# Subjective Performance Evaluation, Influence Activities, and Bureaucratic Work Behavior: Evidence from China

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

Shaoda Wang†, Qiong Zhang▪

*[August 2021]*

Subjective performance evaluation is widely used by firms and governments to provide work incentives. However, delegating evaluation power to local leadership could induce influence activities: employees might devote too much effort to impressing or pleasing their supervisors, relative to working toward the goals of the organization. In this paper, we conduct a large-scale randomized field experiment among Chinese local civil servants and provide the first rigorous empirical evidence on the existence and implications of influence activities. We find that civil servants do engage in evaluator-specific influence to affect evaluation outcomes. These activities involve both reallocating work efforts toward job tasks that are more important and observable to the evaluator and delivering personal favors and flattery outside of official mandates. Importantly, we find that introducing uncertainty in the identity of the evaluator discourages evaluator-specific influence activities and significantly improves the work performance of local civil servants.

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

JEL: M12; D73; F63

* University of California, Berkeley. Email: alain@berkeley.edu.
• University of Hong Kong. Email: gjhe@hku.hk.
‡ University of California, Berkeley. Email: esadoulet@berkeley.edu.
† University of Chicago. Email: shaoda@uchicago.edu.
▪ Renmin University of China. Email: zhangqiong8@ruc.edu.cn.

## I. Introduction

For a large share of jobs in modern economies, 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 measuring individual achievements and the multiplicity of tasks for most civil service jobs (Olken and Pande, 2013; Finan et al., 2015).

While subjective performance measures might 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 assessment in their favor, which could potentially be detrimental to the interests of the organization (Milgrom and Roberts, 1988; Milgrom, 1988). Specifically, as noted by Milgrom and Roberts (1988), influence activities can be categorized into two types: productive activities, such as putting extra effort into tasks that are more visible to the evaluator; and non-productive activities, such as "buttering up" the evaluator with personal favors. While a rich theoretical literature has investigated the formation and consequences of influence activities, these theoretical arguments have rarely been confronted with rigorous empirical analyses (Oyer and Schaefer, 2011; Lazear and Oyer, 2012).

Empirically studying influence activities is challenging for at least three reasons. First, spending extra effort on tasks that are more visible to the evaluator, or trying to personally benefit the evaluator, can be regarded unfavorably by others, which means that the agent might try to hide such behaviors. Second, even if such behaviors are observed, it is difficult to infer that they are driven by intentions of improving evaluation outcomes (instead of simply being hardworking or friendly), making it difficult to classify them exclusively as influence activities. Third, even if the existence of influence activities is established, quantifying their effects on work performance still requires exogenous variation in behavior across agents.

In this paper, we conduct a large-scale field experiment in two Chinese provinces, which aims at addressing these three challenges and providing the first rigorous empirical evidence on the existence and consequences of influence activities in the workplace. Our experiment focuses on China's "3+1 Supports" program, a large national "human capital reallocation" initiative that hires more than 30,000 college graduates annually to work in rural township governments on two-year contracts. These junior government employees are referred to in this paper as College Graduate Civil Servants (CGCSs).

A distinct institutional feature of the Chinese governance system is its dual-leadership arrangement (Shirk, 1993), whereby every government organization/subsidiary has two leaders: a "party leader" (i.e., party secretaries at various levels) and an "administrative leader" (i.e., the head in a village, the mayor in a city).[1] As a result of this dual system, every CGCS reports to two supervisors who both assign her job tasks and provide performance feedback on a regular basis. Under the *status quo,* every CGCS is evaluated by one of her two supervisors every year.[2] The evaluation outcome will determine whether the CGCS can be awarded a permanent contract upon completing her two-year term, a highly sought-after outcome for most CGCSs due to the prestige of permanent civil service jobs in China. Under the current arrangement, rich anecdotal evidence suggests that many CGCSs exert substantial efforts trying to please their specific evaluating supervisor, in both productive and non-productive ways, in the expectation of better evaluation outcomes. The prevalence of influence activities is concerning for the government, which worries that such efforts to please specific evaluators might crowd out their efforts on the productive tasks deemed important by the organization.

To examine the existence of influence activities in this environment and understand the impacts of such activities on CGCS work behavior, we collaborated with two provincial governments in China and randomized two performance evaluation schemes among their 3,785 CGCSs working in more than 200 counties. In both schemes, we randomly selected one of the two supervisors to be the evaluator. The only difference is that, in the "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 her promotion. In the "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 as having a 50% chance of determining her promotion. In both arms, we do not inform the supervisors about who will be chosen as the evaluator.

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. This result is consistent with a scenario where the agent is able to engage in evaluator-specific influence activities – either productive or non-productive – to improve evaluation outcomes. In comparison, we find no such asymmetry in supervisor assessments in the masked scheme. Exploiting the random assignment

[1] The two leaders have large overlaps in their responsibilities, introducing *de facto* checks and balances in employee supervision. See Li (2018) for information on the institutional details of the dual system.

[2] Sixty-two percent of the CGCSs are females, so we use the female pronouns (she/her/hers) for the CGCS throughout this paper. In contrast, the majority of the CGCSs' supervisors and colleagues are males, so we use the male pronouns (he/him/his) when referring to them throughout this paper.

of the two evaluation schemes, we find that masking the evaluator's identity incentivizes the CGCSs to reallocate their efforts from evaluator-specific influence activities to common productive tasks that are valued by both supervisors, and can thus significantly improve CGCS work performance according to a series of different measures.

To better understand the underlying mechanisms, we examine the prevalence of both productive and non-productive influence activities as described by Milgrom and Roberts (1988). The former can be viewed as a special form of multi-tasking, where the agent needs to complete different tasks assigned by different supervisors and chooses to work harder on the dimensions that are better observed/appreciated by the evaluating supervisor. Such productive influence activities do not necessarily harm the organization. In contrast, non-productive influence activities only benefit the evaluator personally but not the organization as a whole, and are generally deemed undesirable by the organization.

We provide evidence that both types of influence activities are prevalent in the revealed scheme. First, we find that, under the revealed scheme, the CGCS devotes more efforts to the job tasks assigned by her evaluator, and her work performance improves more in areas that are valued highly by the evaluator. These results suggest the existence of productive influence activities. Second, by eliciting individual beliefs, we find that CGCSs in the revealed scheme are more likely to think that "handling the personal relationship with the supervisor is a very challenging part of the job", while CGCSs in the masked scheme are more likely to think "civil service is meritocratic and hard work pays off." These results are consistent with the existence of non-productive influence activities.

We conduct a battery of additional tests to rule out alternative interpretations of our findings. For example, we find that the assessment asymmetry under the revealed scheme is not driven by any behavioral change of the evaluator, nor by any additional information about CGCS performance being presented to him. We also find that the improved performance under the masked scheme cannot be explained by the CGCS engaging in even more influence activities toward her two supervisors and other colleagues. Taken together, the empirical results consistently support our proposed mechanism – the CGCS's differential efforts at impressing the evaluator instead of the non-evaluator – over other explanations.

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 the workplace. As pointed out by Lazear and Oyer (2012), while a large theoretical literature has studied how agents try to engage in 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 is a lack of rigorous empirical evidence, aside from anecdotes and case studies, to verify these arguments.[3] Our paper fills this gap by providing field experimental evidence on influence activities among Chinese local government employees, as well as quantifying the causal impact of reducing influence activities on job performance.[4] More broadly, subjective performance evaluation is ubiquitous across both the private and public sectors and has been investigated extensively by a large body of theoretical work (Gibbons and Murphy, 1992; Baker et al., 1994; Prendergast and Topel, 1996; MacLeod, 2003, Maestri, 2012; Deb et al., 2016). 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 can undermine the effectiveness of subjective performance evaluations.

Second, this paper adds to a growing experimental literature on the personnel economics of the developing state, specifically on incentivizing public employees (Finan et al., 2016). Most of the existing field experiments on this topic focus on the role of financial incentives,[5] with only a few exceptions studying non-pecuniary incentives, such as transfers and postings (Banerjee et al., 2012), social incentives (Ashraf and Bandiera, 2018), and intrinsic motivation (Ashraf at al., 2014). Our paper adds to this line of work by exogenously varying the (implicit) career incentive involved in performance evaluations; this is a prevalent form of motivation in the public sector due to an often compressed wage structure, but has rarely been studied in the literature until recently (Deserranno et al., 2021).[6] In addition, we show that, holding the career reward fixed, a slight refinement of the performance evaluation practice can lead to a substantial improvement in bureaucratic performance, indicating a highly cost-effective way of enhancing state effectiveness.

[3] Rasul and Rogger (2016) find a negative correlation between incentives/monitoring practices and public project completion in Nigeria, which is stronger for more experienced bureaucrats. This empirical pattern is consistent with bureaucrats learning to engage in influence activities over time. Our paper complements Rasul and Rogger (2016) by experimentally altering the bureaucrats' incentives to engage in influence activities, which allows us to causally evaluate the existence and consequences of these activities in the public sector.

[4] A related paper is Wu (2017), which shows that, in a newspaper context, when both mid-level editors and top editors make editorial decisions, the bottom-level reporters have improved work performance. Our paper complements Wu (2017) by not only randomizing the authority for evaluation between two supervisors at the same level, but also cross-randomizing the employee's knowledge of the randomized evaluator's identity. This design allows us to better understand the underlying mechanisms through which the allocation of authority within organizations affects work performance, and to highlight the role of influence activities.

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

[6] Previous research focuses on the selection effect of career incentives; see Ashraf et al. (2020) for example. Our paper complements this line of work by investigating the "intensive margin" impact of career incentives, while holding selection fixed.

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 the design of various performance indicators, such as fiscal revenue (Lü and Landry, 2014), environmental standards (He et al., 2020), and population control (Serrato et al., 2019), can affect the behaviors of provincial and prefectural leaders in China. However, existing evidence has focused almost exclusively on high-level government officials, leaving incentives and constraints for the vast majority of local bureaucrats under-researched, even though they could differ substantially from those for high-level leaders.[7] Our paper intends to unravel the black box of incentive schemes for grassroots bureaucrats in China, who are the building blocks of state capacity and play key roles in 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 lay out a simple conceptual framework to help rationalize the experimental design, and generate testable hypotheses to guide the empirical analysis. In Section IV, we present experimental results testing the theoretical predictions. In Section V, we discuss potential alternative interpretations for 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 public employee assignment programs, which have hired more than one million college graduates to work with local governments in rural areas, in the hope that their modern human capital and independence from local interest groups could improve state effectiveness at the grassroots level. For example, in the College Graduate Village Officials (CGVO) program, new college graduates were hired as village officials on a contractual basis, and the arrival of CGVOs in rural villages has been shown to improve policy implementation and reduce leakages in poverty subsidy distribution (He and Wang, 2017).

[7] For instance, a key distinction is that job tasks for low-level bureaucrats are much more difficult to quantify with objective measures such as GDP growth and environmental quality. As a result, most grassroot bureaucrats are only rewarded based on subjective evaluations by their supervisors.

In this paper, we focus on the "3+1 Supports" initiative – a human capital building program for local governments launched in 2006 by the Ministry of Human Resources and Social Security.[8] Through this program, college graduates are hired to work as temporary civil servants in rural townships. They assume four types of positions: township government clerks focusing on poverty alleviation, township government clerks focusing on agricultural support, 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 "College Graduate Civil Servants" (CGCSs) through this program.

The CGCSs are recruited nationwide on a yearly basis. In May, before the end of the school year, each provincial government announces vacancies on its website and invites college graduates to apply. In most provinces, the procedure for CGCS recruitment is similar to that for recruiting regular state employees. Applicants first take a comprehensive written exam, which is similar to the Administrative Aptitude and Essay Writing Tests on the National Civil Service Exam. High-scoring applicants are then interviewed, and top-ranked candidates (based on combined scores) are recruited. Some provinces forgo tests and interviews, and screen applicants simply based on their application materials.

Admission to CGCSs is highly competitive. In most provinces, the acceptance rate for the "3+1 Supports" program is consistently below 10%. For example, Shandong province had around 1,500 positions in 2017 and attracted over 31,000 applicants (acceptance rate < 5%); in Guangxi province, the government planned to hire 800 CGCSs in 2017 and the total number of applicants exceeded 13,600 (acceptance rate < 6%). Such intense competition ensures the high quality of selected CGCSs.

The job tasks of a CGCS are similar to those of a regular entry-level township civil servant. Specifically, for CGCSs in clerical positions – as in the case of regular rural civil servants – job tasks tend to be a combination of routine paperwork, visits to villages, interactions with villagers, and other case-based assignments from supervisors. They are also responsible for policy propaganda, policy enforcement, and identifying and screening beneficiaries for various social assistance programs.

For CGCSs in more specialized positions, such as township clinic nurses or primary school teachers, job tasks are also similar to those of their colleagues who are formal public employees. CGCS teachers typically teach multiple courses, help with administrative work, and assist the regular teachers in various *ad hoc* tasks. CGCS nurses work in township clinics; their daily tasks involve assisting with

[8] In Chinese, the initiative corresponds to the "*San Zhi Yi Fu* (三支一扶)" program. Six other 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.

diagnosis and treatment, visiting villages to provide health consultations and check-ups, managing patients with chronic diseases, and providing health education. While some dimensions of these teaching and nursing jobs are better defined than those of clerical jobs, objective performance evaluation remains difficult. For example, due to the non-permanent nature of the CGCS positions, CGCS teachers are often assigned to teach non-core courses (such as art or music) or lower grades (1st to 3rd grades), where there are no school-wide exams to test student performance, and thus student scores cannot be used to objectively evaluate the performance of these teachers.

Since the multi-dimensional and vaguely-defined nature of CGCS job tasks makes it infeasible to objectively compare job performance across different individuals, the evaluation of a CGCS relies solely on the evaluating supervisor's subjective assessment. This is also the norm for the vast majority of regular civil service jobs in China and across the world.

The only major difference between a CGCS position and a regular civil servant position is that the former is based on a two-year contract while the latter is "tenured."[9] The majority of CGCSs are eager to be promoted to tenured positions upon finishing their two-year terms, which can only be approved by the government with a satisfactory supervisor evaluation.[10] As a result, aspiring CGCSs have exceptionally strong incentives to impress their supervisors. While such incentives could encourage greater work efforts, they might also induce influence activities that are misaligned with the government's interest. Simple examples of such non-productive influence activities include "buttering up" the supervisors, picking up the supervisors' kids from school, making coffee for supervisors, doing personal chores for supervisors, etc. In more extreme cases, CGCSs might try to explicitly bribe their evaluating supervisors in exchange for better evaluation outcomes.

Under the dual-leadership governance structure, every CGCS reports to both a party leader and an administrative leader. In principle, the administrative leader is in charge of the day-to-day operation of the government entity, while the party leader oversees the process and has the final say in the most high-stakes decisions. 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 becomes less clear, and there tends to be substantial overlap in their roles. This dual arrangement provides *de facto* checks

[9] In this setting, "tenure" corresponds to the "*Bian Zhi* (编制)" status, which is essentially a permanent contract provided by the government.

[10] In the provinces where our study took place, on average about 40% of the CGCSs subsequently become permanent civil servants.

and balances in local governance, including employee supervision (Li, 2018). It is prevalent in many levels of administrative units, ranging from the central government to the village committees. It is also implemented 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, she is explicitly told that the Department of Human Resources has designated one of the two leaders as the "evaluator" who is responsible for evaluating the CGCS's performance at the end of the year.[11] The CGCS, therefore, knows whose opinion matters for her career development, starting at the beginning of her appointment. Nevertheless, the CGCS is hired to work for the entire organization rather than the specific evaluator, which means that she is expected to respond to the job tasks assigned by both leaders, even though only one of them will matter for her evaluation outcomes.

### *B. Experimental Design*

In this section, we explain the experimental design and discuss the intuitions for our main hypotheses. A formal rationalization of the experiment is presented with a conceptual framework in Section III.

In collaboration with two provincial governments in China, in 2017, we randomly assigned the "revealed" and "masked" subjective performance evaluation schemes across all their 3,785 CGCSs employed in 2017. For every CGCS in our sample, one of her two supervisors was randomly selected to be the evaluator, meaning that this supervisor's assessment was given 100% weight in the final evaluation outcome. For the randomized non-evaluator, we also collected his assessment of CGCS performance, but this assessment was given no weight in determining the actual evaluation. In both schemes, we never directly informed a supervisor whether or not he was chosen as the evaluator, nor did we inform the CGCS's colleagues.

Two-thirds of the CGCSs in our sample were assigned to the "revealed" scheme. In this scheme, we informed the CGCS about the identity of her evaluating supervisor at the beginning of the evaluation cycle. This essentially mimics the current system of CGCS performance evaluation, where the agent is informed *ex-ante* about the evaluating supervisor. The key difference is that, in the current system, the evaluator is endogenously chosen from the two supervisors, typically through an opaque

[11] In our field interviews, we learned that the government decided to choose only one of the two supervisors to evaluate the CGCS in order to avoid potentially sensitive cases where the two supervisors give drastically different evaluations regarding the same CGCS, which might cause some political or legal troubles for the government itself.

process combining supervisor opinions, division of labor between supervisors, and other idiosyncratic factors. In our "revealed" scheme, by randomly selecting the evaluator, endogeneity in evaluator selection is eliminated.

We exploit the revealed scheme to test whether knowing the evaluator's identity generates asymmetry in supervisor assessments. Since the evaluator is randomly selected from the two supervisors, in the absence of any evaluator-specific influence activities, both supervisors, should on average, give similar assessments of CGCS performance. However, if the CGCS indeed engages in evaluator-specific influence activities, we should observe an asymmetry in the two supervisors' assessments of the same CGCS.

The remaining one-third of the CGCSs were assigned to the "masked" scheme. In this scheme, while we still randomized one of the two supervisors as the evaluator, we did 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 would have a 50% chance of determining her evaluation outcome. Compared to the revealed scheme, the masked scheme reduced the relative return to supervisor-specific influence activities: if the CGCS put effort into influencing a specific supervisor, there was a 50% chance that this supervisor would not end up evaluating her performance, significantly reducing the expected benefit from engaging in influence activities. As a result, under the masked scheme, a CGCS had incentives to reallocate her efforts from influence activities toward productive activities that would be appreciated by both supervisors; this should improve overall work performance.

Exploiting the randomization of CGCSs into the "revealed" vs. "masked" schemes, we can test whether introducing uncertainty in evaluator identity improves CGCS performance. Our benchmark performance indicator is the average assessment given by other colleagues. We define "colleagues" as co-workers in the same office as the CGCS who were not hired through the "3+1 Supports" program. We consider the colleagues' assessments as a credible performance measure in this context for three reasons. First, the colleagues were randomly chosen from the same office where the CGCSs work. They work closely with the CGCSs and can thus best observe the CGCSs' performance. Second, there is no obvious conflict of interest between the CGCSs and their colleagues. Unlike CGCSs, who work under two-year contracts, most colleagues already have tenure and have worked in the office for many years. As a result, the CGCSs and their colleagues do not directly compete with each other for career advancement. Finally, the CGCSs do not have obvious incentives to influence their colleagues for

evaluation purposes; at the beginning of the experiment, the provincial governments explicitly told each CGCS that only the evaluating supervisor's opinion would count for promotion.[12]

To complement colleagues' assessments, we also measured CGCS performance in three other ways. First, we elicited performance assessments from both the evaluating supervisor and the non-evaluating supervisor. Second, to make sure that the performance results were not driven by "cheap talk," we asked the supervisors and colleagues to make a recommendation to the provincial government as to whether the CGCS should be promoted to a permanent position in the current work unit. Third, we tried 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 nature of most CGCS job tasks, for some CGCS positions, a modest amount of "monthly bonus" is explicitly linked to some well-defined objective performance indicators.[13] Therefore, we can compare the actual salaries received by CGCSs between the two schemes, and infer the differences in objective performance measures (based on the bonus pay algorithms).

### *C. Implementation*

Our experiment was conducted in collaboration with the governments of two of the largest provinces in China, with a combined population of more than 150 million. Province A is coastal and more developed, while Province B is inland with a lower average income. Our sample covers all 3,785 CGCSs employed by these two provinces 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" programs to help pilot new performance evaluation schemes. The provincial governments officially informed all the CGCSs of this pilot. This high-level endorsement helped ensure that the vast majority of CGCSs were well aware of the high stakes involved in the evaluation outcomes under our newly introduced evaluation schemes.

The baseline survey was carried out in September 2017, one month after the 2017 CGCS cohort 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.[14] Different CGCSs working in the same unit (i.e., an organization branch led by

[12] Most CGCSs, in fact, did not even expect that we would survey their colleagues until the enumerators were sent to their workplaces at the end of the experiment.

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

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

the same set of supervisors) were assigned to the same scheme. This was at the request of our government partners to ensure that the evaluation outcomes of CGCSs working in the same unit could be fairly compared to each other. In this setting, because 83.9% of the work units had only one CGCS assigned, randomizing at the work unit level instead of the individual level did not make any substantial difference statistically.

Based on the randomization, in September 2017, we informed every CGCS about the evaluation scheme to which she had been assigned. Specifically, if a CGCS was randomized into the revealed scheme, we notified her 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 into the masked scheme, we notified her 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 individualized notification letters are translated in Appendix B.

To ensure the credibility of our intervention, the two provincial governments sent formal notifications with official stamps to every CGCS. The government notifications emphasized the importance of this "third-party" performance evaluation and confirmed the design of 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 enumerators visited the office where a CGCS worked, if there were fewer than five colleagues in the office, all of them were invited to fill in the colleague questionnaire; if there were more than five colleagues, the surveyor randomly sampled five of them to fill in the colleague questionnaire, using a random number generator.[15] To protect the privacy of colleagues and encourage truth-telling, colleague questionnaires were strictly anonymous, and CGCSs were not allowed to communicate with colleagues during the entire process. The CGCS survey was also conducted on-site, but independently from the colleague survey to avoid interference.

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

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 recommended 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 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. Table 1 reports the summary statistics of the CGCSs' characteristics and the differences in these variables between the revealed and masked schemes. All the characteristics are balanced across the two schemes, suggesting our randomization is well executed. In Appendix Tables A1–A2, we also report balance tests for supervisor characteristics and colleague characteristics. Again, supervisor and colleague characteristics are well balanced across the two groups. Finally, in Appendix Table A3, we further test whether supervisor characteristics in the revealed scheme are balanced between the evaluating and non-evaluating supervisors and reach the same conclusion.

Between the baseline and the end-line surveys, we lost 929 (24.5%) CGCSs in the sample. The main cause for attrition was 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 retire or be promoted or rotated to other institutions. Such job changes would break the supervisor-subordinate relationship defined by our intervention and thus invalidate the experimental design. In addition, some CGCSs passed the formal civil service exams or got admitted to graduate schools and thus decided to quit their jobs during our experiment (7.4%).

To test whether our experiment suffers from any attrition bias, we regress the attrition status on the treatment status in Appendix Table A4. We find that the masked scheme does not increase overall attrition, nor does it affect each specific type of attrition. To further investigate the potential impacts

of CGCS attrition on our findings, in Appendix Table A5, we regress a CGCS's attrition status on her baseline characteristics and find that three covariates are significantly correlated with future attrition: gender, college type, and parental education. As explained in Section IV, we conduct robustness checks in which we control for the interactions between these three covariates and our treatment variable. The main findings remain the same. Taken together, we conclude that, while CGCS attrition is not low in this institutional context, it has little impact on the empirical analyses presented in this paper.

## III. Conceptual Framework

In this section, we present a conceptual framework to rationalize the experiment and derive the main propositions that will guide our empirical analysis.

Assume a CGCS's work performance can be (at least partially) observed by her supervisors and co-workers but cannot be verified quantitatively. The organization therefore relies on a subjective performance evaluation scheme, where the agent's reward depends on the assessments given by her supervisors. To mimic our empirical setting, we assume that there are two supervisors, $j \in \{1, 2\}$.

The CGCS allocates her efforts across three dimensions. First, she can work on the "common productive dimensions" of the job ($X$), which can be observed and appreciated by both supervisors. Second, she can work on "supervisor-specific productive tasks" ($x_j$), which are assigned or observed solely by supervisor $j$. Finally, she can exert non-productive efforts to personally flatter a supervisor ($u_j$). Following Milgrom and Roberts (1988), we categorize $x_j$ as "productive influence activities" and $u_j$ as "non-productive influence activities."

Under this setup, the assessment score of supervisor $j$ is given by:

$$Y_j = X + x_j + \alpha u_j, \quad j = 1,2$$

where $\alpha > 0$ measures the marginal return to engaging in non-productive influence activities to please supervisor $j$.

We assume that the cost of working on productive activities that are appreciated by both supervisors ($X$) is $G(X)$, where $G(\mathrm{X})$ is strictly increasing and strictly convex; the cost of working on supervisor-specific productive activities ($x_1$, $x_2$) is $g(x_1 + x_2)$, where $g(\mathrm{x})$ is strictly increasing and strictly convex; the cost of non-productive influence activities is $b \cdot (u_1 + u_2)$, where $b > 0$ is a constant that measures the marginal cost of engaging in non-productive influence activities.

Each CGCS maximizes her utility subject to a time constraint:

$$Max_{X,x,u}\, V = \sum_{j=1,2} s_j \cdot \left(X + x_j + \alpha u_j\right) - G(X) - g(x_1 + x_2) - b \cdot (u_1 + u_2)$$

s.t.

$$X + \sum_j x_j + \sum_j u_j \leq T$$

where $s_j$ is the probability of each supervisor $j$'s assessment being used to determine the CGCS's reward in the performance evaluation scheme ($\sum_{j\in\{1,2\}} s_j = 1$). $T$ is the total time budget for an individual.

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 her career development: $s_1 = 1, s_2 = 0$ or $s_1 = 0, s_2 = 1$. 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 her career development: $s_1 = s_2 = \frac{1}{2}$. Solving the CGCS's maximization problem in the two schemes, under simple regularity conditions, we can derive the main testable hypotheses that will guide the empirical investigations.

**Proposition 1:** *Under the revealed scheme, the agent engages in evaluator-specific influence activities, and the evaluating supervisor gives a higher assessment than the non-evaluating supervisor.*

Without loss of generality, assume that $s_1 = 1, s_2 = 0$. We can rewrite the problem as:

$$Max_{X,x_1,u_1}\, V = (X + x_1 + \alpha u_1) - G(X) - g(x_1) - bu_1$$

s.t.

$$X + x_1 + u_1 = T$$

Solving the maximization problem, the evaluation outcome of Supervisor 1 is $E_1^{reveal} = \alpha T + (1-\alpha)G'^{-1}(1-\alpha+b) + (1-\alpha)g'^{-1}(1-\alpha+b)$, while the evaluation outcome of Supervisor 2 is $E_2^{reveal} = X^{reveal} = G'^{-1}(1-\alpha+b)$.[16] Clearly, we have $E_1^{reveal} > E_2^{reveal}$. The intuition is that, when the agent knows the identity of the evaluator, she exerts more evaluator-specific influence

[16] We assume the following singularity conditions to avoid corner solutions: $\max(0, \alpha - 1) < b < \alpha$; $0 < G'(0) < 1$, $0 < g'(0) < 1$, and $T > G'^{-1}(1-\alpha+b) + g'^{-1}(1-\alpha+b)$.

(both on the productive and non-productive dimensions), which leads to a more positive assessment from the evaluating supervisor.

**Proposition 2:** *Under the masked scheme, total influence activities decrease, while efforts on common productive activities increase.*

Since $s_1 = s_2 = \frac{1}{2}$ under the masked scheme, we can re-write the problem as:

$$Max_{X,x_1,x_2,u_1,u_2} V = X + \frac{1}{2}(x_1 + x_2) + \frac{\alpha}{2}(u_1 + u_2) - G(X) - g(x_1 + x_2) - b \cdot (u_1 + u_2)$$

s.t.

$$X + x_1 + x_2 + u_1 + u_2 \leq T$$

Solving the model, we have $X^{mask} = G'^{-1}\left(1 + \mathrm{b} - \frac{\alpha}{2}\right) > X^{reveal} = G'^{-1}(1 - \alpha + b)$, i.e., efforts on the common productive activities increase under the masked scheme.[17] In other words, when the CGCS cannot discern which supervisor will evaluate her, she spends more efforts on the tasks that can be commonly observed/appreciated by both supervisors. Given the binding time constraint, an increase in $X$ also means a reduction in total influence activities, but whether such a reduction is driven by productive influence activities or non-productive influence activities depends on the relative marginal returns to these two types of influences ($\alpha$).[18]

If we investigate the assessments of the evaluator and non-evaluator respectively, our model suggests that, when switching from the revealed scheme to the masked scheme, the non-evaluating supervisor's assessment will strictly increase, while the change in the evaluating supervisor's assessment is ambiguous: depending on the values of $\mathrm{b}, \alpha, T$, and the functional form of $G()$, $E_1^{mask}$ can be either larger or smaller than $E_1^{reveal}$.[19]

[17] We assume the following singularity conditions to avoid corner solutions: $\max(0, \alpha - 1, \frac{\alpha}{2} - \frac{1}{2}) < b < \frac{\alpha}{2}$; $0 < G'(0) < 1, 0 < g'(0) < 1/2$, and $T > G'^{-1}\left(1 + \mathrm{b} - \frac{\alpha}{2}\right) - g'^{-1}\left(\frac{1}{2} + \mathrm{b} - \frac{\alpha}{2}\right)$.

[18] We have $x^{mask} > x^{reveal}$ if $\alpha > 1$, $x^{mask} < x^{reveal}$ if $\alpha < 1$. The intuition is that, when $\alpha > 1$, the marginal return of working on non-productive influence activities $u$ is larger than the marginal return of working on advisor-specific tasks $x$, so the agent will exert more efforts on flattering her supervisor. The comparison between $u^{mask}$ and $u^{reveal}$ is uncertain; it depends on the values of $\mathrm{b}, \alpha, T$, and the functional forms of $G(X)$ and $g(x)$.

[19] Later in Section IV, we find that the evaluating supervisor's assessment score under the masking scheme is slightly higher than the evaluating supervisor's assessment score under the revealed scheme. We provide a numerical example of the model that can rationalize this finding in Appendix C.

## IV. Main Results

In this section, we present a series of experimental results to verify the main propositions of our conceptual framework. In the revealed scheme, we find that the assessment given by the (randomized) evaluating supervisor is substantially higher than that given by the (randomized) non-evaluating supervisor, and this asymmetry in supervisor assessments is correctly predicted by the CGCS's colleagues, confirming *Proposition 1*. When switching from the revealed scheme to the masked scheme, the asymmetries in supervisor assessment and colleague perceptions no longer exist. Instead, we find significant improvements in various measures of CGCS work performance and suggestive evidence of reduced influence activities and increased work efforts, which, together, consistently support *Proposition 2*. Moreover, we investigate the underlying mechanisms for the effects of masking evaluator identity, and document suggestive evidence that both productive influence activities and non-productive influence activities are reduced under the masked scheme.

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

First, we investigate *Proposition 1*: whether revealing the identity of the evaluator to the CGCS will cause the evaluator to be more positive about the CGCS than the non-evaluator. Specifically, for each CGCS's two supervisors, we randomly label them as "Supervisor 1" and "Supervisor 2," and then use the subsample of CGCSs in the revealed scheme to estimate the following econometric model:

$$Sup1_Edge_{icst} = \alpha \cdot Sup1_Eval_i + \gamma_c + \lambda_s + \phi_t + \epsilon_{icst}$$

where the outcome variable $Sup1_Edge_{icsy}$ is defined as "Supervisor 1's assessment score minus Supervisor 2's assessment score for CGCS $i$," who is in county $c$, cohort $t$, and serves as CGCS type $s$.[20] $Sup1_Eval_i$ is a dummy variable indicating whether CGCS $i$ is being evaluated by supervisor 1 (instead of supervisor 2). $\gamma_c$, $\lambda_s$, and $\phi_t$ represent county FE, CGCS type FE, and cohort FE, respectively. Standard errors are clustered at the work unit level. Under this specification, since the evaluator is randomly chosen among the two supervisors for each CGCS, $\alpha$ causally identifies the additional positiveness due to being assigned as evaluator.

As shown in Column (1) of Table 2, for CGCSs in the revealed scheme, if a supervisor was chosen as the evaluator at the baseline, he indeed gave a more positive assessment at the end line, and the

[20] This represents the four types of CGCS positions: township government clerks focusing on poverty alleviation, township government clerks focusing on agricultural support, teachers in township primary schools, and nurses in township clinics.

magnitude of this "evaluator edge" in assessment scores is as large as 0.24 standard deviation. In Column (2), we include a rich set of control variables in the regression, and the estimated coefficient remains unchanged, confirming that the treatment of "Supervisor 1 Evaluating" was randomly assigned. This asymmetry in supervisor assessments is consistent with the agent engaging in evaluator-specific influence activities to improve evaluation outcomes.

If the "assessment asymmetries" documented in Columns (1)–(2) are indeed caused by evaluator-specific influence activities, as we have argued, they should only exist when the CGCS knows who the "target" is. Under the masked scheme, when the CGCS no longer knows the identity of the evaluator, there should be no asymmetry in supervisor assessments. Therefore, as a placebo test, in Columns (3)–(4), we focus on the masked scheme where the randomly chosen evaluator's identity was not announced until the end of the evaluation cycle. As we can see, in the masked scheme, being selected as the evaluator indeed no longer leads to more positive assessments than the other non-evaluating supervisor.

In addition, if the results on scoring asymmetries are indeed driven by CGCSs engaging in influence activities, to the extent that such behaviors can at least partially be 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 the CGCS's performance. To put it another way, we expect that, under the revealed scheme, the colleagues would systematically point to the evaluator as the more positive supervisor among the two, despite the fact that we never informed the colleagues who the evaluator is. In contrast, since the CGCS should not try to impress or flatter the two supervisors differentially under the masked scheme, we expect that the colleagues will no longer systematically point to the evaluator as being more positive.

We test this prediction in Table 3, where the outcome variable is the ratio of a CGCS's colleagues who think that Supervisor 1 would be more positive than Supervisor 2 about the CGCS's performance. We see that, under the revealed scheme, when Supervisor 1 is randomly selected as the evaluator, colleagues are more likely to (correctly) think he is going to give more positive assessments than Supervisor 2; by contrast, under the masked scheme, colleagues will (again, correctly) not expect the evaluator to be more positive about CGCS performance. These results are consistent with *Proposition 1,* and suggest that influence activities can be at least partially observed by the co-workers of the agent.

### *B. Proposition 2: Improved Work Performance under the Masked Scheme*

As suggested by *Proposition 2*, the masked scheme makes a CGCS reallocate her efforts from evaluator-specific influence activities toward common productive tasks, leading to improved performance. To test this hypothesis, we evaluate the impacts of masking the evaluator's identity on a series of work performance measures. Specifically, we use the full sample of CGCSs (in both revealed and masked schemes), and estimate the following econometric model:

$$Y_{icst} = \alpha \cdot Mask_i + \gamma_c + \lambda_s + \phi_t + \epsilon_{icst}$$

where $Y_{icst}$ is a performance measure for CGCS $i$, and $Mask_i$ is a dummy variable indicating whether CGCS $i$ belongs to the masked scheme. Since the CGCSs were randomly assigned between the two evaluation schemes, $\alpha$ identifies the causal effect of being assigned to the masked scheme (relative to the revealed scheme). The standard errors are clustered at the work unit level.

The multi-dimensional and subjective nature of the CGCS jobs means that we are unable to collect comprehensive objective performance indicators that are interpersonally comparable across all CGCSs, which is why the government used subjective evaluation schemes for CGCS promotion in the first place. That being said, we try our best to paint a more complete picture of the CGCSs' performance under different evaluation schemes, by investigating a series of different performance indicators collected in our end-line surveys.

***Colleague Assessments***

First, we investigate how colleague assessment of the CGCS performance varies between the revealed and masked schemes. As explained in Section II(B), we consider colleague assessment to be a credible measure of a CGCS's performance in the common productive tasks that benefit the organization, which, according to *Proposition 2*, should improve under the masked scheme.[21]

In Table 4 Columns (1)–(2), the dependent variable is the average colleague assessment of the CGCS's performance, which is framed relative to other civil servants employed in the same work unit. The assessment score in the questionnaire ranges from 1 to 7, representing different categories from "worse than all other colleagues" to "better than all other colleagues." As we can see, being assigned to the masked scheme leads to significantly higher colleague assessment scores. In Appendix Table A7, we find that the improvement in colleague assessment associated with masking the evaluator's identity is comparable to the performance gap between four-year regular college graduates and three-

[21] Colleagues observe CGCS performance closely, but their opinions are not included in the performance evaluation scheme, so the agent is not incentivized to adjust her efforts to improve colleague assessments.

year community college graduates, suggesting that the treatment effect of the masked scheme is economically significant.

This is corroborated by Columns (3) and (4), in which we ask the colleagues whether they think the CGCS's performance ranks in the top 10% of the organization, and find that CGCSs in the masked scheme are significantly more likely to be recognized as top performers in the organization. In Columns (5)–(6), we also find that colleagues think the CGCSs in the masked scheme are more hardworking. Taken together, these results are consistent with *Proposition 2*.[22]

***Supervisor Assessments***

Second, we investigate how the assessments given by the evaluating supervisor and the non-evaluating supervisor, respectively, change in the masked scheme. According to *Proposition 2,* we expect the non-evaluating supervisor to become more positive about CGCS performance under the masked scheme, while the change in the evaluating supervisor's assessment is theoretically ambiguous. In Appendix C, we use numerical examples to demonstrate the intuitions behind the ambiguous impacts of the masked scheme on the evaluator's assessment.

In Table 5 Column (1), the outcome variable is the mean assessment of the two supervisors. We find that masking the identity of the evaluator significantly improves the average supervisor assessment. In Columns (2) to (4), we find that the non-evaluating supervisor's assessment increases by a significant margin when the evaluator's identity is masked, while the evaluating supervisor's assessment also slightly increases (but this is statistically nonsignificant). As a result, the deviation in assessments also decreases in the masked scheme. These results are consistent with our conceptual framework, where masking the evaluator's identity leads to less evaluator-specific influence activity and higher overall work performance.

***Promotion Recommendations and Performance Pays***

Third, in addition to the subjective assessments given by colleagues and supervisors, we explore a set of "revealed-preference" and arguably more "objective" performance measures to further support our findings.

In Column (1) of Table 6, we directly asked colleagues whether they recommend to the provincial government that the CGCS be promoted to a tenured position in this office after finishing her two-year term. Using this "revealed-preference" measure as an alternative outcome, we find that masking

[22] To investigate whether the results are affected by the attrition of CGCSs, in Appendix Table A6, we control for the interactions between the masked scheme dummy and the dimensions of CGCS characteristics that predict attrition. As we can see, all the main findings remain, and, if anything, have larger magnitudes.

evaluator identity makes more colleagues report that the CGCS deserves tenure, which is consistent with the previous finding that colleagues think the CGCS shows better work efforts and performance under the masked scheme.

We also asked each CGCS to report their total monthly remuneration, including basic wages and performance bonuses (if any), which we later verified using administrative information provided by the provincial governments. The basic wage is set by the upper-level government and matches the entry-level permanent civil servant wage, so it should be exactly the same for all CGCSs, conditional on the county of residence, enrollment year, and position type. In addition to the basic wage, some work units have some discretion over a modest amount of performance bonuses to reward the best performing employees (based on their own criteria). In Table 6 Columns (3) and (4), we observe that, on average, the CGCSs in the masked scheme earn 50-RMB (2.3%) higher salaries than those in the revealed scheme. Since the basic salary for CGCSs is fixed, this income gap reflects the difference in performance bonuses.

During our field interviews, we were informed that the CGCSs who work as nurses in township clinics enjoy the most substantial performance bonuses, because these clinics have a "business" feature and can keep some profits to reward the most hard-working staff. For nurses, the most important factor determining their bonus differentials is the number of night shifts taken each month: every additional night shift is rewarded by about 20 RMB (about $3). In Columns (5) and (6), when we restrict the sample to CGCSs working as nurses, we find a larger than 110-RMB (6.2%) income gap between the two schemes. The compensation differential between the revealed and masked groups is therefore equivalent to nearly six additional night shifts per month. This result suggests that the performance improvement caused by the masked scheme is indeed substantial when benchmarked objectively.

As reflected by various measures, the evidence consistently suggests that CGCS performance along the commonly observable productive tasks improved in an economically significant manner under the masked scheme, confirming *Proposition 2* of our model.

### *C. Evidence for Productive and Non-Productive Influence Activities*

Following Milgrom and Roberts (1988), our conceptual framework sketches two types of influence activities: (1) productive influence activities, where the CGCS spends more efforts on job tasks deemed important by the evaluator (relative to those deemed important by the non-evaluator); (2) non-productive influence activities, where the CGCS tries to befriend the evaluator by flattering him

or providing personal favors to him. In this section, we provide additional evidence on the relevance of each type of influence activity.

In Table 7, we investigate the existence of productive influence activities under the revealed scheme. In the end-line survey, we asked each CGCS, "among all the job tasks you need to do, what is the proportion that is assigned by Supervisor 1/Supervisor 2?" In Columns (1)–(2), we find that, if Supervisor 1 is chosen as the evaluator, the CGCS will report that she has more job tasks assigned by Supervisor 1. In addition, we asked each CGCS, "is your most important job task assigned by Supervisor 1 or Supervisor 2?" It turns out that, when Supervisor 1 is chosen as the evaluator, the GCGS is more likely to think her most important job task is assigned by Supervisor 1 (Columns (3)–(4)). Finally, we asked the CGCS, "among all the job dimensions, in which dimension do you think you improved most in the past year?" We find that, when a job dimension is deemed as important by the evaluating supervisor, the CGCS is more likely to improve along this specific dimension (Columns (5)–(6)). Since the evaluator is randomly selected between the two supervisors, and the selection is blinded to both supervisors throughout the evaluation cycle, these results reflect behavioral changes on the CGCS's side to cater towards the preference of the evaluating supervisor. In contrast, as we can see in Appendix Table A8, when we mask the identity of the evaluator, none of the empirical patterns in Table 7 still exists, which is consistent with an alleviation of "multi-tasking" problems under the masked scheme.

To investigate the existence of non-productive influence activities, we asked a set of additional questions about each CGCS's work experience in the end-line survey. First, we asked the CGCS, "what was the most challenging part of your CGCS experience?"[23] As shown in Table 8 Column (1), the CGCSs under the revealed scheme are significantly more likely to report that "handling personal relationships with the supervisors" is the most challenging part of their experience, as compared to their peers in the masked scheme. In contrast, as can be seen from Column (2), the proportion of CGCSs identifying "handling personal relationships with colleagues" as the most challenging part of the experience is the same across the two schemes. These two results are consistent with our model, in which the CGCS engages in more non-productive influence activities under the revealed scheme than under the masked scheme, and does not have incentives to influence her colleagues under either

[23] The choices included "familiarizing myself with the local governance system," "handling the personal relationship with my supervisor," "handling personal relationships with my colleagues," "adjusting to life in the rural area," "working on tasks unrelated to my college major," "adjusting to unfamiliar work and life conditions," "getting useful work feedback," and "other challenges."

scheme.[24] We then asked each CGCS, "do you think the civil service system is meritocratic?" In Column (3), we find that the CGCSs under the masked scheme are significantly more likely to give a positive answer to this question. In Column (4), we also asked the CGCS, "do you think hard work pays off in your position?" Similarly, we find that the masked scheme makes the CGCS more likely to believe that working hard pays off. These results are consistent with our interpretation that the CGCS reallocates her efforts from non-productive influence activities to common productive tasks under the masked scheme.

Taken together, our results suggest that both productive and non-productive influence activities play important roles within China's local governance system.

### *D. Influence Activities as the Source of Hometown Favoritism*

To better understand the economic significance of influence activities in our context, we benchmark the impacts of influence activities against the magnitude of "hometown favoritism" between employer and employees, which is well-documented as playing an important role in China's bureaucratic system (Fisman and Wang, 2015; Fisman et al., 2020). Specifically, we define hometown favoritism as the gap in assessment scores between a "same-hometown evaluator" and a "different-hometown evaluator" regarding the same CGCS. Such favoritism can be decomposed into two parts: (1) the "top-down" preference, meaning that an evaluator would spontaneously assess a same-hometown CGCS more positively; and (2) the "bottom-up" influence activities, meaning that a CGCS would find it easier to influence an evaluator who is from the same hometown.

Because the evaluator was randomly assigned between the two supervisors, our experiment introduced exogenous variations in whether the supervisors and the CGCS share the same hometown. In addition, in the revealed scheme, the evaluator and the CGCS are made aware of each other's identity, so both "top-down preference" and "bottom-up influence" could be at work. In the masked scheme, however, the evaluator is aware of the identity of the CGCS, but not the other way around, which keeps the "top-down preference" while alleviating "bottom-up influence." Therefore, by

[24] Relatedly, as shown in Appendix Table A8, the "unexplained positiveness in evaluator assessment," defined as the residual obtained from regressing "evaluator assessment" on "non-evaluator assessment" and "colleague assessments," is significantly lower in the masked scheme. This is also consistent with our interpretation that the CGCS engages less in influence activities under the masked scheme. For this result to be confounded, an alternative interpretation would need to explain why the evaluating supervisor can appreciate the achievements of the CGCS better than either the non-evaluator or any colleague does, and why such appreciation only exists in the revealed scheme.

comparing the magnitude of "hometown favoritism" across the revealed vs. masked schemes, we can infer the relative importance of influence activities.

The results are presented in Table 9.[25] In Column (1), we find that, if the CGCS shares the same hometown with the evaluator, the evaluator indeed gives a higher performance assessment, confirming the existence of hometown favoritism in this setting. Then, we examine this favoritism separately for the revealed-scheme sample and the masked-scheme sample in Columns (2) and (3). We find that the hometown favoritism can only be observed under the revealed scheme, but not under the masked scheme. Additionally, for the non-evaluating supervisor, sharing the same hometown with the CGCS does not affect the assessment score, regardless of whether the identity of the evaluator is revealed or masked (Columns (4)–(6)).

We can draw two conclusions from these results. First, because the "top-down" preference should remain the same across different evaluation schemes, the results suggest that the "bottom-up" influence activities are likely driving the observed hometown favoritism in our data. Second, because the assessment scores of both supervisors are uncorrelated with the hometown tie in the masked scheme, we can infer that, without influence activities from the CGCS, the hometown tie alone cannot generate meaningful favoritism. These findings further testify to the importance of influence activities in China's bureaucratic system.

## V. Alternative Interpretations

In this section, we discuss the main alternative interpretations to our main findings.

### *A. Alternative Interpretations of Asymmetric Supervisor Assessments under the Revealed Scheme*

Our interpretation of the findings in Table 2 is based on *Proposition 1*: in the revealed scheme, the CGCS is able to perform evaluator-specific influence activities. There are three potential confounding explanations.

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

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 behavior. This concern is partly alleviated by our experimental design, as supervisors were not informed by the

[25] The sample size is smaller due to missing values for supervisors' hometowns.

research team about their roles in the evaluation. However, it is still possible that the CGCS might have informed her evaluator about his role.

To investigate this possibility, in our end-line survey, we directly asked each evaluator whether he was aware of his responsibility of evaluating the CGCS. It turns out that the majority of them (more than 65%) did not know whether they were chosen as evaluators until after they had finished the evaluations. In Panel A of Table 10, we re-estimate the specifications in Table 2 separately for the subsample in which supervisors did not know their (evaluator) roles, and the subsample in which supervisors did know about their evaluator roles. We find that the asymmetry in supervisor assessments under the revealed scheme is almost identical in these two subsamples, suggesting that our results are not driven by supervisor behavioral changes due to being selected as the evaluator.

Moreover, in Panel B of Table 10, 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 our end-line survey; (2) the number of words a supervisor writes when assessing the CGCS's performance; and (3) the frequency with which the supervisor meets with the CGCS. Our hypothesis is that, if the evaluating supervisor indeed paid more attention to the CGCS's performance, he should be more likely to answer the survey, write more words in the assessment, and meet with the CGCS more frequently. Again, our results do not support this interpretation.

### *2. More Information for the Evaluating Supervisor*

Another confounding story is that the evaluating supervisor might receive more information regarding CGCS performance from various sources; the CGCS, colleagues, and the other (non-evaluating) supervisor might all try to send signals to help him better evaluate. This increase in information might improve the evaluator's assessment and thus create the scoring asymmetry shown in Table 2. Again, we think this interpretation is unlikely to be of first-order importance, given that we never directly informed the colleagues or the supervisors about the identity of the evaluator.

Nevertheless, we explicitly examine this alternative interpretation. In our end-line survey, we asked each supervisor, "how frequently did the CGCS, the colleagues of the CGCS, or the other supervisor discuss the CGCS's performance with you?" We are interested in whether the evaluating supervisor received more information than the non-evaluating supervisor from these three sources. In Panel C of Table 10, we find that, relative to the non-evaluator, the evaluator did not gain extra information

from any of these sources.[26] Therefore, the asymmetry in supervisor assessments under the revealed scheme cannot be explained by the difference in information between the two supervisors.

### *B. Alternative Interpretations of Improved Assessments under the Masked Scheme*

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

#### *1. CGCS Influencing Both Supervisors More*

The first alternative interpretation is that, under the masked scheme, the CGCS does not work harder on productive dimensions. Instead, she simply extends more influence activities toward both supervisors, which is why we see improved average supervisor assessment. However, this interpretation is inconsistent with a series of 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 incentive to influence her colleagues; every CGCS is clearly informed that only her 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, rather than working harder, there should not be a significant improvement in average colleague assessment.

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

Third, as documented in Table 8, under the masked scheme, the CGCSs are less worried about handling personal relationships with supervisors, as compared to their peers under the revealed

[26] If anything, the evaluator is 3% less likely to receive information regarding CGCS performance from colleagues, although the coefficient is small in magnitude and only marginally significant.

scheme. This also suggests a reduction in influence activities, rather than extending influence activities to both supervisors, under the masked scheme.

### *2. CGCS Influencing Colleagues under the Masked Scheme*

Suppose that the CGCSs, for whatever reason, are trying to influence their colleagues, and they do so to a larger extent under the masked scheme. Could this be confounding our results on improved colleague assessments under the masked scheme? To begin with, this interpretation is inconsistent with the result in Table 8 Column (2), which shows that the proportion of CGCSs worrying about "handling personal relationships with colleagues" remains the same across both schemes. In addition, we conduct two placebo tests to further rule out this confounding interpretation.

In the first placebo test, we hypothesize that, under the masked scheme, additional influence activities toward colleagues will result in more interactions between the CGCSs and their colleagues, especially on non-professional occasions. However, in Panel A of Table 11, we find that, according to colleagues, CGCSs did not communicate with them more frequently under the masked scheme (Columns (1) and (2)), nor did they become more familiar with the CGCSs' work or personal lives under the masked scheme (Columns (3) and (4)). These results are inconsistent with the alternative interpretation.

In the second placebo test, we examine whether there exists hometown favoritism in colleague assessment. Recall that, in Table 9, we document the existence of hometown favoritism in supervisor assessment and show that "bottom-up influence activities" are driving such favoritism. We conduct a similar exercise using colleague assessment and check whether the CGCSs have incentives to influence their colleagues under the masked scheme. As shown in Appendix Table A10, a "same hometown colleague" does not show differential positiveness across the two schemes. Taken together, we conclude that the CGCSs do not engage in additional influence activities toward colleagues under the masked scheme.

### *3. Higher Information Quality under the Masked Scheme*

Another possibility is that supervisors in the masked scheme get better information on CGCS performance, which might explain the increase in average supervisor assessment. To address this concern, in Panel B of Table 11, we examine whether supervisors get additional information on CGCS performance under the masked scheme, either from colleagues or from the CGCS herself. We find that the masked scheme does not increase the frequency of CGCSs and other colleagues reporting to either the evaluating supervisor or the non-evaluating supervisor regarding CGCS performance. This

suggests that improved supervisor assessments in the masked scheme cannot be explained by changes in information quality.

*4. CGCS Gets Discouraged when Matched to "Hostile Evaluator" under Revealed Scheme*

A remaining possibility is that, under the revealed scheme, some CGCSs might be matched with an evaluator whom they perceive as hostile, in that, no matter how hard the CGCS works, efforts will not be appreciated by this evaluator. As a result, the CGCSs get discouraged and put little effort into productive tasks, which might explain why performance is higher under the masked scheme.

In our baseline survey, before the randomizations of schemes and evaluators were realized, we asked each CGCS, "among the two supervisors, whom would you prefer to be your evaluator?" Due to randomization, half of the CGCSs under the revealed scheme would be evaluated by their "non-preferred" supervisor, and the other half evaluated by their "preferred" supervisor. Since the "discouragement" mechanism should operate only through those evaluated by the non-preferred supervisor, we can compare performance differences between CGCSs facing the preferred supervisor under the revealed scheme and those under the masked scheme. If discouragement was driving the observed improvement in CGCS performance, we should expect the performance improvement under the masked scheme to disappear in this restricted comparison. However, as shown in Appendix Table A11, the masking effect actually becomes slightly stronger in this subsample analysis, providing strong evidence against the "discouragement" interpretation.

## VI. Conclusion

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

To shed light on this topic, we conducted a large-scale field experiment, where we helped the government randomize two subjective performance evaluation schemes among 3,785 junior state employees in China. In the "revealed" scheme, we randomly chose one of the two supervisors as the performance evaluator and informed the subordinate *ex-ante* about the evaluator's identity. We find

that, under this scheme, subordinates are indeed induced to engage in evaluator-specific influence activities to improve their evaluation outcomes. Such favoritism can be predicted by the co-workers of the agent, suggesting that co-workers are able to observe influence activities in the workplace.

In the "masked" scheme, we randomly chose one of the two supervisors as the performance evaluator, but the identity of the evaluator was not disclosed to the subordinate, which reduced the expected return to supervisor-specific influence activities. Therefore, masking evaluator identity should encourage the subordinate to reallocate her efforts from influence activities toward common productive dimensions that could be appreciated by both supervisors. We find that this evaluation scheme indeed improves the subordinate's work performance, as measured by average colleague assessments, average supervisor assessments, likelihood of being recommended for "tenure," and monthly bonus payments determined by objective performance indicators.

We also distinguish between two types of influence activities. On the one hand, there will be multi-tasking issues, where the agent works harder on tasks that are assigned or better observed by the evaluator. On the other hand, the agent will also try to benefit the evaluator personally through non-productive influence activities that go beyond her mandated tasks. Our analyses suggest that both types of influence activities are prevalent in China's local bureaucratic system.

Our findings also reveal the mechanisms behind the formation of hometown favoritism in China's bureaucratic system. We show that "bottom-up" influence activities (i.e., subordinates trying to benefit supervisors coming from the same hometown) are likely the driving force behind the observed hometown favoritism, while "top-down" preference (i.e., supervisors have an intrinsic bias towards subordinates from the same hometown) plays a negligible role.

In addition to providing the first rigorous empirical evidence on the existence and implications of influence activities, our findings have important policy implications. We find that by introducing uncertainty in the evaluator's identity, which has minimal implementation cost, the government can significantly improve the job performance of its employees. Further, this uncertainty causes more state employees to believe that hard work pays off and that the bureaucratic system is meritocratic. These belief changes should have far-reaching consequences for the working culture and ethics of the Chinese government. Our findings not only have direct policy implications for the more than 50 million state employees in China, but also are likely to be relevant for many other contexts where high-stakes rewards are linked to the subjective opinions of designated evaluators.

## REFERENCES


Ashraf, Nava, and Oriana Bandiera. "Social incentives in organizations." *Annual Review of Economics* 10 (2018): 439-463.

Ashraf, Nava, Oriana Bandiera, Edward Davenport, and Scott S. Lee. "Losing prosociality in the quest for talent? Sorting, selection, and productivity in the delivery of public services." *American Economic Review* 110, no. 5 (2020): 1355-94.

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.

Deserranno, Erika, and Gianmarco León Ciliotta. "Promotions and productivity: The role of meritocracy and pay progression in the public sector." (2021).

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.

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. "Watering down environmental regulation in China." *The Quarterly Journal of Economics* 135, no. 4 (2020): 2135-2185.

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.

Lü, Xiaobo, and Pierre F. Landry. "Show me the money: Interjurisdiction political competition and fiscal extraction in China." *American Political Science Review* (2014): 706-722.

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

Maestri, Lucas. "Bonus payments versus efficiency wages in the repeated principal-agent model with subjective evaluations." *American Economic Journal: Microeconomics* 4, no. 3 (2012): 34-56.

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.

Powell, Michael. "An influence-cost model of organizational practices and firm boundaries." *The Journal of Law, Economics, and Organization* 31, no. suppl_1 (2015): i104-i142.

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

Rasul, Imran, and Daniel Rogger. "Management of bureaucrats and public service delivery: Evidence from the Nigerian civil service." *The Economic Journal* 128, no. 608 (2018): 413-446.

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.

Serrato, Juan Carlos Suárez, Xiao Yu Wang, and Shuang Zhang. "The limits of meritocracy: screening bureaucrats under imperfect verifiability." *Journal of Development Economics* 140 (2019): 223-241.

Shirk, Susan L. *The Political Logic of Economic Reform in China*. Vol. 24. Univ of California Press, 1993.

Wu, Yanhui. "Authority, incentives, and performance: Evidence from a Chinese newspaper." *Review of Economics and Statistics* 99, no. 1 (2017): 16-31.

**Table 1. Balance Check: CGCS Characteristics**

| | Revealed Scheme | Masked Scheme | Difference |
|---|---|---|---|
| | (1) | (2) | (3) |
| Age | 24.868 | 24.928 | 0.039 |
| (at Endline) | (1.630) | (1.604) | (0.061) |
| Gender | 0.592 | 0.600 | 0.009 |
| (=1 if Female) | (0.492) | (0.490) | (0.019) |
| Social Science Major | 0.555 | 0.545 | -0.015 |
| (=1 if Yes) | (0.497) | (0.498) | (0.020) |
| 4-Year College or Above | 0.723 | 0.724 | -0.004 |
| (=1 if Yes) | (0.448) | (0.447) | (0.017) |
| STEM Students in High School | 0.347 | 0.342 | -0.006 |
| (=1 if Yes) | (0.476) | (0.475) | (0.020) |
| Party Member | 0.217 | 0.218 | -0.002 |
| (=1 if Yes) | (0.412) | (0.413) | (0.017) |
| Parent Completing College | 0.288 | 0.285 | -0.005 |
| (=1 if Yes) | (0.453) | (0.452) | (0.019) |
| Work in Village | 0.160 | 0.150 | -0.012 |
| (=1 if Yes) | (0.366) | (0.357) | (0.015) |
| CEE Score | 4.813 | 4.856 | 0.057 |
| (100 Points) | (0.744) | (0.723) | (0.036) |
| Risk Averse | 0.471 | 0.477 | -0.000 |
| (=1 if Yes) | (0.499) | (0.500) | (0.021) |
| Obs. | 1,935 | 919 | 2,854 |

*Notes*: The first two columns summarize the mean and standard deviation of CGCS characteristics. Column (1) uses the sample of CGCSs in the revealed scheme, Column (2) uses the sample of CGCSs in the masked scheme. Column (3) checks the covariate balance between the revealed group and the masked group, controlling for county FE, CGCS type FE, and cohort FE, with standard errors clustered at the work unit level.

**Table 2. Asymmetry in Supervisor Assessments**

| | Supervisor 1's Score Minus Supervisor 2's Score | | | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| Supervisor 1 Evaluating | 0.310*** | 0.310*** | -0.097 | -0.091 |
| | (0.082) | (0.082) | (0.121) | (0.125) |
| Sample | Revealed | Revealed | Masked | Masked |
| DV Mean | -0.02 | -0.02 | 0.00 | 0.00 |
| DV S.D. | 1.30 | 1.30 | 1.22 | 1.22 |
| Controls | N | Y | N | Y |
| County FE | Y | Y | Y | Y |
| Type FE | Y | Y | Y | Y |
| Cohort FE | Y | Y | Y | Y |
| Obs. | 1,301 | 1,301 | 580 | 580 |

*Notes:* Each column represents a separate regression. In Columns (1) and (2), the sample is for the revealed scheme only; in Columns (3) and (4), the sample is for the masked scheme only. Standard errors clustered at the work unit level are reported below the coefficients. Control variables include CGCS characteristics presented in Table 1, and colleague characteristics presented in Appendix Tables A1. * significant at 10% ** significant at 5% *** significant at 1%.

**Table 3. Influence Activities and Favoritism Observed by Colleagues**

| | Colleagues Predict Sup1 More Positive than Sup 2 | | | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| Supervisor 1 Evaluating | 0.042** | 0.042* | -0.001 | 0.001 |
| | (0.021) | (0.021) | (0.033) | (0.033) |
| Sample | Revealed | Revealed | Masked | Masked |
| DV Mean | 0.54 | 0.54 | 0.53 | 0.53 |
| DV S.D. | 0.50 | 0.50 | 0.50 | 0.50 |
| Controls | N | Y | N | Y |
| County FE | Y | Y | Y | Y |
| Type FE | Y | Y | Y | Y |
| Cohort FE | Y | Y | Y | Y |
| Obs. | 1,815 | 1,815 | 824 | 824 |

*Notes:* Each column represents a separate regression. In Columns (1) and (2), the sample is for the revealed scheme only; in Columns (3) and (4), the sample is for the masked scheme only. Standard errors clustered at the work unit level are reported below the coefficients. Control variables include CGCS characteristics presented in Table 1, and colleague characteristics presented in Appendix Tables A1. * significant at 10% ** significant at 5% *** significant at 1%.

**Table 4. Masking and Colleague Assessments**

| | Performance (1-7) | | Top 10% | | Hardworking | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Masking | 0.217*** | 0.217*** | 0.077*** | 0.077*** | 0.028** | 0.028** |
| | (0.035) | (0.035) | (0.013) | (0.013) | (0.012) | (0.012) |
| DV Mean | 5.22 | 5.22 | 0.70 | 0.70 | 0.44 | 0.44 |
| DV S.D. | 0.91 | 0.91 | 0.32 | 0.32 | 0.43 | 0.43 |
| Controls | N | Y | N | Y | N | Y |
| County FE | Y | Y | Y | Y | Y | Y |
| Type FE | Y | Y | Y | Y | Y | Y |
| Cohort FE | Y | Y | Y | Y | Y | Y |
| Obs. | 2,837 | 2,837 | 2,837 | 2,837 | 2,837 | 2,837 |

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

**Table 5. Masking and Supervisor Assessments**

| | Mean Assessment | Evaluator Assessment | Non-Evaluator Assessment | Assessment Deviation |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| Masking | 0.138*** | 0.060 | 0.216*** | -0.098** |
| | (0.046) | (0.056) | (0.059) | (0.050) |
| DV Mean | 5.14 | 5.18 | 5.11 | 0.90 |
| DV S.D. | 0.91 | 1.13 | 1.10 | 0.93 |
| Controls | Y | Y | Y | Y |
| County FE | Y | Y | Y | Y |
| Type FE | Y | Y | Y | Y |
| Cohort FE | Y | Y | Y | Y |
| Obs. | 1,940 | 1,940 | 1,940 | 1,940 |

*Notes*: Each column represents a separate regression. We keep the subsample of CGCSs for whom both supervisors completed our assessment form. Standard errors clustered at work unit level are reported below the coefficients. Control variables include CGCS characteristics presented in Table 1. * significant at 10% ** significant at 5% *** significant at 1%.

**Table 6. Treatment Effects on "Revealed Preference" and Objective Measures**

| | Qualify for Tenure | | Wage | | Wage (Medical Support) | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Masking | 0.036*** | 0.035*** | 48.81** | 49.42** | 115.54* | 121.58* |
| | (0.011) | (0.011) | (22.41) | (22.43) | (61.94) | (65.98) |
| DV Mean | 0.87 | 0.87 | 2113 | 2113 | 1892 | 1892 |
| DV S.D. | 0.33 | 0.33 | 672 | 672 | 452 | 452 |
| Controls | N | Y | N | Y | N | Y |
| County FE | Y | Y | Y | Y | Y | Y |
| Type FE | Y | Y | Y | Y | Y | Y |
| Cohort FE | Y | Y | Y | Y | Y | Y |
| Obs. | 2,839 | 2,839 | 2,750 | 2,750 | 193 | 193 |

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

**Table 7. Multi-Tasking under the Revealed Scheme**

| | Ratio of CGCS's Job Tasks Assigned by Supervisor 1 | | CGCS's Most Important Job Task is Guided by Supervisor 1 | | CGCS Improved More in Areas Deemed Important by Sup 1 (Relative to Sup 2) | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Supervisor 1 Evaluating | 0.031** | 0.032** | 0.051* | 0.051* | 0.124** | 0.124** |
| *(Revealed)* | (0.014) | (0.014) | (0.030) | (0.030) | (0.056) | (0.056) |
| DV Mean | 0.48 | 0.48 | 0.50 | 0.50 | -0.02 | -0.02 |
| DV S.D. | 0.24 | 0.24 | 0.47 | 0.47 | 0.94 | 0.94 |
| Controls | N | Y | N | Y | N | Y |
| County FE | Y | Y | Y | Y | Y | Y |
| Type FE | Y | Y | Y | Y | Y | Y |
| Cohort FE | Y | Y | Y | Y | Y | Y |
| Obs. | 1,482 | 1,482 | 1,134 | 1,134 | 1,300 | 1,300 |

*Notes:* Each column represents a separate regression. The sample is for the revealed scheme only. Standard errors clustered at the work unit level are reported below the coefficients. Control variables include CGCS characteristics presented in Table 1, and colleague characteristics presented in Appendix Tables A1. * significant at 10% ** significant at 5% *** significant at 1%.

**Table 8. Treatment Effects on Influence Activities and Work Efforts**

| | CGCS Challenge: Supervisor Relationship | CGCS Challenge: Colleague Relationship | CGCS Belief: Civil Service is Meritocratic | CGCS Belief: Hardwork Pays Off |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| Masking | -0.030** | -0.003 | 0.024** | 0.017* |
| | (0.014) | (0.009) | (0.012) | (0.009) |
| DV Mean | 0.15 | 0.05 | 0.90 | 0.94 |
| DV S.D. | 0.36 | 0.22 | 0.30 | 0.24 |
| Controls | Y | Y | Y | Y |
| County FE | Y | Y | Y | Y |
| Type FE | Y | Y | Y | Y |
| Cohort FE | Y | Y | Y | Y |
| Obs. | 2,839 | 2,839 | 2,839 | 2,839 |

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

**Table 9. Hometown Favoritism and Influence Activities**

| | Evaluator Assessment Score | | | Non-Evaluator Assessment Score | | |
|---|---|---|---|---|---|---|
| | Full Sample | Revealed Sample | Masked Sample | Full Sample | Revealed Sample | Masked Sample |
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Same Home Town | 0.102** | 0.192*** | -0.083 | -0.020 | -0.052 | 0.042 |
| | (0.052) | (0.068) | (0.089) | (0.049) | (0.060) | (0.101) |
| DV Mean | 5.16 | 5.13 | 5.21 | 5.11 | 5.04 | 5.26 |
| DV S.D. | 1.13 | 1.14 | 1.10 | 1.10 | 1.09 | 1.11 |
| Controls | Y | Y | Y | Y | Y | Y |
| County FE | Y | Y | Y | Y | Y | Y |
| Type FE | Y | Y | Y | Y | Y | Y |
| Cohort FE | Y | Y | Y | Y | Y | Y |
| Obs. | 2,291 | 1,549 | 686 | 2,274 | 1,542 | 676 |

*Notes*: Each column represents a separate regression. Standard errors clustered at work unit level are reported below the coefficients. Columns (1) and (4) use the full sample of CGCSs, Columns (2) and (5) use the sample of CGCSs in the revealed scheme, Columns (3) and (6) use the sample of CGCSs in the masked scheme. * significant at 10% ** significant at 5% *** significant at 1%.

**Table 10. 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. (Revealed) | 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 |
| *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 Meets the CGCS More Often |
| Supervisor 1 Eva. (Revealed) | -0.010 | 0.649 | -0.020 |
| | (0.019) | (0.431) | (0.042) |
| Obs. | 1,910 | 1,910 | 1,910 |
| *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. (Revealed) | 0.000 | -0.031* | 0.022 |
| | (0.019) | (0.017) | (0.020) |
| Obs. | 1,910 | 1,910 | 1,910 |
| County FE | Y | Y | Y |
| Type FE | Y | Y | Y |
| Enrol Year FE | Y | Y | Y |

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

**Table 11. Alternative Explanations to the Improved Performance**

| | (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 | -0.024 | 0.006 | 0.059 | 0.080 |
| | (0.015) | (0.021) | (0.036) | (0.063) |
| Obs. | 2,837 | 2,839 | 2,835 | 2,835 |
| *Panel B. Masking the Identity Does Not Lead to Information Difference* | | | | |
| | Evaluator Information | | Non-Evaluator Information | |
| Masking | 0.014 | 0.013 | 0.000 | -0.018 |
| | (0.020) | (0.016) | (0.019) | (0.016) |
| Information from | CGCSs | Colleagues | CGCSs | Colleagues |
| Obs. | 2,839 | 2,839 | 2,839 | 2,839 |
| County FE | Y | Y | Y | Y |
| Type FE | Y | Y | Y | Y |
| Cohort FE | Y | Y | Y | Y |

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

# Appendix to "Performance Evaluation, Influence Activities, and Bureaucratic Work Behavior: Evidence from China"

## APPENDIX A

**Table A1. Characteristics of CGCSs' Colleagues and Balance Checks**

| | Revealed Scheme | Masked Scheme | Difference |
|---|---|---|---|
| | (1) | (2) | (3) |
| Colleague Age | 34.568 | 34.401 | -0.28 |
| | (8.993) | (8.780) | (0.26) |
| Colleague Gender | 0.571 | 0.568 | -0.01 |
| (=1 if Female) | (0.495) | (0.495) | (0.01) |
| Colleague Education | 3.467 | 3.444 | -0.02 |
| | (0.721) | (0.700) | (0.02) |
| Colleague Tenured | 0.730 | 0.732 | 0.00 |
| | (0.444) | (0.443) | (0.01) |
| Meet Frequency with CGCS | 4.745 | 4.759 | 0.01 |
| Weekly | (0.731) | (0.691) | (0.02) |
| Know CGCS Well (Work) | 9.272 | 9.305 | 0.02 |
| (0-10) | (1.257) | (1.229) | (0.03) |
| Know CGCS Well (Life) | 8.300 | 8.383 | 0.07 |
| (0-10) | (2.046) | (1.998) | (0.06) |
| Colleague Self Assessment | 4.445 | 4.500 | 0.05* |
| (1-7) | (1.215) | (1.206) | (0.03) |
| Obs. | 6,374 | 2,981 | 9,355 |

*Notes*: Column (1) summarizes the mean and standard deviation of colleagues' characteristics in the revealed scheme. Column (2) summarizes the mean and standard deviation of colleagues' characteristics in the masked scheme. Column (3) checks the covariate balances between the revealed group and the masked group. 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 A2. Characteristics of Supervisors and Balance Checks**

| | Revealed Scheme | Masked Scheme | Difference |
|---|---|---|---|
| | (1) | (2) | (3) |
| Evaluator Gender | 0.234 | 0.205 | -0.021 |
| (=1 if Female) | (0.423) | (0.404) | (0.019) |
| Evaluator Age | 43.438 | 43.119 | -0.292 |
| | (7.544) | (7.304) | (0.347) |
| Evaluator Work Experience | 7.085 | 6.988 | 0.004 |
| (Years) | (3.350) | (3.459) | (0.167) |
| Evaluator Education | 4.656 | 4.700 | 0.035 |
| | (0.601) | (0.574) | (0.028) |
| Evaluator Title | 0.550 | 0.520 | -0.040 |
| (=0 if Party, =1 if Admin) | (0.498) | (0.500) | (0.026) |
| Non-Evaluator Gender | 0.240 | 0.226 | -0.005 |
| (=1 if Female) | (0.427) | (0.419) | (0.020) |
| Non-Evaluator Age | 43.327 | 43.148 | -0.095 |
| | (7.990) | (7.596) | (0.358) |
| Non-Evaluator Work Experience | 43.327 | 43.148 | -0.095 |
| (Years) | (7.990) | (7.596) | (0.358) |
| Non-Evaluator Education | 4.690 | 4.654 | -0.025 |
| | (0.601) | (0.585) | (0.029) |
| Non-Evaluator Title | 0.583 | 0.563 | -0.005 |
| (=0 if Party, =1 if Admin) | (0.493) | (0.497) | (0.030) |
| Obs. | 1,935 | 919 | 2,854 |

*Notes:* Column (1) summarizes the mean and standard deviation of supervisors' characteristics in the revealed scheme. Column (2) summarizes the mean and standard deviation of supervisors' characteristics in the masked scheme. Column (3) checks the covariate balances between the revealed group and the masked group. 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 the work unit level are reported in the parentheses.

**Table A3. Evaluator and Non-Evaluator Characteristics under the Revealed Scheme**

| | Evaluator (Revealed) | Non-Evaluator (Revealed) | Difference |
|---|---|---|---|
| | (1) | (2) | (3) |
| Gender | 0.240 | 0.234 | -0.017 |
| (=1 if Female) | (0.427) | (0.423) | (0.024) |
| Age | 43.327 | 43.438 | 0.220 |
| | (7.990) | (7.544) | (0.415) |
| Work Experience | 6.822 | 7.085 | 0.340* |
| (Years) | (3.396) | (3.350) | (0.205) |
| Education | 4.690 | 4.656 | -0.035 |
| | (0.601) | (0.601) | (0.034) |
| Title | 0.583 | 0.550 | -0.176 |
| (=0 if Party, =1 if Admin) | (0.493) | (0.498) | (0.225) |
| Obs. | 1,935 | 1,935 | 3870 |

*Notes:* We keep the subsample of all CGCS supervisors under the revealed scheme. Column (1) summarizes the mean and standard deviation of evaluating supervisors' characteristics. Column (2) summarizes the mean and standard deviation of non-evaluating supervisors' characteristics. Column (3) checks the covariate balances between the two groups controlling for CGCS FE. 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 the work unit level are reported in parentheses.

**Table A4. Test for Attrition**

| | Total Attrition | Re-assignment | Quitting |
|---|---|---|---|
| | (1) | (2) | (3) |
| Masking | -0.010 | -0.008 | 0.005 |
| | (0.016) | (0.011) | (0.009) |
| Constant | 0.249*** | 0.142*** | 0.076*** |
| | (0.016) | (0.010) | (0.007) |
| Obs. | 3,785 | 3,785 | 3,785 |
| R-Squared | 0.000 | 0.111 | 0.116 |

*Notes*: This table test if masking the identify of the evaluator affects the attrition of CGCSs. 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. Attrition by Baseline Covariates**

| | Attrition |
|---|---|
| | (1) |
| Age | 0.007 |
| (at Endline) | (0.006) |
| Gender | -0.035* |
| (=1 if Female) | (0.020) |
| Social Science Major | -0.025 |
| (=1 if Yes) | (0.015) |
| 4-Year College or Above | 0.068*** |
| (=1 if Yes) | (0.018) |
| STEM Students in High School | -0.024 |
| (=1 if Yes) | (0.021) |
| Party Member | -0.008 |
| (=1 if Yes) | (0.024) |
| Parent Completing College | 0.043** |
| (=1 if Yes) | (0.022) |
| Work in Village | 0.018 |
| (=1 if Yes) | (0.022) |
| CEE Score | -0.000 |
| (100 Points) | (0.000) |
| Risk Averse | -0.008 |
| (=1 if Yes) | (0.014) |
| Obs. | 3,778 |

*Notes*: This table regresses the attrition status of a CGCS on her baseline characteristics, controlling for county FE, CGCS type FE, and cohort FE, with standard errors clustered at the work unit level.

**Table A6. Masking and Colleague Assessments: Robustness Check**

| | Performance (1-7) | Top 10% | Hardworking |
|---|---|---|---|
| | (1) | (2) | (3) |
| Masking | 0.217*** | 0.077*** | 0.028** |
| | (0.035) | (0.013) | (0.012) |
| DV Mean | 5.22 | 0.70 | 0.44 |
| DV S.D. | 0.91 | 0.32 | 0.43 |
| Controls | Y | Y | Y |
| Female*Masking | Y | Y | Y |
| 4-Year College*Masking | Y | Y | Y |
| Parent College*Masking | Y | Y | Y |
| County FE | Y | Y | Y |
| Type FE | Y | Y | Y |
| Cohort FE | Y | Y | Y |
| Obs. | 2,837 | 2,837 | 2,837 |

*Notes*: Each column represents a separate regression. Standard errors clustered at work unit level are reported below the coefficients. Control variables include CGCS characteristics presented in Table 1. We further control for the interactions between the dummy variable for the masked scheme and the three pre-RCT covariates that are correlated with CGCS attrition. Standard errors are clustered at the work unit level. * significant at 10% ** significant at 5% *** significant at 1%.

**Table A7. Partial Correlations between CGCS Characteristics and Performance**

| | Performance (1-7) | |
|---|---|---|
| | by Colleague | Supervisor |
| | (1) | (2) |
| Age | 0.074*** | 0.074*** |
| | (0.010) | (0.016) |
| Gender | -0.055 | -0.085* |
| | (0.040) | (0.049) |
| Social Science | -0.018 | -0.028 |
| | (0.036) | (0.043) |
| 4-Year College | 0.222*** | 0.260*** |
| | (0.041) | (0.048) |
| STEM Students | -0.028 | 0.051 |
| | (0.037) | (0.044) |
| Party Member | 0.256*** | 0.250*** |
| | (0.042) | (0.053) |
| Parent High Sch. | 0.038 | 0.102** |
| | (0.035) | (0.044) |
| Parent College | -0.037 | 0.059 |
| | (0.040) | (0.048) |
| Work in Village | 0.042 | 0.154** |
| | (0.059) | (0.061) |
| CEE Score | 0.039 | 0.100*** |
| | (0.027) | (0.034) |
| Risk Averse | -0.033 | -0.036 |
| | (0.031) | (0.046) |

*Notes*: Each cell represents a separate regression between the outcome variable and the CGCS's certain characteristics. No control is included in any of these partial regressions. In Column (1), the outcome variable is the colleague assessments of CGCS performance; in Column (2), the outcome variable is the supervisor assessments of CGCS performance. Standard errors clustered at the work unit level are reported in parentheses. * significant at 10% ** significant at 5% *** significant at 1%.

**Table A8. Multi-Tasking under the Masked Scheme**

| | Ratio of CGCS's Job Tasks Assigned by Supervisor 1 | | CGCS's Most Important Job Task is Guided by Supervisor 1 | | CGCS Improved More in Areas Deemed Important by Sup 1 (Relative to Sup 2) | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Supervisor 1 Evaluating | -0.016 | -0.011 | -0.042 | -0.041 | -0.020 | -0.026 |
| *(Masked)* | (0.021) | (0.021) | (0.049) | (0.050) | (0.084) | (0.084) |
| DV Mean | 0.49 | 0.49 | 0.51 | 0.51 | -0.05 | -0.05 |
| DV S.D. | 0.24 | 0.24 | 0.47 | 0.47 | 0.89 | 0.89 |
| Controls | N | Y | N | Y | N | Y |
| County FE | Y | Y | Y | Y | Y | Y |
| Type FE | Y | Y | Y | Y | Y | Y |
| Cohort FE | Y | Y | Y | Y | Y | Y |
| Obs. | 659 | 659 | 529 | 529 | 580 | 580 |

*Notes:* Each column represents a separate regression. The sample is for the revealed scheme only. Standard errors clustered at the work unit level are reported below the coefficients. Control variables include CGCS characteristics presented in Table 1, and colleague characteristics presented in Appendix Tables A1. * significant at 10% ** significant at 5% *** significant at 1%.

**Table A9. Unexplained Supervisor Positiveness**

| | Residualized Supervisor Positiveness | |
|---|---|---|
| | (1) | (2) |
| Masked | -0.104** | -0.098** |
| | (0.042) | (0.043) |
| Controls | N | Y |
| County FE | Y | Y |
| Type FE | Y | Y |
| Cohort FE | Y | Y |
| Obs. | 2,037 | 1,935 |
| R-Squared | 0.145 | 0.150 |

*Notes*: Each column represents a separate regression. "Residualized Supervisor Positiveness" is the residual obtained from regressing evaluator assessment on non-evaluator assessment and colleague assessments. Control variables include CGCS characteristics listed in Table 1. Standard errors clustered at the work unit level are reported below the coefficients. * significant at 10% ** significant at 5% *** significant at 1%.

**Table A10. Hometown Favoritism among Colleagues**

| | Colleague Assessment Score | | |
|---|---|---|---|
| | Full Sample | Revealed Sample | Masked Sample |
| | (1) | (2) | (3) |
| Same Home Town | 0.051* | 0.054 | 0.056 |
| (Colleague) | (0.028) | (0.034) | (0.051) |
| DV Mean | 5.20 | 5.13 | 5.35 |
| DV S.D. | 1.22 | 1.23 | 1.19 |
| Controls | Y | Y | Y |
| County FE | Y | Y | Y |
| Type FE | Y | Y | Y |
| Cohort FE | Y | Y | Y |
| Obs. | 9,252 | 6,286 | 2,954 |

*Notes*: Each column represents a separate regression. Standard errors clustered at work unit level are reported below the coefficients. Columns (1) uses the full sample of CGCSs, Columns (2) uses the sample of CGCSs in the revealed scheme, Columns (3) uses the sample of CGCSs in the masked scheme. * significant at 10% ** significant at 5% *** significant at 1%.

**Table A11. Discouragement Effect**

| | Colleague Assessment Score | |
|---|---|---|
| | Full Sample | Masking vs. Being Evaluated by Preferred Leader |
| | (1) | (2) |
| Masking | 0.220*** | 0.247*** |
| | (0.033) | (0.038) |
| County FE | Y | Y |
| Type FE | Y | Y |
| Enrol Year FE | Y | Y |
| Obs. | 9,256 | 6,206 |
| R-Squared | 0.130 | 0.158 |

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

## APPENDIX B

*Sample notification letter (Revealed Scheme):*

Dear Mr. [redacted]:

Greetings!

Per the request of the provincial human resources department, we, a research team based at Renmin University in China, will be conducting a "third-party evaluation" of CGCS performance in this fiscal year. The results of this third-party evaluation will be used by the provincial human resources department for decision making.

In June 2018, we will send a team of enumerators to your work unit ([redacted] department in [redacted] township), to collect information about your work performance in the past year. **Specifically, among your two supervisors, Mr. [redacted] and Mr. [redacted], we have randomly selected Mr. [redacted] to be the evaluator. We will collect his assessments of your work performance by the end of the evaluation cycle, and provide that information to the provincial human resources department.**

The performance information will be used only by the research team and the provincial human resources department. Under no circumstance will we provide your personal information to other irrelevant parties. If you have any questions, please contact us at:

Email: [redacted]

WeChat: [redacted]

Phone: [redacted]

Regards,

Renmin University, School of Public Administration

*Sample notification letter (Masked Scheme):*

Dear Mr. [redacted]:

Greetings!

Per the request of the provincial human resources department, we, a research team based at Renmin University in China, will be conducting a "third-party evaluation" of CGCS performance in this fiscal year. The results of this third-party evaluation will be used by the provincial human resources department for decision making.

In June 2018, we will send a team of enumerators to your work unit ([redacted] department in [redacted] township), to collect information about your work performance in the past year. **Specifically, among your two supervisors, Mr. [redacted] and Mr. [redacted], we will randomly select one of them to be the evaluator. We will collect this evaluator's assessments of your work performance by the end of the evaluation cycle, and provide that information to the provincial human resources department.**

The performance information will be used only by the research team and the provincial human resources department. Under no circumstance will we provide your personal information to other irrelevant parties. If you have any questions, please contact us at:

Email: [redacted]

WeChat: [redacted]

Phone: [redacted]

Regards,

Renmin University, School of Public Administration

## APPENDIX C

The CGCS is maximizing her utility function, which consists of positive returns to her evaluation scores and negative returns to her efforts in productive and influence activities. Therefore, there exists a scenario in which within the range of our variables, the cost function of the "common productive tasks" is more convex than the cost functions of "influence activities," so that under the masked scheme, when the CGCS reallocates her efforts from influence activities to the common productive tasks, she will receive higher assessment score from her evaluating supervisor, but also experience a reduction in her overall utility.

To illustrate this possible equilibrium, we construct a concrete numerical example. Specifically, let $T = 2,\ b = 0.145,\ \alpha = 0.3,\ G(X) = \frac{10}{11}X^{\frac{11}{10}},\ g(x) = \frac{1}{11}x^{11}$. Given these parametric values, in the revealed scheme, we have: $X^{reveal} = G'^{-1}(1-\alpha+b) = 0.1856$; $x_1^{reveal} = g'^{-1}(1-\alpha+b) = 0.9833$ ; $u_1^{reveal} = T - G'^{-1}(1-\alpha+b) - g'^{-1}(1-\alpha+b) = 0.8311$; and $E_1^{reveal} = \alpha T + (1-\alpha)G'^{-1}(1-\alpha+b) + (1-\alpha)g'^{-1}(1-\alpha+b) = 1.4182$. In contrast, in the masked scheme, we have: $X^{mask} = G'^{-1}\left(1-\frac{\alpha}{2}+b\right)$; $x_1^{mask} = g'^{-1}\left(\frac{1}{2}-\frac{\alpha}{2}+b\right) = 0.9321$ ; $u_1^{mask} = T - G'^{-1}\left(1-\frac{\alpha}{2}+b\right) - g'^{-1}\left(\frac{1}{2}-\frac{\alpha}{2}+b\right) = 0.1167$; and $E_1^{mask} = \frac{1}{2}\alpha T + \left(1-\frac{1}{2}\alpha\right)G'^{-1}\left(1-\frac{\alpha}{2}+b\right) + \left(\frac{1}{2}-\frac{\alpha}{2}+b\right)g'^{-1}\left(\frac{1}{2}-\frac{\alpha}{2}+b\right) = 1.4347$.

As we can see, in this numerical example, we have: $E_1^{mask} > E_1^{reveal}$; $x_1^{mask} < x_1^{reveal}$ and $u_1^{mask} < u_1^{reveal}$, which are consistent with the empirical patterns documented in our experimental data.