	3,785	3,773	3,773	3,773
R-Squared	0.000	0.141	0.145	0.145

Notes: Each column represents a separate OLS regression. Standard errors clustered at county level are reported below the coefficients. * significant at 10% ** significant at 5% *** significant at 1%.

Table A5. Correlations between CGCS Characteristics and Assessment Scores

	Performance (1-7)	
	Colleague	Supervisor
	(1)	(2)
Age	0.049***	0.058***
	(0.011)	(0.015)
Gender	-0.069*	-0.055
	(0.038)	(0.040)
Social Science	-0.005	-0.004
	(0.038)	(0.040)
4-Year College	0.153***	0.129***
	(0.048)	(0.048)
STEM Students	-0.065	-0.009
	(0.041)	(0.042)
Party Member	0.173***	0.143***
	(0.041)	(0.051)
Parent High Sch.	0.033	0.044
	(0.043)	(0.047)
Parent College	-0.030	0.096*
	(0.047)	(0.051)
Work in Village	0.004	0.090*
	(0.060)	(0.051)
CEE Score	-0.038	0.005
	(0.032)	(0.035)
Risk Averse	-0.009	-0.009
	(0.030)	(0.040)
Obs.	8,871	2,556
R-Squared	0.017	0.032

Notes: Each column represents a separate regression. Column (1) reports the correlation between CGCSs performances evaluated by their colleagues and the CGCSs' personal characteristics. Column (2) reports the correlation between CGCSs' performances evaluated by their supervisors and the CGCSs' personal characteristics. Column report the correlation between CGCS's performances evaluated by their colleagues and the colleagues personal characteristics. Standard errors clustered at work unit level are reported in the parentheses. * significant at 10% ** significant at 5% *** significant at 1%.

Table A6. Treatment Effects for Different Cohorts

	Performance (1-7)		Top 10%	
	(1)	(2)	(3)	(4)
Panel A. 2016 Cohort				
Masking	0.244***	0.189***	0.082***	0.071***
	(0.052)	(0.043)	(0.018)	(0.017)
Obs.	4,158	4,124	4,158	4,124
R-Square	0.146	0.374	0.107	0.275
Panel B. 2017 Cohort				
Masking	0.217***	0.190***	0.088***	0.077***
	(0.044)	(0.036)	(0.017)	(0.015)
Obs.	5,095	5,035	5,095	5,035
R-Square	0.158	0.406	0.105	0.300
Controls	N	Y	N	Y
County FE	Y	Y	Y	Y
Type FE	Y	Y	Y	Y

Notes: Each column represents a separate OLS regression. Control variables include CGCS and colleagues' characteristics listed in Tables A1 and A2. Standard errors clustered at work unit level are reported below the coefficients. * significant at 10% ** significant at 5% *** significant at 1%.

Table A7. Treatment Effects for Different Provinces

	Performance (1-7)		Top 10%	
	(1)	(2)	(3)	(4)
Panel A. Richer Province				
Masking	0.196***	0.152***	0.064***	0.050***
	(0.054)	(0.045)	(0.016)	(0.015)
Obs.	3,401	3,377	3,401	3,377
R-Square	0.124	0.346	0.092	0.256
Panel B. Poorer Province				
Masking	0.230***	0.190***	0.090***	0.079***
	(0.041)	(0.033)	(0.017)	(0.014)
Obs.	5,855	5,780	5,855	5,780
R-Square	0.053	0.343	0.046	0.262
Controls	N	Y	N	Y
County FE	Y	Y	Y	Y
Type FE	Y	Y	Y	Y

Notes: Each column represents a separate OLS regression. Control variables include CGCS and colleagues' characteristics listed in Tables A1 and A2. Standard errors clustered at work unit level are reported below the coefficients. * significant at 10% ** significant at 5% *** significant at 1%.

Table A8. Heterogeneity w.r.t. CGCSs' Characteristics

	Performance (1-7)						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Masking*X	0.089*	-0.008	-0.030	-0.040	-0.059	-0.052	0.130*
	(0.052)	(0.052)	(0.057)	(0.056)	(0.062)	(0.059)	(0.077)
Individual Characteristics X	0.009	0.001	0.071**	0.017	0.081**	-0.047	0.005
	(0.032)	(0.030)	(0.036)	(0.033)	(0.037)	(0.033)	(0.046)
Masking	0.129***	0.189***	0.207***	0.198***	0.198***	0.199***	0.167***
	(0.042)	(0.039)	(0.048)	(0.033)	(0.030)	(0.031)	(0.029)
Individual Characteristics (X)	Female	Social Science Major	4-Year College or Above	STEM Students in High School	Party Member	Parent Completed College	Work in Village
Controls	Y	Y	Y	Y	Y	Y	Y
County FE	Y	Y	Y	Y	Y	Y	Y
Type FE	Y	Y	Y	Y	Y	Y	Y
Obs.	9,167	9,167	9,167	9,167	9,167	9,167	9,167
R-Square	0.371	0.370	0.370	0.370	0.371	0.371	0.370

Notes: Each column represents a separate OLS regression. Control variables include CGCS and colleagues' characteristics listed in Tables A1 and A2. Standard errors clustered at work unit level are reported below the coefficients. * significant at 10% ** significant at 5% *** significant at 1%.

Table A9. Heterogeneity w.r.t. CGCS's Motivation

	Performance (1-7)					
	(1)	(2)	(3)	(4)	(5)	(6)
Masking*X	0.033*	0.042**	0.003	-0.013	0.017	-0.000
	(0.018)	(0.017)	(0.028)	(0.027)	(0.017)	(0.020)
Motivation X	0.007	-0.007	0.014	0.036**	0.005	-0.030***
	(0.010)	(0.010)	(0.016)	(0.015)	(0.010)	(0.011)
Masking	0.075	0.053	0.174	0.241**	0.141***	0.183***
	(0.066)	(0.061)	(0.126)	(0.121)	(0.053)	(0.047)
Motivation (X)	Want to Work Close to Family	Want Bonus Points for Civil Exam and Graduate School	Want to Work in the Government	Want to Serve the Grassroots and Gain Experience	Want to Get Entrepreneurship Assistance	No Other Choices
Controls	Y	Y	Y	Y	Y	Y
County FE	Y	Y	Y	Y	Y	Y
Type FE	Y	Y	Y	Y	Y	Y
Obs.	9,167	9,167	9,2167	9,167	9,167	9,167
R-Square	0.372	0.371	0.371	0.371	0.371	0.372

Notes: Each column represents a separate OLS regression. Control variables include CGCS and colleagues' characteristics listed in Tables A1 and A2. Standard errors clustered at work unit level are reported below the coefficients. * significant at 10% ** significant at 5% *** significant at 1%.

Table A10. Heterogeneity w.r.t. CGCS's Attitudes

	Performance (1-7)						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Masking*X	-0.076**	-0.119**	-0.062**	0.045	0.034	0.047*	0.016
	(0.030)	(0.048)	(0.026)	(0.029)	(0.027)	(0.027)	(0.025)
Attitude X	0.045**	0.055**	0.007	-0.039**	-0.012	-0.032**	-0.018
	(0.019)	(0.025)	(0.015)	(0.017)	(0.015)	(0.015)	(0.014)
Masking	0.413***	0.620***	0.322***	0.136***	0.152***	0.126***	0.170***
	(0.093)	(0.180)	(0.065)	(0.043)	(0.037)	(0.042)	(0.036)
Attitude (X)	Satisfied with the Work Performance	Always Try My Best to Finish the Work	Have Fast Pace of Life	Interested in Public Policy	No Interest in Being a Leader	Often Feel Anxious	Sometimes Feel Myself Worthless
Controls	Y	Y	Y	Y	Y	Y	Y
County FE	Y	Y	Y	Y	Y	Y	Y
Type FE	Y	Y	Y	Y	Y	Y	Y
Obs.	9,167	9,167	9,167	9,167	9,167	9,167	9,167
R-Square	0.131	0.372	0.372	0.371	0.371	0.371	0.371

Notes: Each column represents a separate OLS regression. Control variables include CGCS and colleagues' characteristics listed in Tables A1 and A2. Standard errors clustered at work unit level are reported below the coefficients. * significant at 10% ** significant at 5% *** significant at 1%.

Table A11. Revealing Evaluator's Identity Causes Scoring Asymmetry

	Colleague Assessment Score		
	Full Sample	Revealed Sample	Masked Sample
	(1)	(2)	(3)
Colleague from Same Hometown	0.051*	0.054	0.056
	(0.028)	(0.034)	(0.051)
County FE	Y	Y	Y
Type FE	Y	Y	Y
Enrol Year FE	Y	Y	Y
Obs.	9,252	6,286	2,954
R-Squared	0.326	0.350	0.340

Notes: * significant at 10% ** significant at 5% *** significant at 1%.