Online Appendix: Information Systems, Service Delivery, and Corruption: Evidence From the Bangladesh Civil Service

By Martin Mattsson *

A. Theory

1. *Model of Reputation and Bureaucrat Behavior*

In this Appendix, I present the formal model described in Section V.

Model Setup. — I assume that bureaucrats have the following utility function:

$$U(E_i, B_i) = D(E_i) + M(B_i) + t_i R(B_i, v_i P(E_i)) \tag{A1}$$

where the subscript i represents an individual bureaucrat. The disutility of effort $D(E)$ is a strictly concave function decreasing in effort E. The utility of bribe money $M(B)$ is a strictly concave function increasing in bribes B. $R(B, vP(E))$ is the utility the bureaucrat receives from reputation or pride in her work; it depends on B and visible performance $vP(E)$. The performance function $P(E)$ is concave and increasing in effort E. v is the visibility of performance; I assume v increases with the scorecards. v is not binary, so the model allows for performance to matter for reputation even without the scorecards, as long as the scorecards *increase* the visibility. $R(.)$ is a strictly concave function decreasing in bribes ($R_1(B, vP(E)) < 0$) and increasing in visible performance ($R_2(B, vP(E)) > 0$). The intuition for why bribes reduce the bureaucrat's reputation is that for bureaucrats consistently asking for higher-than-normal bribes, a reputation will build over time that this bureaucrat is corrupt. Similarly, if the bureaucrat consistently outperforms peers in terms of processing times, this will build a good reputation over time.

I assume that the only innate difference between bureaucrats is the weight t put on the reputation term in their utility function; this is what generates over- and underperformers in the model. It is not essential for the predictions of the model that the differences in performance are generated by differences in t; it could also have been generated by differences in ability or some other factor. I choose to model the differences as differences in the valuation of reputation because it is consistent with the observation that overperforming bureaucrats collect fewer bribes than underperforming bureaucrats in the control group, as shown in Table 3.

An important assumption is that bribes and visible performance are complements in generating reputation, i.e. that $R_{12}(B, vP(E)) > 0$. I discuss the basis for this assumption in Section V.A.

* National University of Singapore (email: martin.mattsson@nus.edu.sg).

Another important assumption is that $P(E)$ can be both positive and negative. Specifically, I assume that $P(E)$ is positive when E is above median effort and negative when E is below median effort. This is realistic when performance is measured in a relative way, as it is in the scorecards' percentile rankings. This means that for bureaucrats performing above the median, increased visibility improves their visible performance, while for those performing below the median, it decreases visible performance by making it more negative.

As discussed in Section V.A, I do not formally model applicants' behavior, but I assume that they have no choice but to accept the bureaucrats' bribe requests.

Abstractions From Reality. — In the model, the applicants simply pay the bribe amount that the bureaucrats are demanding. While this is a simplification, Appendix Figure D2 shows that even the largest estimate for the average bribe is just 0.12% of the average stated land value. Furthermore, when asked directly about the value of the record of rights, the largest estimate for the average bribe is just 0.14% of the average stated value of the record of rights.[1] This indicates that applicants' willingness to pay for the service is not an important determinant of the bribe value. Instead, what determines the amount of bribes in the model is the bureaucrat's trade-off between bribe money and reputational concerns.

The model also abstracts away from bureaucrats buying reputation or career advancement using money. This avoids bureaucrats taking as-high-as-possible bribes and then using the money to regain their reputation. The results would be the same if bureaucrats would pay for their position in the first place but then could not bribe their way to future career advancement.[2]

Solution to the Bureaucrats' Problem. — Bureaucrats choose E and B to maximize $U(E, B)$. The first-order conditions to the bureaucrat's maximization problem are:

$$D'\left(E_i^*\right) + t_i R_2\left(B_i^*, v_i P\left(E_i^*\right)\right) v_i P'\left(E_i^*\right) = 0 \tag{A2}$$

where E_i^* and B_i^* represent the choices of E and B that maximize utility for bureaucrat i:

$$M'\left(B_i^*\right) + t_i R_1\left(B_i^*, v_i P\left(E_i^*\right)\right) = 0 \tag{A3}$$

At the optimum, the marginal disutility of effort equals the marginal utility of effort's effect on reputational concerns (Equation A2), and the marginal utility of bribe money equals the marginal disutility from a decrease in reputation due to bribes (Equation A3).

[1] 56% of respondents stated that the value of the record of rights is as high as the value of the land itself, indicating that they used the land value as a proxy for the value of the record of rights.

[2] Weaver (2021) analyses the effects of such bribes in the allocation of job applicants to positions in public-service delivery.

Effect of Scorecards on Effort and Performance. — Differentiating the first-order conditions with respect to v and solving for $\frac{\partial E_i^*}{\partial v}$ provides the expression in Equation A4. For ease of exposition, I henceforth drop the parentheses describing the variables that each function depends on, the subscript i, and the star superscript. All mentions of E and B refer to the values at the optimum.

$$\frac{\partial E}{\partial v} = P' \frac{-R_2 \left(M'' + tR_{11}\right) + vP\left(t\left(\left(R_{12}\right)^2 - R_{11}R_{22}\right) - M''R_{22}\right)}{\left(M'' + tR_{11}\right)\left(\frac{D''}{t} + R_{22}\left(vP'\right)^2 + R_2 vP''\right) - t\left(R_{12}vP'\right)^2} \tag{A4}$$

Using the assumptions that $M(.)$,$D(.)$, and $R(.)$ are concave, the denominator is always positive. The first term in the numerator of Equation A4 reflects the incentive effect of the changed visibility of effort. I.e., it is the effect the scorecards have through changing the visibility of performance, holding the reputation level constant. This effect is always positive. The second term in the numerator represents the reputation-level effect. I.e., the effect on effort stemming from the changed reputation level due to the change in visibility of the performance that was there before the scorecards.[3] If $P < 0$, then this term is also positive, making the whole expression positive. Since effort is the sole determinant of performance, an increase in effort also implies a performance improvement.

Prediction 1: If $P(E) < 0$ then $\frac{dE}{dv} > 0$

The overall effect is ambiguous when $P(E) > 0$ since the reputational level effect is negative and may be larger or smaller than the incentive effect. The ambiguous effect is analogous to the ambiguous effect of a wage increase on labor supply. The income effect and substitution effect go in different directions and depending on which dominates, the overall effect may be positive or negative.

Effect of the Scorecards on Bribes. — Differentiating the first-order conditions with respect to v and solving for $\frac{\partial B_i^*}{\partial v}$ gives us the following expression:

[3] To define the incentive and reputation-level effects formally, I introduce two types of visibility, $\underline{v}$, which determines the level of reputation, and $\bar{v}$, which determines the incentive to improve performance. I then rewrite the FOCs as:

$$D'(E) + tR_2\left(B, \underline{v}P(E)\right)\bar{v}P'(E) = 0$$

$$M'(B) + tR_1\left(B, \underline{v}P(E)\right) = 0$$

The incentive effect is:

$$\frac{\partial E^*}{\partial \bar{v}} = P' \frac{-R_2\left(M'' + tR_{11}\right)}{\left(M'' + tR_{11}\right)\left(\frac{D''}{t} + R_{22}\left(vP'\right)^2 + R_2 vP''\right) - t\left(R_{12}vP'\right)^2} > 0$$

The reputation-level effect is:

$$\frac{\partial E^*}{\partial \underline{v}} = P' \frac{vP\left(t\left(\left(R_{12}\right)^2 - R_{11}R_{22}\right) - M''R_{22}\right)}{\left(M'' + tR_{11}\right)\left(\frac{D''}{t} + R_{22}\left(vP'\right)^2 + R_2 vP''\right) - t\left(R_{12}vP'\right)^2}$$

$$\frac{\partial B}{\partial v} = R_{12}\frac{tvR_2\left(P'\right)^2 - P\left(D''\left(E\right) + tvR_2P''\right)}{\left(M'' + tR_{11}\right)\left(\frac{D''}{t} + R_{22}\left(vP'\right)^2 + R_2vP''\right) - t\left(R_{12}vP'\right)^2} \tag{A5}$$

The first term in the numerator is the impact of the incentive effect increasing effort. This leads to an improvement in visible performance, which in turn leads to an increase in bribes because of the complementary between visible performance and bribes. The second term is the impact of the reputation-level effect, which also affects bribes due to the complementarity between visible performance and bribes.

For bureaucrats with a performance above the median ($P\left(E\right) > 0$) the overall effect is positive. Their positive performance becomes more visible, so visible performance increases and the complementarity decreases the marginal disutility, through the reputation channel, from collecting bribes. Conversely, for bureaucrats with below-median effort, this effect is negative. Hence, for bureaucrats with above-median effort, the scorecards lead to higher bribes ($\frac{dB}{dv} > 0$), while for bureaucrats with below-median effort, the effect is ambiguous.

Prediction 2: If $P\left(E\right) > 0$ then $\frac{dB}{dv} > 0$

2. *Theories of a Causal Relationship Between the Speed of Service Delivery and Corruption*

There are several reasons why bribes may be causally related to processing times. In the literature on this relationship, two opposing views lead to drastically different policy conclusions. One view is that bribes "grease the wheels" by providing incentives to bureaucrats and allowing for excessive red tape to be circumvented (Leff, 1964; Huntington, 1968). In this view, rooting out corruption would decrease the speed of service delivery and increase inefficiencies of excessive bureaucratic control. On the other hand, improving the speed of service delivery for everyone through better technology or more personnel would decrease bribery, because the need for bribes would decrease.

Another view is that opportunities for corruption are the reason for excessive red tape and delays in public services, as additional hurdles allow officials to extract more bribes in exchange for avoiding these hurdles (Rose-Ackerman, 1978; Kaufmann and Wei, 1999; Mattsson, 2023). Under these circumstances, providing the bureaucracy with more resources would not improve processing times, because the bureaucrats are intentionally slowing down service delivery. However, service delivery could be improved by eliminating corruption and corruption could be reduced by making it more costly for bureaucrats to slow down service delivery. Below I empirical implications of one such theory.

Monopolistic Price Discrimination Model. — The scorecard experiment was originally designed to test a model of how opportunities for corruption may slow down service delivery. Mattsson (2023) provides a full exposition of the model. For a more detailed description of the model's predictions with regards to the experimental intervention in this paper, please see the pre-analysis plan. Here I outline the intuition behind the model and its predictions for the experiment.

The model is based on an asymmetric information model of price discrimination under monopoly, where the bureaucrat is the monopolist (Mussa and Rosen, 1978). Applicants get utility from having their application processed; the faster the application is processed, the more utility the processing generates. Applicants differ only in their willingness to pay for the speed of processing their applications. The bureaucrat can ask for different bribe payments from the applicants and can offer the service with different processing times or refuse to provide the service. Once a processing time and bribe payment are agreed upon, the applicant pays the bribe and the bureaucrat must honor the agreement. The bureaucrat gets utility from receiving bribes. It can be costly for the bureaucrats to process applications faster, although this is not necessary for the main conclusions of the model.

Perfect information means that the bureaucrat can perfectly observe the applicants' willingness to pay for having the application processed within a certain time. Under perfect information, applicants will have their applications processed at a Pareto optimal speed, where the marginal benefit of having the application processed faster is the same as the marginal cost of processing the application faster for the bureaucrat.

In an asymmetric information setting the bureaucrat cannot observe each individual applicant's willingness to pay, but the bureaucrat knows the distribution of willingness to pay across all applications. Thus, the bureaucrat has to offer the same menu of processing times and bribe payments to all applicants. Under asymmetric information, only the applicants with the highest willingness to pay get their application processed at the Pareto optimal speed. All other applicants have their applications slowed down, as the bureaucrats trade-off providing faster processing for applicants with lower willingness to pay, against how large a bribe they can charge from applicants with a higher willingness to pay.

A simple example is where it is costless for the bureaucrat to process the application immediately and there are two types of applicants, one with a higher willingness to pay for fast processing. Under full information, the bureaucrat can simply make a take-it-or-leave-it offer to all applicants at exactly their willingness to pay to have the application processed immediately. The applicants will pay their respective willingness to pay because they have no better outside option, and the bureaucrat will process the applications immediately.

Under asymmetric information, the bureaucrat cannot ex-ante differentiate between applicants of different types. It is therefore optimal, from the bureaucrat's perspective, to offer to process the application immediately at a higher bribe payment and slower at a lower bribe payment. The applications for those with low willingness to pay are now intentionally delayed, despite that processing them immediately does not cost the bureaucrat anything.

The Price Discrimination Model's Predictions for the Experiment. — The scorecards encourage bureaucrats to process applications within 45 working days. Section III shows that the scorecards led to an increase in the number of applications processed within 45 working days and that the effect was driven by offices that were underperforming at the start of the experiment. Under full information, an increase in processing speed is predicted to lead to a slight increase in bribe payments among those whose applications were processed faster. This is because for these applications, the value of the processing increased and they are now willing to pay more for it. Under asymmetric information, an increase in processing speeds is predicted to reduce the bribe payments among

those with the highest willingness to pay for getting their applications processed quickly. This is because the bureaucrat has to make the menu option of having applications being processed quickly more attractive in order for these applicants to continue to pay for it, now that the processing speed of the option to pay less has increased.

Testing the Model's Predictions . — The main testable prediction of this model under asymmetric information is that when processing times are improved, bribe payments should decrease among those who are getting their application processed quickly. Appendix Table D5 shows that the results from the experiment are inconsistent with this prediction. Column (1) shows the effect on bribes among those who had their applications processed quickly, using 25 working days as the cutoff. Among these applicants, the scorecards increased bribe payments by BDT 477. Column (2) shows that for applications processed outside of 25 working days, the estimated effect was a BDT 390 increase, but this effect is not statistically significant. Column (3) shows that even for offices that were underperforming at baseline, the effect of the scorecards on bribes for applications processed within 25 working days is estimated to be an increase of BDT 375. Although the effect is not statistically significantly different from zero, a large negative effect can be rejected.

One potential explanation for the results within the framework of the monopolistic price discrimination model is that government officials have full information about applicants' willingness to pay for processing speed. However, even under full information, the bribe payments are not predicted to increase for those with the highest willingness to pay, and not among the overperforming offices that did not change processing times. The effects of the scorecards on bribes shown in Appendix Table D5 are therefore not consistent with the predictions of the model under any information setting.

Another explanation for why the results are inconsistent with the predictions of the model is that information about the increase in processing speeds had not yet been disseminated to applicants by the time of the survey. The information treatment is designed to alleviate this problem. However, Column (4) of Appendix Table D5 shows that even for applicants who received the information in offices that were underperforming at baseline, no negative effect on bribes can be found for applications processed within 25 working days.

Implications for Other Theories . — The increase in bribes among the offices that were overperforming at baseline is also inconsistent with models where it is applicants' outside option or ability to pay that determines the bribe levels (e.g. Svensson, 2003; Niehaus and Sukhtankar, 2013*b*). If the bribe level was fully determined by the applicants' outside option or ability to pay, it could not have been increased by a positive scorecard, without any change in service quality. This result is probably dependent on the market structure of the interaction in which the bribe is paid. In this context, the land office is the only institution that can make the required land-record change, and there are no close substitutes for this service. Had there been competition for applicants between land offices, or a close alternative to a land-record change, it is plausible that outside options would have been more important in determining the bribe level (Svensson, 2003; Bai et al., 2019).

A different class of models exists, in which government officials could extract more bribes if they wanted to, but they choose not to do so because they face a trade-off between taking bribes and achieving some other objective. This trade-off could be between bribe money and the risk of getting

caught (Becker and Stigler, 1974; Olken, 2007; Niehaus and Sukhtankar, 2013*a*), but it could also be between bribes and altruistic motivations or reputation concerns (Do, Van Nguyen and Tran, 2021). The model described in Section V is within this class of models.

B. Additional Details on the Experiment and Data

1. *Associations between Performance, Corruption, and Promotions*

Appendix Table D21 shows the associations between performance, corruption, and tenure on one side, and promotions and unattractive postings on the other. Overall the Table shows suggestive evidence that delays in promotions and unattractive posting are correlated with bureaucrats being underperforming and corrupt, although the associations are weak and only a few are statistically significant. I measure underperformance as 100 minus the average percentile ranking of the land office during the bureaucrats' tenure in the land office, regardless if that was before, during, or after the scorecard experiment. The measure goes from zero, if the bureaucrat was the top performer on both performance indicators each month of their tenure, to one hundred, if the bureaucrat was the worst performer on both performance indicators each month of their tenure. The median underperformance is 44 and the difference between the 75th and the 25th percentile is 20. The two bribe measures used are the averages of the typical and reported bribes from the survey data expressed in thousands of BDT.

In Panel A, the outcome variable is if the Bureaucrat was promoted before March 2020. In Panel B, the outcome is if the bureaucrat was ever suspended or placed as an "Officer on Special Duty" (OSD) at the Ministry of Public Administration during the time period after the start of the experiment. OSD postings are when a bureaucrat does not have any direct responsibilities but is waiting to get a real posting. OSD postings are typically seen as a punishment and negative for one's career within the civil service.[4] In media reports, bureaucrats have described being an OSD as a "punishment" , and "a curse for my professional life."[5] Anecdotally, OSD postings are associated with the bureaucrat being incapable of satisfactorily performing duties, having been caught with some type of misconduct, or being perceived as insufficiently loyal towards the political leadership. However, there are also other reasons for being OSD, such as a lack of suitable open postings for the officer's rank.

Column (1) of Appendix Table D21 shows that underperformance, as measured by the performance scorecards, has a negative association with being promoted and a positive association with being suspended or placed as an OSD. Although it is only the association with being suspended or placed as an OSD that is statistically significantly different from zero, the magnitudes of the associations are substantial. For example, moving from the 25th to the 75th percentile of underperformance is associated with a reduction in the share of bureaucrats promoted of -4.3 percentage points (equivalent to a 13% increase in the risk of not being promoted) and an increase in the share of being suspended or placed on OSD of 5.8 percentage points (equivalent to a 98% increase).

[4] OSD is also used when a bureaucrat is taking a leave from the service for further studies or for other special appointments, I do not classify such positions as OSD as I want a measure of undesirable postings.

[5] https://www.dhakatribune.com/bangladesh/133721/how-does-an-osd-spend-his-day

Columns (2) and (3) of Appendix Figure D21 show that both measures of average bribes are negatively associated with promotion and positively associated with being suspended or placed as an OSD. Similarly to the associations with performance, the magnitudes are economically large but only one of the four associations is statistically significant. For example, moving from the 25th to the 75th percentile of average typical bribe payments is associated with a reduction in the share of bureaucrats promoted of -4.3 percentage points (equivalent to a 21% increase in the risk of not being promoted) and an increase in the share of being suspended or placed on OSD of 11.7 percentage points (equivalent to a 261% increase).

Columns (4) and (5) show that the interaction term between underperformance and corruption has a negative association with promotion and a positive association with being suspended or placed as an OSD, although only one of the four coefficients is statistically significantly different from zero. This is consistent with the model assumption that honesty and performance are substitutes and that it is worse being both corrupt and underperforming than the coefficients on underperformance and corruption alone would suggest.

Finally, Column (6) shows that tenure in the civil service, as measured by the year in which the bureaucrat joined the civil service, is strongly associated with promotion but not with the being with being suspended or placed as an OSD. This is expected as tenure in the civil service is the main determinant of promotions at the junior level of the civil service, as described in Section I.A.

The results in Appendix Table D21 come with two important caveats. First, the results are only associations and one should be careful not to interpret the coefficients as causal effects. Second, the results are noisy and despite the magnitudes of the coefficients being large, only a few are statistically significantly different from zero.

2. *Randomization of Scorecards Treatment Assignment*

Half of the land offices were randomly assigned to receive the scorecards. The randomization was done at the office level and implemented by the author. The first-wave randomization was done separately for the group of land offices classified by the government as having a complete implementation of the e-governance system at the start of the experiment and for the group with partial implementation at the start of the experiment. The randomization strata were created using: the number of applications processed within 45 working days in June and July 2018; and applications pending for more than 45 working days at the end of July 2018.

The second-wave randomization was done separately for the group of land offices having received above/below the median number of applications in February and March 2019. The randomization strata were created using: the number of applications processed within 45 working days in March 2019; and applications pending for more than 45 working days at the end of March 2019.

The peer performance list was randomized in the same way as the second wave of scorecard randomizations using data as of September 2019. The randomization only took place among the 155 land offices receiving the scorecards.

3. Randomization of Information Intervention to Applicants

The information intervention was randomly assigned to half of the survey days. The randomization was stratified by the weekday pairs. During the treated days all applicants taking the in-person survey received the intervention. Noncompliance with the treatment assignment is discussed in Appendix C.8.

4. Scorecards Design and Performance Indicators

The two performance indicators were chosen after multiple discussions with both the Ministry of Land and the Access to Information (a2i) agency of the Government of Bangladesh. The government logotypes in the design of the scorecard and the information we provided made it clear that the scorecards were created in collaboration with the relevant government agencies and endorsed by those agencies while also informing them that the project was in a trial phase and that a research team was involved in the implementation.

While working closely with the implementing partners in the Government of Bangladesh required several rounds of redesigning the scorecards and some compromises to build a consensus around the final design, it also ensured that the scorecards were designed similarly to how they would have been designed if the government agency had initiated the process themselves. This is valuable as the effect of the government's own design is that of the highest policy relevance.

The performance scorecards are based on applications made in the land office, regardless of which bureaucrat was assigned to the office when the application was made. Percentile rankings are calculated using the performance indicators for offices in both the treatment and control groups. If several offices had the same value of the indicator, they all received the highest percentile in the percentile range covered by the offices. A thumbs-up symbol is shown for rankings above the 60th percentile while a thumbs-down symbol is shown for rankings below the 40th percentile.

Both indicators are absolute and not relative to the total number of applications or the past performance of the land office. After deliberations, the Ministry of Land, a2i, and I chose to base the performance indicators on absolute numbers for four reasons. First, a relative number could have created incentives to not let applicants apply, and then focus on processing the applications of those allowed to apply. Second, a measure relative to the number of applications received was perceived as being unfair toward larger land offices processing more applications. Third, making performance indicators relative to the past performance of an office was perceived as unfair toward offices performing well before the start of the experiment and would not have been possible for the offices that had installed the e-governance system very recently. Fourth, we could not come up with an intuitive rule for which applications should be counted towards a monthly measure like " share of applications processed within 45 working days". This is because we only know this share once at least 45 working days have passed. Therefore, to create a monthly indicator with the share of applications processed within 45 working days, we would have to base the measure on applications made approximately 45-67 working days ago (assuming the month had 22 working days). But, if we were to base the indicator on those applications, an application processed just before the scorecard was issued but that was started less than 45 working days ago, would not enter the scorecard for

one or even two months. We worried that such a long lag between the bureaucrats' actions and the indicator shown in the scorecard would be confusing and cause the scorecards to be less effective.

Instead, we choose one absolute measure benefiting larger land offices (the number of applications processed on time) and another benefiting smaller land offices (the number of applications pending beyond the time limit). Appendix Sub-Figures D4b, D4c, and D4d show that this worked in the intended way and that the average percentile ranking is not correlated with the number of applications received.

How Much of the Performance is Determined by the Bureaucrat?. — An important question is to which extent bureaucrats can control the performance of their office and how much of it that is predetermined by factors outside the control of the bureaucrat, such as the location of the office and seasonal patterns. I address this question using the adjusted R-squared from fixed-effects regressions where the monthly average performance ranking is the outcome variable, for land offices that had more than one bureaucrat during the span of my data. Including fixed effects for the land office and month yields an adjusted R-squared of 0.44 suggesting that less than half of the variation in the performance score is determined by land office characteristics and seasonality. Including bureaucrat fixed effects increases the adjusted R-squared to 0.63. This suggests that bureaucrats have a significant level of control over their performance scores and that there is variation in bureaucrat performance that can explain a substantial part of the variation in performance scores.

5. *Data*

Administrative Data From the E-governance System. — The administrative data was transferred to the author each month from August 2018 until December 2020. The data includes all applications made in the e-governance system since its inception. Due to privacy concerns, personal identifying information was not shared.

To calculate the number of working days between the start of the application and the date it was processed, I use data on holidays in Bangladesh (Time and Date, 2021).[6] I then convert the number of calendar days to working days. For the very small number of applications that were processed within one day, I use the exact time of the application and processing to create a measure of what fraction of a day it took to process the application.

It would be very difficult for any individual bureaucrat to improve their performance scorecard by manipulating the administrative data. The data is stored on a central server that the bureaucrats do not have access to. While it would be possible to create fake applications in the e-governance system, to process these applications with an acceptance, the processing fee would have to be paid. Creating fake unprocessed applications would decrease, not increase, a bureaucrat's performance ranking.

During the training of bureaucrats in using the e-governance system, example applications were made in two land offices that had not yet installed the e-governance system, making it appear as if the e-governance system was active in these two offices. Thus, these offices were included in the

[6]The data was retrieved from https://www.timeanddate.com/holidays/bangladesh/.

first wave of randomization, one was assigned treatment and one control. In September 2018, I found out that these two offices had not yet installed the e-governance system, and I removed all applications from these offices. I also found out that some other applications in the e-governance system are the result of examples created in training. Using information provided about the dates of the training (Mattsson, 2019), I removed applications made before the first wave of randomization suspected to be the result of training. I did not remove any applications made after the start of the experiment.

In one case, the name of one land office was incorrectly entered into the e-governance database. This led to the Scorecard being sent to the wrong office during the first 6 months of the experiment, effectively treating a different office. This was detected during the expansion of the experiment when both of the offices were randomized into the treatment group. After the second wave of randomization, both offices received the correct treatment and scorecards containing the correct data. During the survey, we surveyed the office that had received the treatment and I use data from these surveys in my analysis. However, the main results are all virtually identical when excluding all data from these two offices from the analysis.

The administrative data contain the size of the land for which the land-record change was made. Some of the values in this variable are unrealistically high, most likely due to mistakes in what unit was used when inputting the land value. I clean this variable by setting any observation with a value of above 10 acres to missing as land transactions above this threshold are very rare. The largest land-record change in the survey data is 7.3 acres.

Survey Data . — The survey was carried out in two stages: in the first stage, 3,213 individuals approaching the land office for the purpose of a land-record change were surveyed in person. The follow-up interview was conducted by phone. Enumerators attempted to reach applicants by phone three times in a day and then made another three attempts the next day. If all attempts failed, the phone number was kept for another round of attempts one to two months later. Out of 3,213 in-person surveys 89% were successfully interviewed in the follow-up survey, yielding a final sample of 2,869 respondents. The average time between the in-person and phone interviews is 3.3 months.[7]

Two types of duplicate observations are excluded from the survey data. One comes from the same person being interviewed more than once, probably as a result of them visiting the land office multiple times during the survey period, or multiple applicants providing the same phone number to be contacted on for the follow-up survey, probably because these individuals were in the same households. These observations may refer to different applications for land-record changes by the same individual or individuals in the same household, but since it was not possible to distinguish between the applications in the phone surveys, I keep only the first follow-up phone survey for each phone number. The main results are similar when including multiple follow-up surveys from the same phone number.

The second type of duplicate excluded from the data stems from multiple calls being made as follow-ups to the same in-person interview. This is natural since many applicants were not reached on the first attempt. However, in some cases there were multiple interviews where the applicant

[7]Interviewees were given BDT 50 (USD 0.6) in the form of a mobile phone recharge for a completed in-person interview and BDT 100 (USD 1.2) for a completed phone interview.

started providing answers to interview questions but where the interview was redone because it was interrupted. I keep only the data from the final follow-up interview with a complete set of answers for each in-person interview.

Processing times in the survey data for applicants for whom the application was not processed at the time of the follow-up interview are imputed using the procedure described in Section I.D. For applicants who did not answer the question about their monthly income, I impute their income using the income predicted from a regression of income on per capita household expenditure.

Measuring Bribes. — Corruption is notoriously difficult to measure precisely (Olken and Pande, 2012). Due to the sensitive nature of the questions about corruption, I only asked applicants about bribes in the phone interview, not when interviewing them outside the land office. I asked about bribe payments in two different ways. The first question was phrased "How much do you think it is normal for a person like yourself to pay in order to get the mutation processed and receive the khatian? Include all extra payments or gifts to agents, government officials, and other individuals, but do not include the 1,150 taka official fee." This is the question I refer to as the "typical payment" and it is my preferred measure of bribe payments as it has fewer responses being zero, suggesting that more respondents were comfortable answering the question. The next question was phrased as, "Did you pay any fees or give any gifts to anyone working for the Upazila Land Office or Union Parishad Land Office?" if the applicant responded yes, the enumerator asked about to whom the payment was made and "How much taka did you pay or what was the monetary value of the gift that you paid to [recipient]?" The "reported payment" outcome variable is the sum of all such reported payments. I also asked why the payment was made, with the results being presented in Appendix Figure D3, this had the additional benefit of correcting individuals who may have misperceived the question as asking about how much they paid as an official fee.

Each one taka increase in the "reported payment" measure is associated with a highly statistically significant 0.73 taka increase in the "typical payment" measure. This association is consistent with the results of Reid, Kabasababu and Weigel (2019), who show–in the context of tool booth bribes in the D.R. Congo–that the bribe amount directly stated by drivers is highly correlated to the amount written on a piece of paper and placed in a cardboard box to remove "any direct embarrassment/social desirability associated with admitting to a bribe in the presence of the enumerator".

To further validate the magnitude of my bribe estimates in Appendix Figure D9, I compare them to an independent estimate created by Transparency International Bangladesh (TIB) as part of their nationally representative National Household Survey (Transparency International Bangladesh, 2016). There are two main reasons why this measure may be different from my estimate of the average bribe payments, other than random differences between samples. First, the TIB survey was done in person, potentially allowing enumerators to build more rapport with the respondents. Second, the survey was done in a nationally representative sample for the period between November 2014 and October 2015. In the TIB survey, 605 of the households had made applications for land-record changes, among these 57% reported having paid a bribe. The bribes reported in the TIB survey are on average higher than the bribes reported in the phone survey I conducted, but that difference shrinks substantially when excluding respondents reporting zero bribes. The average typical bribe payment reported in my survey lies between the average payment reported in the TIB

survey and the average nonzero response. Overall, it is reassuring that the two different measures are of similar magnitude, despite using different methodologies, covering different areas, and being done for different periods.

To validate my approach of asking about informal payments in both a direct and indirect way, and to assess which of these approaches generates the most accurate answer, I use the microdata underlying the Viet Nam Provincial Governance and Public Administration Performance Index (PAPI).[8] The PAPI data allows me to compare the responses to both direct and indirect questions about informal payments in the process of obtaining land rights in Vietnam. The survey also contains a list experiment which allows me to generate an estimate of the share of applicants making an informal payment, without respondents having to directly state if they made such an informal payment or not.

In particular the PAPI survey asks respondents an indirect question about if they agree, somewhat agree, or disagree with the statement "People have to pay bribes in order to obtain a land title". I code this variable as zero when the respondent disagrees with the statement, one-half when they somewhat agree, and one when they agree. The survey also asks respondent who bought, sold, or inherited land if they had done any land-use rights related administrative procedures. Respondents who had done such a procedure were asked directly about if they were able to do so without paying an informal fee and if they paid a middleman to help them. Finally, a list experiment was carried out where respondents were asked how many of 3 or 4 actions on a list they had done and it was randomly determined if "Paid informal charges" was added as an action on this the list or not.[9]

To make all the responses comparable I restrict the sample to the 2,344 respondents who responded to all three questions. In this sample, respondents' average answer to the indirectly asked question about people's bribe payments has an average response of 0.30 (61% disagree, 18% somewhat agree, and 21% agree), this figure can be compared to the 75% of respondents in my survey who provided a positive number when asked about the typical payment among people like themselves. When asked directly about if they paid a bribe or paid a middleman 15% of respondents in the PAPI data stated that they made such a payment, which can be compared to the 28% stating that they made a positive payment in my survey. Interestingly, in both surveys, the estimate from the direct question is about half of the estimate from the indirect question. Finally, using the list experiment I estimate that 23% of respondents made an informal payment, which is close to the midpoint between the two other estimates. If one considers the list experiment estimate to be the most reliable way to measure corruption, this suggests that the direct question underestimates corruption while the indirect question overestimates corruption, potentially because corruption perceptions are higher than actual corruption. If this is the case, it justifies my approach to provide data from both measurement methods throughout the paper, and suggests that actual bribe payments are likely somewhere in between the two measurements, which is also what the TIB survey suggests.

Data on Bureaucrat Positions and Promotions . — I use the administrative data from the e-governance system to determine which bureaucrat was assigned to which land office. I take user

[8] Special thanks to Edmund Malesky for providing this data. For more details about the methodology and data collection, see https://papi.org.vn.

[9] See Malesky, Gueorguiev and Jensen (2015) for a discussion of the list experiment technique of measuring corruption.

data from the e-governance system to separate ACLs from other users and then assign a particular ACL to an office if that ACL is the ACL making the largest number of updates to land-record change applications that month. If an office has no updates made by any ACL in a month, I do not assign any ACL to that office in that month, unless an ACL was assigned to that office both prior to and after that month, in which case I conclude that the ACL was assigned to that office without making any updates in that month.

To determine the position and designation of the bureaucrats, I scrape data from historical versions of the Ministry of Public Administrations for every available month between June 2014 and March 2023 using the Wayback Machine by the Internet Archive.[10] This generated a total of 384,056 bureaucrat by post by month observations. These observations were then matched with the bureaucrats in the experiment using names. In some cases, when a bureaucrat shares the same name as other bureaucrats, changed their name[11], or the spelling of their name varies, I used other sources of data, such as local government offices' websites or Facebook pages to confirm the identity and positions of the bureaucrats. Out of 305 bureaucrats, position and designation data was found for 302.

6. Differences from Pre-analysis Plan

The study was registered, and a PAP was published on the AEA RCT Registry on August 18, 2018.[12] As the administrative data was available almost immediately after the implementation of the scorecard intervention, it was important to publish the PAP before the start of the experiment. However, this had the drawback that some aspects of the experiment and data collection changed after the PAP was written. Here are the most important ways in which the paper differs from the PAP:

- The PAP focuses on testing the theory outlined in Appendix A.2. While rejecting this theory is one of the results in this paper, I also propose a model that can explain the results. This model, described in Section V, was not part of the PAP.

- The information treatment for survey respondents was not included in the PAP. While writing the PAP, I was hopeful that I would be able to collect data from applicants only via phone surveys. However, due to privacy concerns, I was not able to access phone numbers for applicants from the administrative data. As phone surveys need to be shorter than in-person surveys an information treatment would not have been feasible. Together with the decision to interview applicants in person, I also developed the information treatment, but this was after the publication of the PAP.

- The peer performance list variation of the scorecards is not included in the PAP. This intervention originated from discussions with my implementation partners in the government who were interested in seeing if there were ways in which we could modify the scorecards to

[10] https://wayback-api.archive.org/
[11] Name changes are common after for example marriage or after receiving a doctorate degree.
[12] www.socialscienceregistry.org/trials/3232

make them more effective. These discussions and the design and implementation of the peer performance list took place after the publication of the PAP.

- The data collection and analysis of bureaucrats' promotions are not included in the PAP. As the focus of the project shifted from the original model in Appendix A.2 to the model in Section V, it became more important to collect additional evidence for the results on bureaucrats' careers.
- The PAP describes the digitization of administrative data on applications made outside of the e-governance system. This turned out to be practically difficult and was canceled. Therefore, the analysis planned with this data could not be conducted.

The main tables of results that follow the PAP are Tables 2, 3, D5, D14, D22, D23, D24, and D25. These tables use the regressions specifications from Sections 4.3.1, 4.3.3. and 4.4 in the PAP. Below I will describe the differences between these tables and the PAP.

Difference from PAP	Justification	Additional notes
An additional 199 offices were added to the experiment in the second randomization wave.	Adding these offices increases the sample size at a small cost to the overall project budget.	Appendix Table D24 shows the effect for 112 offices that were part of the original sample.
When imputing the processing time for unprocessed applications, the PAP describes setting " processing time equal to the average processing time for that [land office] for applications that have been pending for as long as that application" . Instead, I impute the value by taking the average processing time for *all* applications that have been pending for as long as that application.	In some land offices, there are no processed applications with processing times longer than the applications that have been pending for the longest time.	The results using the office-level imputation described in the PAP can be found in column (4) of Appendix Table D15.

Difference from PAP	Justification	Additional notes
The PAP does not describe how observations should be weighted. I weight applications by the inverse of observations in that land office.	With these weights, the regressions estimate the average effect of the scorecards on an office. As there is only one bureaucrat per office, this is the relevant unit for studying changes in bureaucrat behavior. This also makes the analysis of the administrative and survey data comparable. Finally, the weights improve the standard errors of the estimates. See Section II.A for more details.	A uniformly weighted analysis at the office level is shown in Appendix Table D2. Uniformly weighted estimates at the application and respondent level are shown in Appendix Tables D10 and D11.
The PAP only describes the " reported" bribes variable, but not the "typical" bribes variable.	I added the typical bribes question to the survey because of an implausibly high number of respondents answering zero to the reported bribes question.	Throughout the paper, I show the results using both measures of bribe payments.
Satisfaction was not part of the PAP, but I added it to Appendix Table D23.	Satisfaction is an important outcome as it holistically measures the applicant's subjective experience.	
The PAP describes analyzing the hours spent traveling to visit the land office, but I do not include this analysis.	This variable is closely related to the number of visits, which I show the results for.	
I did not collect data on hours spent on the preparation of the application or travel costs.	To shorten the phone survey.	
The PAP describes analyzing the first six months of data separately for each month to measure short-term vs. medium-term effects. In Figures 4, I instead show the difference between treatment and control continuously, using minimally processed data. To measure short-term, medium-term, and long-term effects, I also split the experiment into three parts in Appendix Figure D6.	When writing the PAP I expected the length of the study to be six months, as the experiment went on for longer I used a longer time frame to study the effects over time.	
I do not calculate the Family-Wise Error Rate for all of my results.	Instead, in the analysis of the heterogeneous results I use the Westfall-Young multiple-hypothesis testing method to adjust for that I am testing two hypotheses.	

C. Additional Empirical Analyses

1. Attrition and Non-Responses in the Survey

The attrition from the survey was 11%. Appendix Table D25 provides estimates for the effect of the scorecards and information treatments on the attrition rate. Column (1) of Panel A shows that the scorecards are estimated to have had a positive effect on the attrition rate by 3%. Column (2) of Panel A shows that the information intervention did not affect attrition.

An alternative definition of attrition is if a respondent did not answer a specific question in the survey. This definition classifies applicants who did not provide an answer to a specific question as having attrited, regardless of if these applicants were interviewed for the follow-up survey or not. In columns (3)-(4) of Table D25. I show that when defining attrition in this way for the question about the typical bribe payment, the scorecard treatment did not affect attrition, neither overall nor for over- or underperformers. Finally, columns (5)-(6) show the effect of the scorecards and information treatments on the non-response rate of respondents to the question about typical bribe payments, conditional on the respondent taking the follow-up survey.

Using Lee Bounds to Address Differential Attrition. — If the scorecards caused some applicants to not take the follow-up survey and these applicants, on average, had different values for an outcome variable, this would bias the estimates of the effect of the scorecards on those outcomes. To assess the potential bias stemming from the differential attrition on the estimated effect of the scorecards on bribe payments, I construct lower Lee bounds for the estimated effect (Lee, 2009). Lower Lee bounds are the relevant robustness test, since the effects on bribe payments are positive (overall and for overperforming offices) or non-negative (for underperforming offices). The results are shown in Appendix Table D26. The lower Lee bounds are not qualitatively different from the main results.

2. Effects in Survey Data: Visits, Time, and Satisfaction

In Columns (1)-(3) of Appendix Table D22, I use survey data to show that the scorecards reduced the average number of visits to the land office that each applicant made to process their application, increased the share of applications processed within 45 working days, and decreased overall processing times. Only the results for visits are statistically significant but the direction of the estimates and the heterogeneity between over- and underperforming offices are consistent with the results in the administrative data.

Column (4) of Appendix Table D22 shows the estimated effects of the scorecards on applicant satisfaction, which is negative but small and not statistically significant.[13] The negative effect is driven by offices that were overperforming at baseline, which is consistent with the observation that the scorecards increased bribe payments without improving processing times in these offices.

[13] Satisfaction was measured in the follow-up phone survey by asking applicants "Overall, how satisfied are you with the processing of your application?" The respondent could answer the question on a five-point scale which was then transformed into standard deviations from the control group mean.

3. *External Validity for Offices Outside Survey*

A potential threat to external validity is that the survey and information intervention may have affected the behavior of bureaucrats and applicants differentially in the offices receiving scorecards and the offices in the control group. To rule this out, Appendix Table D16 conducts the main analysis for processing times using only applications that could not have been affected by the survey. I restrict the sample to applications from offices that were never surveyed and applications that were made more than 45 working days before the start of the survey in offices that were eventually surveyed. The estimates in Appendix Table D16 are close to the estimates found in Table 2.

4. *Unintended Consequences of the Scorecards*

A common problem with quantitative performance measures is that they lead to unintended consequences. Below I discuss the potential unintended consequences the performance scorecards could have led to.

Gaming the Performance Indicators. — One potential concern is that land offices may reduce the number of applications, either by refusing to serve some applicants or by processing some applications using the paper-based system. With a smaller number of applications, it may be easier to reach a higher performance. Column (1) in Panel A of Appendix Table D27 shows that the scorecards did not substantially affect the number of applications received in the e-governance system. Column (1) in Panel B shows that the scorecards did not decrease the number of applications more in the offices that were underperforming at baseline and where the scorecards had the largest effect on processing times.

Another potential problem could be that bureaucrats allow applications only from individuals for whom they know it is easier to process the application within the time limit. The size of the land for which the land-record change is being made is positively associated with the processing time.[14] Therefore, if bureaucrats intended to avoid accepting complex applications, we would also expect to see a decrease in applications' average land size. Column (2) in Panel A of Appendix Table D27 shows that the scorecards did not substantially affect the average land size among applications received. Column (2) in Panel B shows a negative point estimate for the effect of the scorecards on application land size in underperforming offices, but that effect is imprecisely measured and does not provide conclusive evidence of the heterogeneous effects.

Quality of Decision Making. — Another potential concern is that the quality of the decisions made by the bureaucrats was reduced by the scorecards. The main decision the bureaucrat makes about the application is whether to accept or reject it. Column (1) in Panel A of Appendix Table D28 shows that the scorecards did not substantially change the percentage of applications rejected overall. Column (1) in Panel B shows that for overperforming offices, the point estimate for the

[14] In a simple regression of log processing time on log land size, controlling for application month and office fixed effects, the coefficient on land size is 0.012, statistically significant at the 1% level.

effect is positive, while for underperforming offices the point estimate is negative. However, both coefficients are small and not statistically significant.

Even if the rejection rate did not change, it is still possible that the quality of the decisions was worse. If this was the case, both more applications that should have been accepted were rejected, and more applications that should have been rejected were accepted. If an application is wrongfully rejected, applicants typically reapply in the same office. Therefore, the percentage of applicants reapplying after having been previously rejected can be used as an indicator for the percentage of incorrect rejections. Column (2) in Panel A of Appendix Table D28 shows that the percentage of applicants stating that they were reapplying, after previously having been rejected, increased with the scorecards, but the estimate is imprecise and not statistically significant. Column (2) in Panel B shows that the increase is driven by an increase in overperforming offices, while there was no increase in underperforming offices.

Together, these results suggest that the scorecards did not lead to a decrease in the quality of decision-making because bureaucrats were pressured to make faster decisions, since both rejections and incorrect rejections were unaffected by the scorecards in underperforming offices, the offices that improved their processing speed. Instead, the results are consistent with applicants in over-performing offices not satisfying the new higher bribe demands in these offices and are therefore rejected. However, these results should be interpreted with caution as the low rate of reapplying applicants causes the estimated effects to be imprecise relative to the control group mean.

Spillover Effects on Applications Not Measured by the Scorecards. — Even after the e-governance system had been installed, not all applications were made in the e-governance system, as described in Section I.A. In the survey data, 24% of applications were not made using the e-governance system and did therefore not count toward the performance indicators in the scorecards. If the bureaucrats had reacted to the scorecards by diverting attention away from any task that was not measured by the scorecards, we would expect a negative spillover from the scorecards on the processing times for these applications. In Appendix Table D23, I estimate the effects of the scorecards on the number of visits needed and processing times for applications made outside of the e-governance system. Due to the decreased sample size, the estimates are imprecise, but overall there is no evidence for negative spillovers. Instead, the point estimates are zero or improvements. Just as for the applications made in the e-governance system, the improvement is driven by offices underperforming at baseline, which is consistent with positive spillovers. While these results do not rule out all potential negative spillovers from the scorecards on other tasks performed by the bureaucrats, it suggests that the bureaucrats did not "game the system" by focusing only on those applications that would improve their scorecards.

5. Differential Effects by Symbols Shown on the First Scorecard

To test how important the thumbs-up and thumbs-down symbols are for the effect of the scorecards, I focus on the thumb symbols shown in the first scorecard, which could not have been affected by the treatment assignment. Appendix Table D6 shows the results from three regressions with heterogeneous results by the thumb symbols. As I only expect the thumb symbols in the first

scorecard to have an effect for the first two to three months, I restrict the sample to applications made within 60 working days of the start of the experiment. Column (1) of Appendix Table D6 shows the overall treatment effect as well as the heterogeneity of the treatment effect by the ranking for processed applications and pending applications, as shown on the scorecard. In Column (2) I show the heterogeneity in the effect by the number of positive or negative thumb symbols shown on the first scorecard. As expected from the overall results in the paper, the offices receiving one or two thumbs-down symbols improve their processing times more as a result of the scorecards, while there are no statistically significant results for offices receiving thumbs-up symbols, no thumbs, or a mix of thumbs-up and thumbs-down symbols. Finally, in Column (3), I try to isolate the effects of the thumb symbols from the effects of different rankings by controlling for each of the rankings as well as the rankings interacted with the treatment. This relies on the assumption that the effect of the scorecards changes approximately linearly with the rankings and thus can be accounted for by linear controls. Using this specification, the effect is still largest for the scorecards receiving one or two negative thumb symbols but the differential effects are only statistically significant at the 10% level. Overall, I interpret the results in Table D6 as suggestive evidence that the thumb symbols may have been important for the bureaucrats' reactions to the scorecards. This could be because the thumbs symbols helped make the scorecards salient and easy to understand.

6. *Effect of Peer Performance List*

Information flows about performance between bureaucrats at the same level in the organizational hierarchy may create an additional incentive for improved performance. I measure such a peer effect, as an addition to the effect of the scorecard, by estimating the effect of the peer performance list sent to a subgroup of the offices receiving scorecards, as described in Section I.C.

The bureaucrats receiving the peer performance list all hold the same position within the bureaucracy and they are "competitors" for the same posting and promotion opportunities. Many of them are "batchmates", i.e. bureaucrats who joined the civil service in the same year and underwent training together." From my qualitative interviews, it is clear bureaucrats typically compare themselves to their batchmates. As there is only one bureaucrat in each land office, there is no competition within offices. Furthermore, as the civil servants are all part of the same national civil service, which is not geographically segmented, it is natural that they compare themselves to other bureaucrats holding the same position, regardless of where in the country these bureaucrats are located.

Appendix Table D17 shows that there is no substantial effect of the peer performance list on processing times. Using only data from land offices receiving performance scorecards, columns (1) and (3) show that the effect on processing times is close to zero. Columns (2) and (4) use the full data set and estimate the effect of being in the scorecard and the peer performance list treatment groups simultaneously, both before and after the performance list intervention started. The point estimate is a minor improvement in performance but it is not statistically significant.[15] There are

[15]The effects from the standard scorecard treatment as measured by the coefficients on *Scorecard* and *Post* × *Scorecard* can account for virtually all of the effect estimated by my main specification in Table 2, showing that the effect of the scorecards is not driven by the inclusion of the Peer Performance Lists.

also no substantial heterogeneous effects by performance, neither when comparing offices to all offices in the Peer Performance List experiment, nor when comparing bureaucrats within their group of batchmates.

7. *Alternative Explanations for the Effects of Scorecards*

Increased Workload on Bureaucrats. — One potential explanation for the scorecards causing higher bribe payments is that the scorecards increase the bureaucrats' workload and the opportunity cost of their time. If bribe payments are made for bureaucrats' time, this could increase bribe payments. Another alternative explanation is that faster processing times cause applicants to be willing to pay more. However, both of these explanations are inconsistent with the results that in the underperforming offices where the changes in processing times are the largest, bribe payments do not change. Instead, bribe payments increase in the overperforming offices where changes in processing times are small. Therefore, it is unlikely that either of these mechanisms is a substantial driver of the effect on bribe payments.

Bureaucrat Transfers . — An alternative explanation that is consistent with the heterogeneity in the effects on processing times and bribes is that overperforming bureaucrats get transferred due to receiving positive scorecards and that they are replaced by average-performing bureaucrats. If the average-performing bureaucrats have both slower processing times and demand higher bribes, we expect that bribe payments would increase in offices overperforming at baseline. Processing times may not change, because the incentive effect of the scorecards may cancel out the effect of the bureaucrat transfers. However, this explanation is refuted by the data.

Appendix Table D19 shows that the scorecards did not affect bureaucrats' transfers. Column (1) shows the effect on the monthly probability of being transferred. Column (2) shows the effect on the duration of the posting for the bureaucrat posted in the office at the start of the experiment. Column (3) shows the effect on not having any ACL assigned to the office. All of the effects are close to zero, both overall and when considering the effects for over- and underperforming offices separately. This shows that the scorecards did not affect the timing or frequency of the bureaucrats' transfers.

Monitoring Efforts Shifted Away from Overperformers. — If the scorecards led supervisors to decrease the monitoring of bureaucrats in the overperforming offices, that could have increased the bribes taken in these offices. However, if this was the main mechanism, we would have expected underperforming offices to have experienced an increase in monitoring and thus a decrease in bribes taken in those offices. Therefore, the results are only consistent with this mechanism if there was an asymmetric response by supervisors shifting monitoring away from overperforming offices but not towards underperforming offices, or if a decrease in monitoring leads to a larger increase in bribe payments than an increase in monitoring leads to a decrease in bribe payments. While I cannot completely rule this out, I find it less plausible than my preferred interpretation.

Supervisors Demanding Higher Bribes. — I cannot observe sharing of bribe money within the bureaucracy, but it is plausible that this happens (Bandiera et al., 2021; de la Sierra et al., 2022). If the scorecards led to the supervisors of overperforming bureaucrats demanding that larger amounts were to be shared with them, this could drive up bribe payments. The most plausible reason for such a demand would be a mechanism similar to the one in the model. In other words, the model would be correctly describing the mechanism, but with the wrong person described as the "bureaucrat." This seems unlikely, since for the supervisors (UNOs), processing land-record applications is a very small share of their responsibilities, while for the bureaucrats (ACLs), it represents 25%-50% of their work. It is therefore more likely that the scorecards cause a change in the bureaucrat's reputation level and incentives, than those of the supervisors.

Bureaucrats Using Scorecards in Negotiations With Applicants. — Another alternative explanation is that positive scorecards help bureaucrats prove to applicants that they can process applications quickly. This could then allow the bureaucrats receiving positive scorecards to charge higher bribes, while it would not affect bribes in offices receiving negative scorecards.

This explanation is implausible for four reasons. First, the coefficients on "Overperform baseline" in columns (4) and (5) of Appendix Table D20, show that the expected processing times are 13% lower in the overperforming offices not receiving the scorecards, suggesting that the applicants are already aware of the faster processing times in these offices. Second, column (4) of Appendix Table D20 shows that the point estimate for the effect of the scorecards on applicants' expectations of processing times in overperforming offices is a 5% increase. Although the positive effect is not statistically significant, if the scorecards helped bureaucrats improve applicants' expectations of processing times, the effect on the expectations should be negative. Third, although I cannot rigorously rule out that no one in the land offices showed the scorecards to applicants, in none of the qualitative interviews done with bureaucrats and applicants was it ever mentioned that the scorecards were shown to applicants. Fourth, the information intervention tried to accomplish what a bureaucrat could have achieved by showing the scorecards to applicants, but Table D18 shows that the information intervention did not substantially increase bribes.

8. Noncompliance With the Information Treatment Assignment

Enumerators implemented the information intervention during the in-person interviews. They were given a different schedule for each land office where each weekday was indicated as a day when the treatment would be delivered or not (Mattsson, 2019). During the interview, the enumerator indicated in the survey if the information had been given or not. According to the enumerators' indications of whether the information was given, in 17% of interviews incorrect information was given. This was due to enumerators misunderstanding the schedule. Among the interviews where the treatment assignment was not followed, 91% were on days when the median interview was not given the information as per the treatment assignment.

The main result related to the information intervention is that it did not decrease bribe payments. In particular, it did not decrease bribe payments in offices receiving the scorecards. Since this is a null result, I make conservative choices regarding the empirical strategy such that the information

intervention has the highest ex-ante probability to cause a decrease in bribe payments. Therefore, in the paper's main analysis, I use the median treatment delivered in a particular land office and survey day as the treatment variable.

Appendix Table D29 shows the robustness of the estimated effects of the information treatment to using alternative definitions of the treatment variable. Column (1) uses my preferred treatment variable based on the median treatment delivered on the office survey day. Column (2) uses the individual treatment delivered to the applicant, as per the enumerator's indication in the survey. Column (3) uses the treatment assigned and column (4) uses the treatment assigned but drops observations where the treatment was not correctly implemented. The null result of the effect on bribe payments is robust across all treatment variable definitions, both overall (Panel A) and in the offices that received the scorecards (Panel B). In Panel C, I show the robustness of the estimated effect of the information intervention on the expected processing time. The estimated effect is the largest when using the median delivered treatment or the actual delivered treatment, confirming that this is the most relevant variable for the analysis of the effect on bribe payments. When using the assigned treatment the effect is small, but when limiting the sample to the applicants for whom the correct treatment was delivered, the effect is similar to the effect estimated using the delivered treatment. This is what we would expect to see if the information intervention had a negative effect on the expected processing time. When using only the treatment assignment, the estimated effect is small, since many applicants in the control group received the treatment and reduced the difference between the treatment and control groups.

9. Estimating the Value of Processing Days Saved

To estimate the total value of the improvements in processing times per year among the treatment offices, I multiply the estimates for the total number of processing days saved per year with the average implied valuation of having the application processed one day faster.

To calculate the total number of processing days per year in the treatment group had there been no scorecards, I take the average processing time in the control offices and multiply it by an estimate for the annual number of applications in the treatment offices.[16] I then multiply this number by the percentage decrease in processing time caused by the scorecards, estimated by a uniformly weighted regression according to the specification in Equation 1.

I calculate the value of having the application processed one day faster by taking the average of the following expression: $\frac{\text{Value of processing in 7 days}}{\text{Expected processing time from survey date}-7}$, across all surveyed applicants. The information on the stated valuations of processing within 7 days and the expected processing time comes from the in-person survey. Applicants with an expected processing time of 7 days or less are excluded from the calculation.

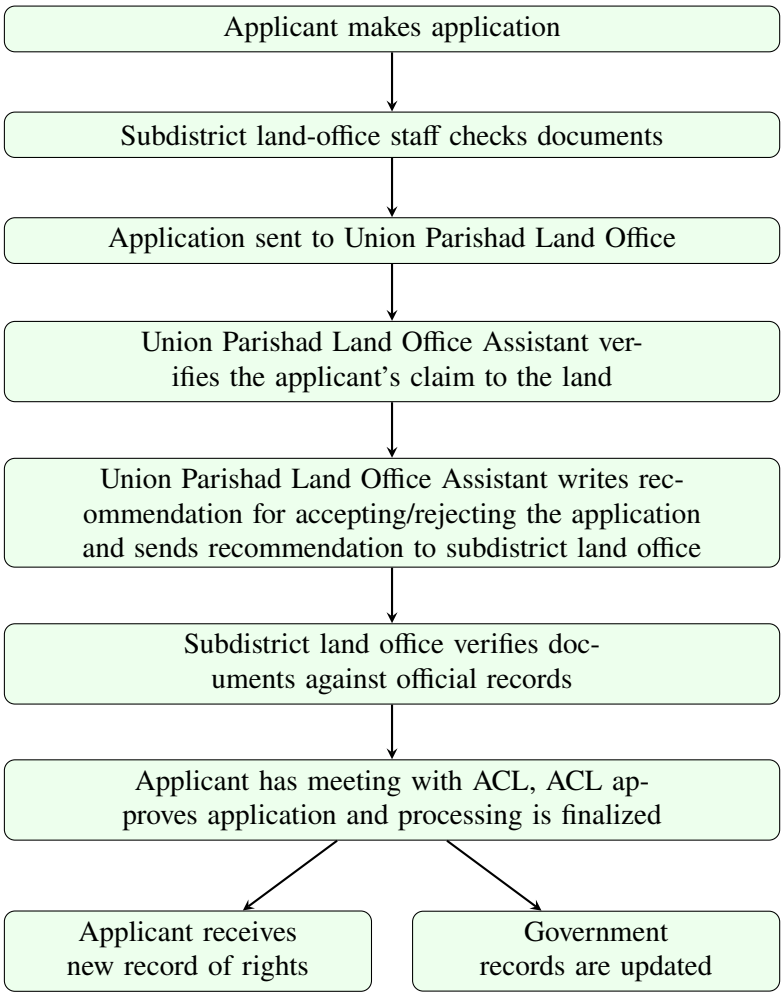
Finally, I multiply the estimated number of processing days saved per year by the average valuation of having the application processed one day faster. The estimate for the total valuation of all processing days saved by the scorecards in the treatment offices is USD 9.7 million per year. The overall cost of implementing the scorecards is approximately USD 40,000. This approximation

[16]The estimate is two times the number of applications in the treatment offices in the last six months of 2019. During this period the usage of the e-governance system had stabilized, but it was before the reduction in applications due to the COVID-19 pandemic.

includes the setup costs and the time the author contributed toward implementing the scorecards, valued at USD 200 per day.

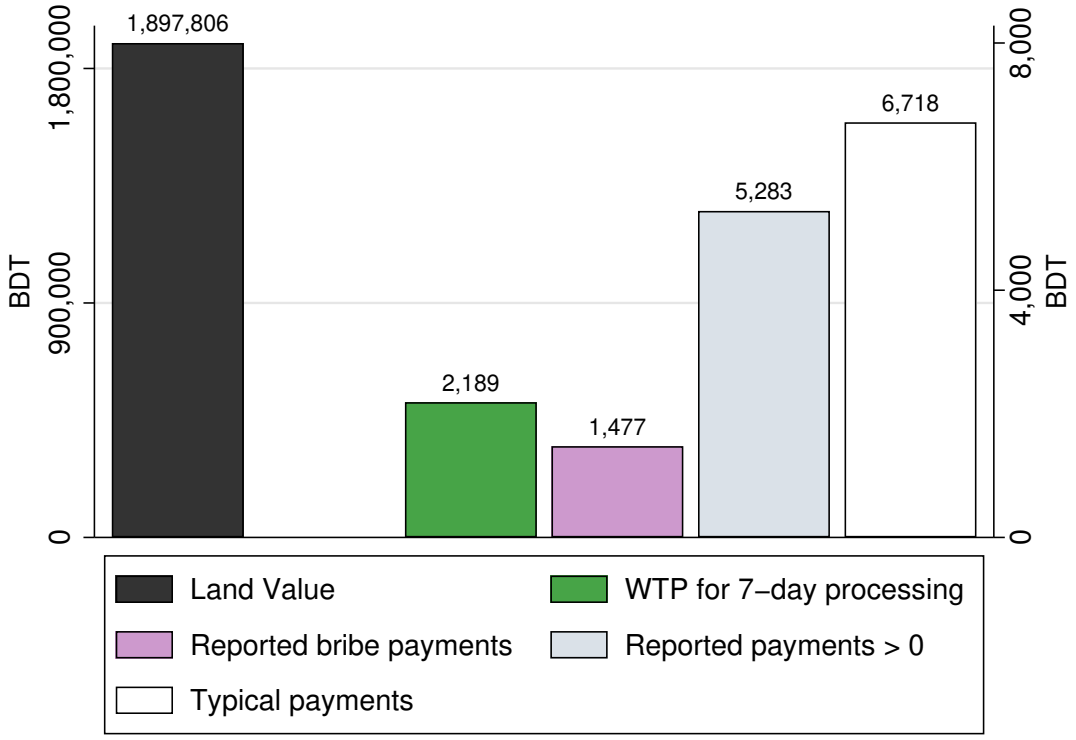
D. Additional Tables and Figures

Figure D1. : Application Process for Successful Application



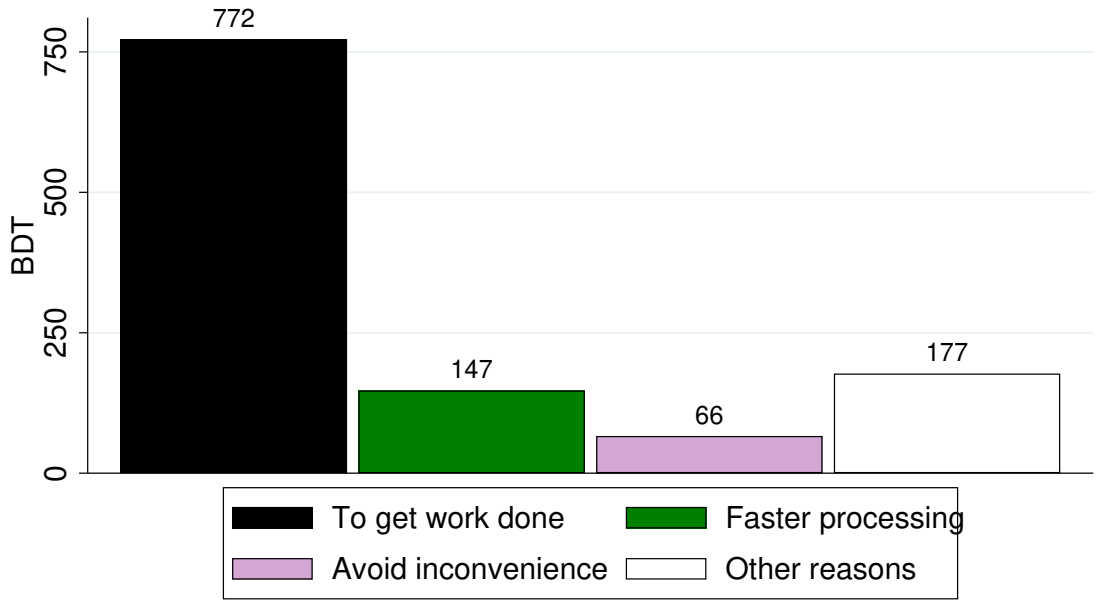
Note: This figure depicts the *de jure* process for a successful land-record change application. See discussion in Section I.A.

Figure D2. : Value of Land, Faster Processing, and Bribes



Note: The figure displays the applicants' stated valuations of the parcels of land for which they are making the record change applications for and compares this to the bribe payments. The first bar shows the average of applicants' stated value their land. The second bar shows the average stated value of getting the application processed within seven days from the time of the first survey. The third bar shows the average value of bribe payments reported by the applicant. The fourth bar shows the average value of reported bribe payments among applicants reporting a nonzero bribe. The fifth bar shows the average response to the question about a typical bribe payment by "a person like yourself". The first bar is measured on the left axis, the next four bars are measured on the right axis. All variables are winsorized at the 99th percentile. Observations are weighted by the inverse of the number of observations in that land office. USD/BDT≈84.3. See discussion in Sections I.A and V.

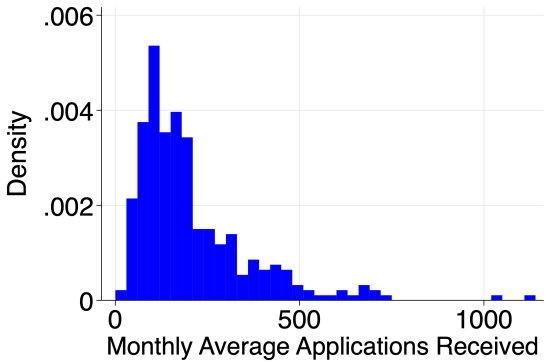
Figure D3. : Stated Reasons for Bribe Payments



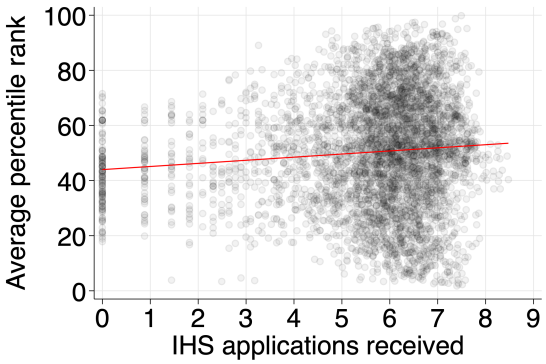
Note: The figure displays average reported bribe amounts for different reasons stated by the applicants'. The amounts include payments to government officials and agents. The question about the reason for paying was open-ended and coded into response categories. All variables are winsorized at the 99th percentile. Observations for which no reason was provided are excluded. Observations are weighted by the inverse of the number of observations in that land office. USD/BDT≈84.3. See discussion in Section I.A.

Figure D4. : Land Office Size and Performance

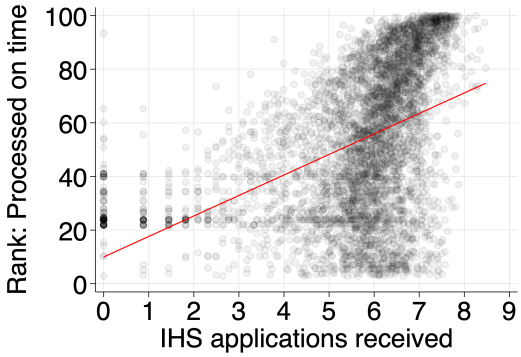
(a) Offices' Monthly Average of Applications Received



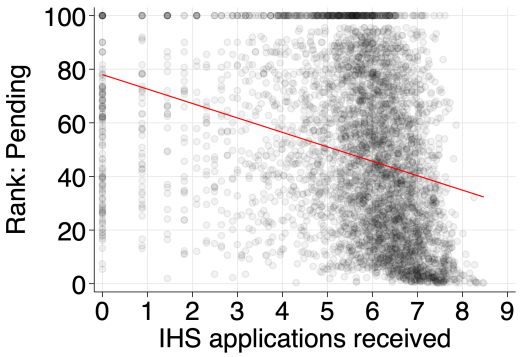
(b) Average Performance Ranking and Applications Received



(c) Ranking of Applications Processed Within 45 Working Days and Applications Received

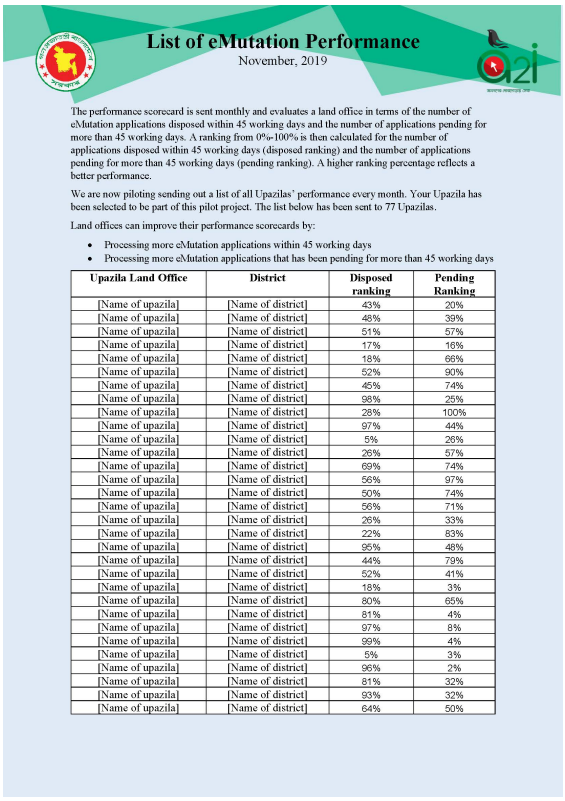


(d) Ranking of Applications Pending Beyond 45 Working Days and Applications Received



Note: The figure displays information on the number of applications received by a land office in a month and that variable's relationship with the performance ranking in that month. Data is from when the office first started using the e-governance system until the end of the experiment. Sub-Figure a) shows a histogram of the offices' monthly average number of applications received. Sub-Figure b) shows the relationship between the average performance ranking in a month and the inverse hyperbolic sine (IHS) transformation of the number of applications received in that month. Sub-Figure c) shows the percentile ranking of the number of applications processed on time and the IHS of the number of applications received. Sub-Figure d) shows the percentile ranking of the number of applications pending beyond the 45-working-day time limit and the number of applications received. See discussion in Section I.B and Appendix B.4.

Figure D5. : Example Peer Performance List and Information Treatment Leaflet



List of eMutation Performance

November, 2019

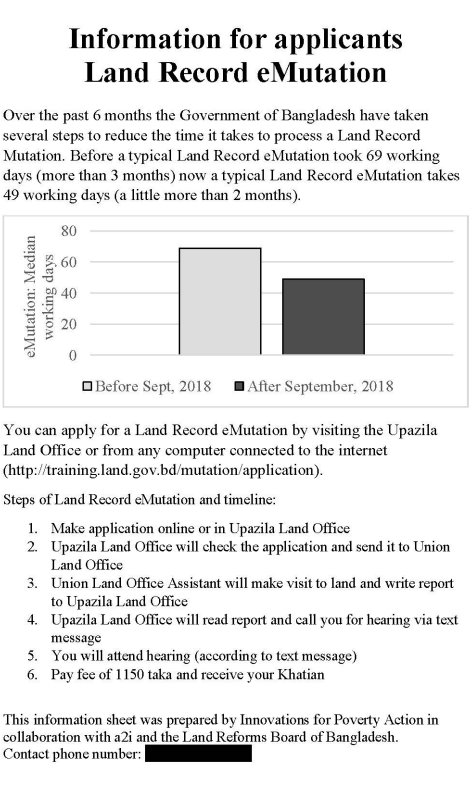
The performance scorecard is sent monthly and evaluates a land office in terms of the number of eMutation applications disposed within 45 working days and the number of applications pending for more than 45 working days. A ranking from 0%-100% is then calculated for the number of applications disposed within 45 working days (disposed ranking) and the number of applications pending for more than 45 working days (pending ranking). A higher ranking percentage reflects a better performance.

We are now piloting sending out a list of all Upazilas' performance every month. Your Upazila has been selected to be part of this pilot project. The list below has been sent to 77 Upazilas.

Land offices can improve their performance scorecards by:

- Processing more eMutation applications within 45 working days
- Processing more eMutation applications that has been pending for more than 45 working days

Upazila Land Office	District	Disposed ranking	Pending Ranking
[Name of upazila]	[Name of district]	43%	20%
[Name of upazila]	[Name of district]	48%	39%
[Name of upazila]	[Name of district]	51%	57%
[Name of upazila]	[Name of district]	17%	16%
[Name of upazila]	[Name of district]	18%	66%
[Name of upazila]	[Name of district]	52%	90%
[Name of upazila]	[Name of district]	45%	74%
[Name of upazila]	[Name of district]	98%	25%
[Name of upazila]	[Name of district]	28%	100%
[Name of upazila]	[Name of district]	97%	44%
[Name of upazila]	[Name of district]	5%	26%
[Name of upazila]	[Name of district]	26%	57%
[Name of upazila]	[Name of district]	69%	74%
[Name of upazila]	[Name of district]	56%	97%
[Name of upazila]	[Name of district]	50%	74%
[Name of upazila]	[Name of district]	56%	71%
[Name of upazila]	[Name of district]	26%	33%
[Name of upazila]	[Name of district]	22%	83%
[Name of upazila]	[Name of district]	95%	48%
[Name of upazila]	[Name of district]	44%	79%
[Name of upazila]	[Name of district]	52%	41%
[Name of upazila]	[Name of district]	18%	3%
[Name of upazila]	[Name of district]	80%	65%
[Name of upazila]	[Name of district]	81%	4%
[Name of upazila]	[Name of district]	97%	8%
[Name of upazila]	[Name of district]	99%	4%
[Name of upazila]	[Name of district]	5%	3%
[Name of upazila]	[Name of district]	96%	2%
[Name of upazila]	[Name of district]	81%	32%
[Name of upazila]	[Name of district]	93%	32%
[Name of upazila]	[Name of district]	64%	50%



Information for applicants
Land Record eMutation

Over the past 6 months the Government of Bangladesh have taken several steps to reduce the time it takes to process a Land Record Mutation. Before a typical Land Record eMutation took 69 working days (more than 3 months) now a typical Land Record eMutation takes 49 working days (a little more than 2 months).

You can apply for a Land Record eMutation by visiting the Upazila Land Office or from any computer connected to the internet (http://training.land.gov.bd/mutation/application).

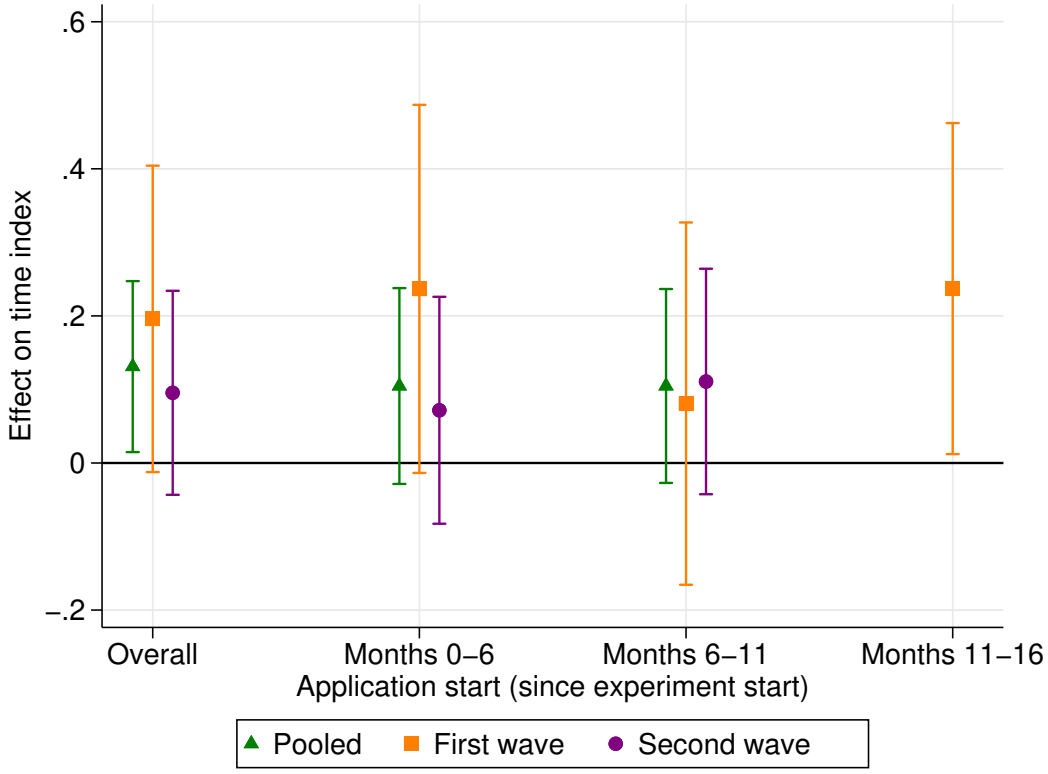
Steps of Land Record eMutation and timeline:

1. Make application online or in Upazila Land Office
2. Upazila Land Office will check the application and send it to Union Land Office
3. Union Land Office Assistant will make visit to land and write report to Upazila Land Office
4. Upazila Land Office will read report and call you for hearing via text message
5. You will attend hearing (according to text message)
6. Pay fee of 1150 taka and receive your Khatian

This information sheet was prepared by Innovations for Poverty Action in collaboration with a2i and the Land Reforms Board of Bangladesh.
Contact phone number: [redacted]

Note: On the left is an example of the first page of a peer performance list. The office and district names have been removed from the peer performance list to preserve anonymity. See discussion in Section I.C. On the right is an English translation of the leaflet used for the information treatment. See discussion in Section I.E.

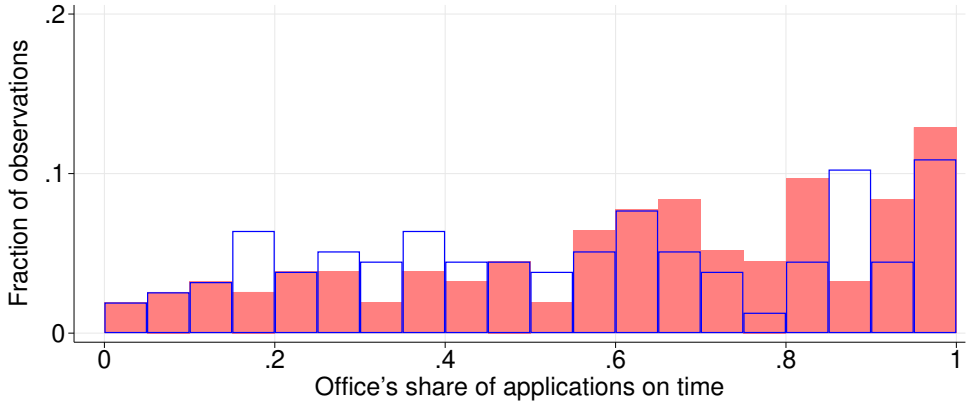
Figure D6. : Scorecards' Effect by Time Since Start of the Experiment



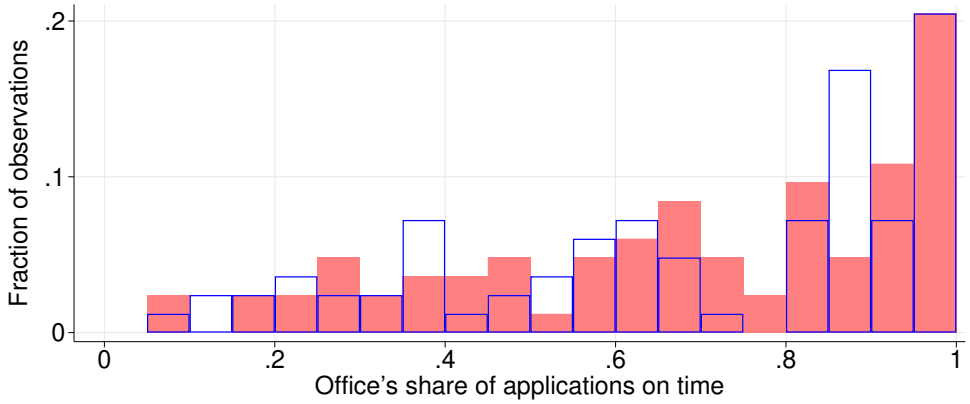
Note: The figure reports the regression coefficients and confidence intervals for regressions using applications made during different periods relative to the start of the experiment. The effects are measured in standard deviations of the time index from column (3) of Table 2. The index is constructed using the variables for whether the application was processed on time and the log of the overall processing time. The estimate for each period comes from a separate regression restricting the sample to applications made in that period. The results are from regressions using data from all offices (triangles), offices in the first randomization wave (squares), and offices in the second randomization wave (circles). The months are numbered relative to when the first scorecards were sent out for that office's randomization wave. Month 0 is the month before the first scorecard. Confidence intervals are constructed using standard errors clustered at the office level. See discussion in Section III.A.

Figure D7. : Distributions of Offices' Share of Applications Processed on Time

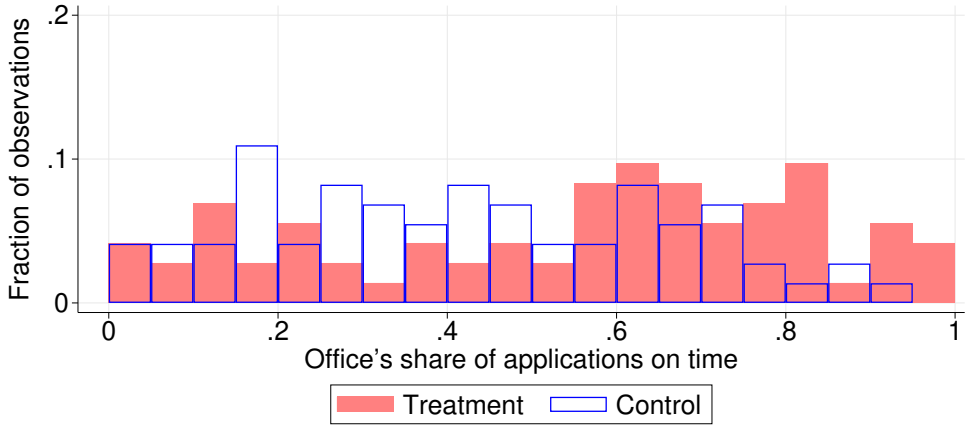
(a) Distributions of Offices' Share of Applications Processed on Time



(b) Distributions for Overperformers at Baseline



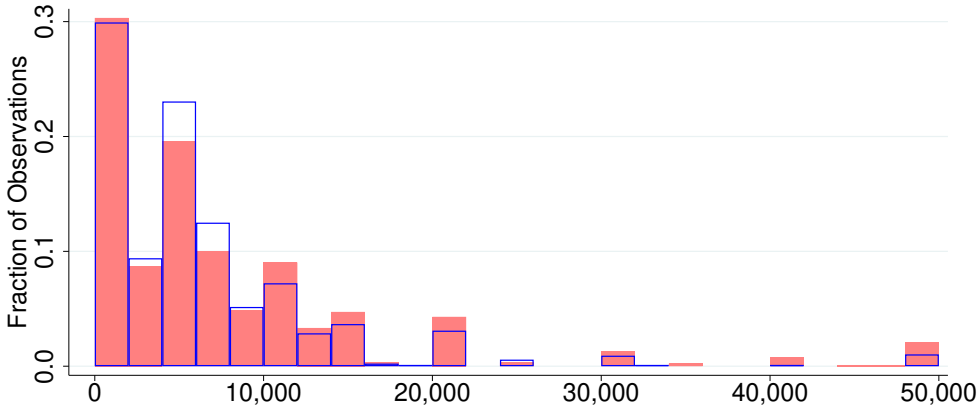
(c) Distributions for Underperformers at Baseline



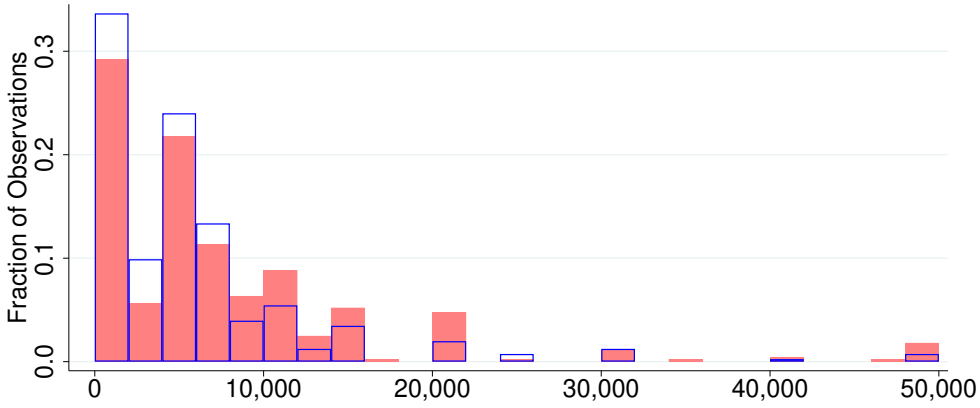
Note: The figures display histograms of each office share of applications processed within the 45 working day time limit for the treatment and control groups separately. Sub-figure a) uses data from all offices. Sub-figure b) and c) use data from offices over- and underperforming at baseline, respectively. See discussion in Section III.A.

Figure D8. : Histogram of Typical Bribes by Treatment

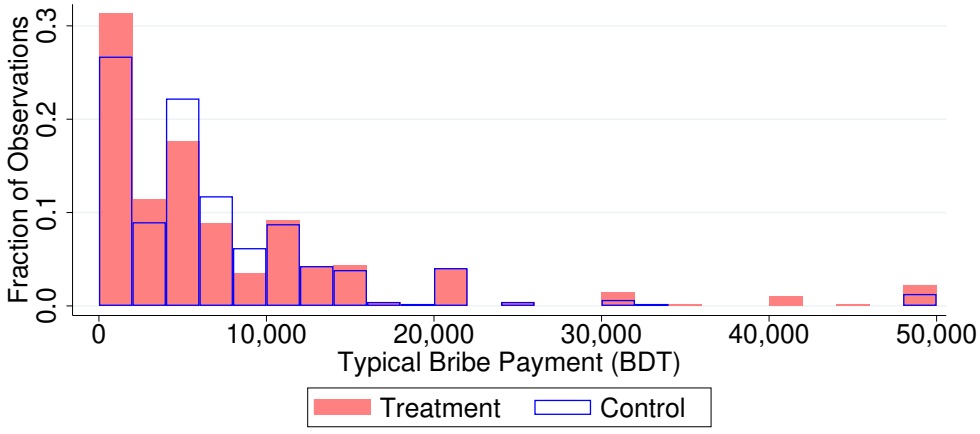
(a) Differences in Distributions of Typical Bribes



(b) Differences in Distributions for Overperformers

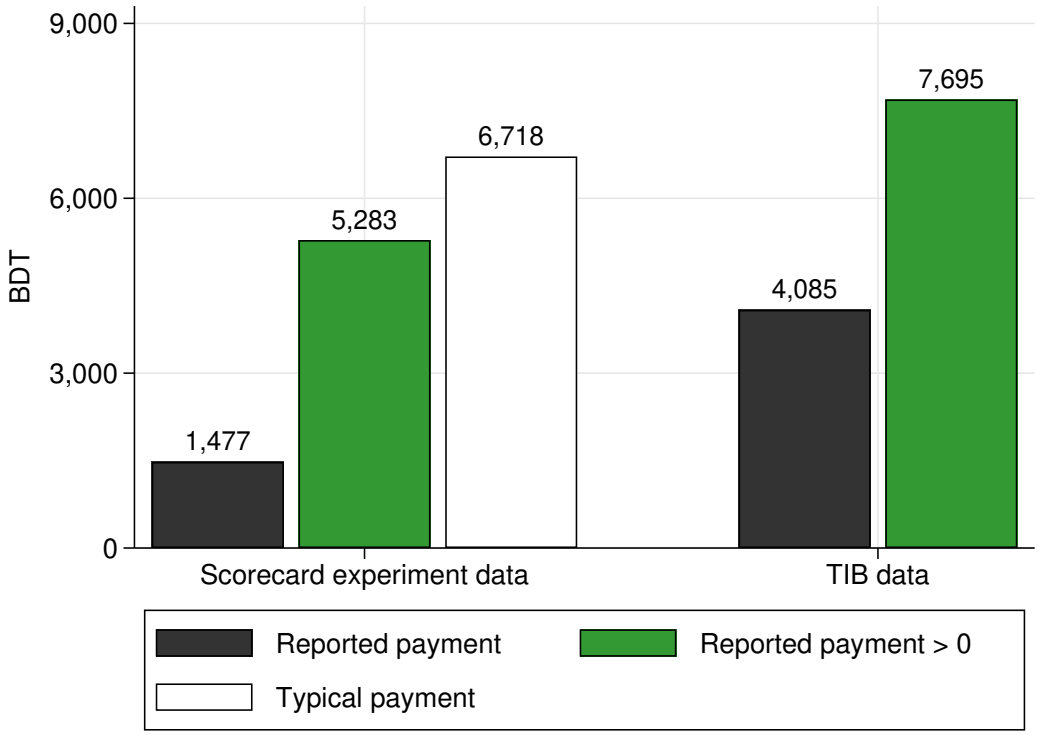


(c) Differences in Distributions for Underperformers



Note: The figures display histograms of typical bribe payments for the treatment and control groups separately. Sub-figure a) uses data from all surveyed offices. Sub-figure b) and c) use data from offices over- and underperforming at baseline, respectively. See discussion in Section III.B.

Figure D9. : Comparison of Estimated Bribes to the TIB Survey



Note: The figure displays the average bribe payments reported in the phone survey conducted to evaluate the scorecard experiment and in an independent survey by Transparency International Bangladesh (TIB). The first bar shows the average value of bribe payments reported by applicants in the scorecard experiment phone survey; 73% of respondents reported having paid no bribes. The second bar shows the average value of bribe payments reported by applicants who reported having paid some bribe in the scorecard experiment phone survey. The third bar shows the average response to the question about the value of a typical bribe payment by "a person like yourself" in the scorecard experiment phone survey; 27% of respondents reported that a typical applicant paid no bribes. The fourth bar shows the average value of bribe payments reported by applicants in the TIB survey; 57% of the respondents reported having paid no bribe for their land record change. The fifth bar shows the average value of bribe payments reported by respondents who reported having paid some bribe in the TIB survey. All variables are winsorized at the 99th percentile. Observations in the three first bars are inversely weighted by the number of observations in that land office. USD/BDT≈84.3. See discussion in Appendix B.5.

Table D1—: Balance of Randomization

	Scorecard		Control		Difference
Panel A. Administrative Data	Obs. (Cluster)	Mean (SD)	Obs. (Cluster)	Mean (SD)	Diff. (SE)
≤45 working days	56,703	0.50	56,564	0.53	-0.028
	(146)	(0.50)	(146)	(0.50)	(0.04)
ln(Process time)	56,703	3.77	56,564	3.69	0.083
	(146)	(1.40)	(146)	(1.44)	(0.13)
Time index	56,703	-0.06	56,564	-0.00	-0.061
	(146)	(0.98)	(146)	(1.00)	(0.09)
Process time (w. days)	56,703	77.60	56,564	76.90	0.700
	(146)	(73.53)	(146)	(79.15)	(6.98)
Approved	36,410	0.70	38,776	0.73	-0.030
	(136)	(0.46)	(141)	(0.44)	(0.03)
Panel B. Survey Data					
Applicant age	1,383	47.34	1,377	47.39	-0.05
	(56)	(13.35)	(56)	(13.82)	(0.64)
Female	1,460	0.06	1,409	0.06	-0.00
	(56)	(0.23)	(56)	(0.24)	(0.01)
Monthly income (BDT 1K)	1,460	24.49	1,409	22.85	1.64
	(56)	(17.62)	(56)	(17.88)	(1.22)
App. status: Applying	1,382	0.23	1,377	0.22	0.01
	(56)	(0.42)	(56)	(0.42)	(0.03)
Ongoing	1,382	0.64	1,377	0.63	0.01
	(56)	(0.49)	(56)	(0.49)	(0.04)
Rejected	1,382	0.00	1,377	0.00	0.00
	(56)	(0.07)	(56)	(0.06)	(0.00)
Approved	1,382	0.13	1,377	0.15	-0.02
	(56)	(0.34)	(56)	(0.36)	(0.03)
Land value (BDT 100K)	1,335	20.70	1,336	17.32	3.38
	(56)	(31.87)	(56)	(29.82)	(2.17)
Land size (acre)	1,379	0.24	1,369	0.24	-0.00
	(56)	(0.41)	(56)	(0.39)	(0.02)

Note: The table reports the balance of randomization. Panel A. uses administrative data from 45 working days before the first scorecard. Due to this restriction, only 292 of the 311 offices are included. For applications not processed by the date of the first scorecard, the processing time is imputed using the procedure in Section I.D. Data on approvals are as per the date of the first scorecard. P-value for the F-test of joint orthogonality: 0.86. Panel B uses survey data from the in-person survey of applicants,which was conducted before the conclusion of the processing of the applicants' applications but after the start of the scorecards. P-value for F-test of joint orthogonality: 0.90. Continuous variables are winsorized at the 99th percentile. In both panels, observations are inversely weighted by the number of observations in that land office. See discussion in Section I.F.

Table D2—: Scorecards' Effect on Offices' Average Outcomes

Panel A: Overall effect	(1) Share ≤45 w. days	(2) Mean ln(w. days)	(3) Mean Typical Payment	(4) Mean Reported Payment
Scorecard	0.060	-0.125	1208.756	381.348
	(0.029)	(0.062)	(640.215)	(197.325)
Panel B: Heterogeneous effects				
Scorecard × Overperform	0.010	-0.042	1995.885	717.403
	(0.039)	(0.084)	(839.566)	(263.567)
Scorecard × Underperform	0.114	-0.217	539.818	124.412
	(0.043)	(0.094)	(973.240)	(278.847)
Overperform baseline	0.192	-0.310	-1465.849	-831.000
	(0.053)	(0.113)	(1011.956)	(329.406)
Stratum FE	Yes	Yes	Yes	Yes
Observations	311	311	112	112
Control mean	0.56	3.70	5,860.93	1,283.79

Note: This table shows the main results when collapsing the data to the land office level and using averages for each office as the outcome variable. Column (1) shows the effect on the offices' share of applications processed within the 45 working day time limit. Column (2) shows the effect on the offices' average log processing time. Column (3) shows the effect on the average response to the question of how much it is "normal for a person like yourself to pay" in bribes. Column (4) shows the effect on reported bribe payments to government officials or agents. The point estimates differ slightly from the main results as month fixed effects cannot be used in this analysis. Heteroskedasticity robust standard errors. See discussion in Section III.D.

Table D3—: Alternative Measures of Baseline Performance

	Time index	Bribe amount		Promotion
Panel A: 3 month baseline	(1)	(2)	(3)	(4)
Scorecard × Overperform 3m baseline	0.017	1880.859	573.722	0.036
	(0.080)	(685.325)	(224.394)	(0.075)
Scorecard × Underperform 3m baseline	0.248	-82.933	64.876	-0.153
	(0.087)	(962.752)	(260.075)	(0.079)
P-value: subgroup diff.	0.05	0.09	0.13	0.03
Panel B: Quartiles of baseline rankings				
Scorecard × 75-100th percentile	-0.027	1695.971	604.207	0.044
	(0.100)	(962.565)	(264.821)	(0.114)
Scorecard × 50-75th percentile	0.044	2434.674	668.988	0.144
	(0.117)	(1163.192)	(382.956)	(0.103)
Scorecard × 25-50th percentile	0.135	116.043	241.714	-0.120
	(0.119)	(1047.747)	(311.788)	(0.110)
Scorecard × 0-25th percentile	0.337	-142.577	-116.390	-0.258
	(0.121)	(1496.050)	(390.573)	(0.106)
Panel C: Continuous baseline ranking				
Scorecard × Baseline ranking	-0.007	38.187	15.666	0.007
	(0.003)	(29.153)	(8.571)	(0.003)
Scorecard	0.115	958.064	329.554	-0.043
	(0.057)	(608.920)	(173.485)	(0.053)
Start-month FE	Yes	Yes	Yes	
Stratum FE	Yes	Yes	Yes	Yes
Baseline performance control	Yes	Yes	Yes	Yes
Observations	1,034,688	1,802	2,869	302
Clusters	311	112	112	
Bribe measure		Typical	Reported	

Note: This table shows the robustness of the results for the heterogeneity of the effect of the scorecards for different baseline performance measures. Column (1) shows the effects on the index of the two main processing-time outcome variables used in column (3) of Table 2; a higher index value indicates a faster processing time. Column (2) shows the effect on the estimate for the typical bribe payment for "a person like yourself to pay". Column (3) shows the effects on the bribe payments reported by the applicant. Panel A shows the heterogeneity in the effect of the scorecards based on the land office having an above- or below-median average ranking across the last three months of the baseline period. Panel B shows the heterogeneity based on the quartile of baseline ranking. Panel C shows the heterogeneity based on the continuous baseline ranking. USD/BDT≈84.3. Standard errors are clustered at the office level, except for in column (4) which uses heteroskedasticity robust standard errors. Observations are inversely weighted by the number of observations in that land office. See discussion in Section III.D.

Table D4—: Differences Between Over- and Underperforming Offices

	Overperformers		Underperformers		Difference
Panel A. Application Admin. Data	Obs. (Cluster)	Mean (SD)	Obs. (Cluster)	Mean (SD)	Diff. (SE)
≤45 w. days	538,686	0.69	496,002	0.48	0.207
	(166)	(0.46)	(145)	(0.50)	(0.03)
Process time (w. days)	538,686	48.87	496,002	74.08	-25.209
	(166)	(58.01)	(145)	(77.81)	(4.65)
App. approved	528,451	0.71	477,814	0.64	0.066
	(166)	(0.46)	(145)	(0.48)	(0.02)
Panel B. Office by Month Admin. Data					
Total applications	2,384	279.30	2,132	286.00	-6.708
	(166)	(265.73)	(145)	(270.37)	(26.32)
Apps. processed ≤45 working days	2,384	173.57	2,132	124.73	48.838
	(166)	(200.83)	(145)	(184.45)	(16.07)
Apps. pending ¿45 working days	2,384	210.81	2,132	523.15	-312.341
	(166)	(472.37)	(145)	(837.45)	(66.87)
No ACL assigned	2,384	0.13	2,132	0.12	0.003
	(166)	(0.33)	(145)	(0.33)	(0.02)
Female ACL	2,080	0.38	1,867	0.29	0.092
	(163)	(0.49)	(143)	(0.45)	(0.05)
Panel C. Sub-District Characteristics					
Population	166	317914.20	145	297982.91	19931.288
	(166)	(199990.32)	(145)	(157828.08)	(20,316.24)
Area (Sq. Km)	166	260.47	145	253.28	7.184
	(166)	(166.92)	(145)	(155.97)	(18.32)
Share of Pop. in Agriculture	166	0.53	145	0.50	0.026
	(166)	(0.23)	(145)	(0.22)	(0.03)
Panel D. Survey Data					
Applicant age	1,280	47.76	1,480	46.99	0.77
	(55)	(13.96)	(57)	(13.21)	(0.63)
Female	1,338	0.06	1,531	0.06	0.00
	(55)	(0.24)	(57)	(0.23)	(0.01)
Monthly income (BDT 1K)	1,338	23.08	1,531	24.24	-1.16
	(55)	(17.21)	(57)	(18.27)	(1.23)
Land value (BDT 100K)	1,243	18.10	1,428	19.84	-1.74
	(55)	(26.94)	(57)	(34.27)	(2.17)
Land size (acre)	1,275	0.25	1,473	0.23	0.02
	(55)	(0.41)	(57)	(0.40)	(0.02)
Typical Bribe Payment	845	6,301.38	957	7,121.51	-820.12
	(55)	(7,986.50)	(57)	(9,017.23)	(696.96)
Reported Bribe Payment	1,338	1,286.04	1,531	1,660.61	-374.57
	(55)	(3,293.74)	(57)	(3,698.31)	(193.91)

Note: The table shows differences in summary statistics between offices that were over- and underperforming at baseline. The data is from the start until the end of the experimental period. Panel A. uses administrative data at the application level. Panel B. uses administrative data aggregated to the office by month level. Panel C. uses in sub-district characteristics from BBS (2014). Panel D uses survey data from the in-person survey of applicants. Continuous variables in the survey data are winsorized at the 99th percentile. In Panels A and D observations are inversely weighted by the number of observations in each land office and standard errors are clustered at the office level. See discussion in Section II.B.

Table D5—: Testing Predictions From Monopolistic Price-Discrimination Model

	Reported payment			
	(1)	(2)	(3)	(4)
Scorecard	477	390		
	(231)	(267)		
Scorecard × Overperform			579	
			(250)	
Scorecard × Underperform			375	
			(399)	
Overperform baseline			-371	-341
			(381)	(374)
Info. x Scorecard x Underperform				7
				(454)
No info. x Scorecard x Underperform				883
				(514)
Info. x Scorecard x Overperform				326
				(341)
No info. x Scorecard x Overperform				858
				(391)
Information treatment				72
				(257)
Start-month FE	Yes	Yes	Yes	Yes
Stratum FE	Yes	Yes	Yes	Yes
Sample	Appr. ≤ 25	Appr. > 25	Appr. ≤ 25	Appr. ≤ 25
Observations	677	1,332	677	677
Clusters	111	111	111	111

Note: This table reports the effects of the scorecard and information treatments on bribes for applicants with different processing speeds. Columns (1), (3), and (4) use data only from applicationsn approved within 25 working days while column (2) uses data from applications that have an approval time of more than 25 working days. No rejected or still ongoing applications are included. USD/BDT≈84.3. Bribe amounts are winsorized at the 99th percentile. Observations are inversely weighted by the number of observations in that land office. Standard errors are clustered at the office level. See discussion in Appendix A.2.

Table D6—: Heterogeneous Effects by Symbols on First Scorecard

	Time index (1)	Time index (2)	Time index (3)
Scorecard	0.137		
	(0.065)		
Scorecard × Baseline Processed Rank	-0.004		-0.001
	(0.003)		(0.005)
Scorecard × Baseline Pending Rank	-0.007		-0.003
	(0.002)		(0.004)
Scorecard × Two Thumbs Up		-0.003	0.102
		(0.139)	(0.288)
Scorecard × One Thumb Up		0.216	0.238
		(0.148)	(0.167)
Scorecard × No Thumbs		-0.130	-0.230
		(0.278)	(0.226)
Scorecard × One Up, One Down		-0.042	-0.068
		(0.116)	(0.111)
Scorecard × One Thumb Down		0.531	0.445
		(0.173)	(0.244)
Scorecard × Two Thumbs Down		0.645	0.638
		(0.205)	(0.367)
Start-month & Stratum FEs	Yes	Yes	Yes
Baseline Performance Controls	Yes	No	Yes
Thumb FE	No	Yes	Yes
Observations	195,136	195,136	195,136
Clusters	310	310	310

Note: This table shows the differential effect of scorecards with different performance rankings and different thumb symbols. Only applications made within 60 working days of the start of the experiment are included in the analysis. Column (1) shows the overall effect as well as the heterogeneity in the effect for offices receiving baseline scorecards with different rankings according to the two performance indicators. Column (2) shows the effect separately for offices receiving baseline scorecards with different numbers of thumbs up and thumbs down symbols. Column (3) shows the effect separately for offices receiving baseline scorecards with different numbers of thumbs up and thumbs down symbols while also controlling for the two performance rankings on the baseline scorecard. Standard errors are clustered at the office level. See discussion in Section III.A and Appendix C.5.

Table D7—: Effect of Scorecards on Bribes by Reason for Payment

Panel A: Overall effect	(1) Get work done	(2) Faster processing	(3) Avoid inconvenience	(4) Other reason	(5) No reason
Scorecard	75	49	5	26	25
	(121)	(39)	(19)	(35)	(10)
Panel B: Heterogeneous effects					
Scorecard × Overperform baseline	187	113	33	63	16
	(159)	(63)	(25)	(46)	(14)
Scorecard × Underperform baseline	6	-4	-18	-5	32
	(177)	(50)	(29)	(51)	(18)
Overperform baseline	-404	-136	-61	-70	24
	(198)	(60)	(30)	(66)	(16)
P-value: subgroup diff.	0.460	0.168	0.193	0.341	0.515
Start-month & Stratum FEs	Yes	Yes	Yes	Yes	Yes
Observations	2,869	2,869	2,869	2,869	2,869
Clusters	112	112	112	112	112
Control mean	718	124	63	159	15

Note: This table shows the effect on reported bribe payments by the reasons stated by the applicants for paying. Column (1) shows the effect on payments made to get the land office to work on the application at all. Columns (2), (3), (4), and (5) show effects on payments made to increase the processing speed, avoid hassle or inconveniences, payments made for other reasons, and payments made without any stated reason, respectively. Outcome variables are winsorized at the 99th percentile. Standard errors clustered at the office level. Observations are inversely weighted by the number of observations in that land office. See discussion in Sections III.B and IV.B.

Table D8—: Heterogeneous Effects on Processing Times, Controlling for Variables Interacted With Treatment

	Time index										
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Scorecard × Underperform baseline	0.204	0.200	0.205	0.208	0.209	0.240	0.218	0.197	0.219	0.205	0.265
	(0.121)	(0.122)	(0.121)	(0.121)	(0.119)	(0.121)	(0.120)	(0.121)	(0.119)	(0.121)	(0.121)
Scorecard × Population (100k)		-0.016									-0.034
		(0.033)									(0.035)
Scorecard × Area (1k km^2)			0.148								0.251
			(0.357)								(0.467)
Scorecard × Population density				0.001							0.001
				(0.005)							(0.006)
Scorecard × House quality index					-0.008						-0.012
					(0.063)						(0.084)
Scorecard × Time used e-gov system						-0.003					-0.003
						(0.016)					(0.016)
Scorecard × Apps. received (1k)							0.199				-0.217
							(0.372)				(0.385)
Scorecard × Female ACL								-0.039			-0.126
								(0.141)			(0.146)
Scorecard × No ACL									-0.297		-0.445
									(0.154)		(0.163)
Scorecard × Land size (st. dev.)										-0.008	-0.004
										(0.013)	(0.011)
Scorecard & Baseline controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Start-month & Stratum FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1,034,688	1,034,688	1,034,688	1,034,688	1,034,688	1,034,688	1,034,688	1,034,688	1,034,688	1,007,014	1,007,014
Clusters	311	311	311	311	311	311	311	311	311	311	311

Note: This table reports the heterogeneous effects on processing time by office baseline performance, controlling for other variables interacted with the treatment. The first row of estimates shows the difference in the effect between under- and over-performing offices. Column (1) shows the main specification. Columns (2)-(5) control for the treatment interacted with the sub-district's population, areas, population density and, as a proxy for wealth, a house quality index constructed using the share of houses at least partly built with concrete or bricks, that has a sanitary toilet, and that has electricity. Data on population and house quality from BBS (2014). Columns (6) and (7) control for the number of months the office had used the e-governance system at baseline and the number of applications the office received in the baseline month. Columns (8)-(10) control for the ACL being female at baseline, no ACL being appointed at baseline, and the size of the land the application was for. Column (11) includes all the controls. Standard errors are clustered at the office level. Observations are inversely weighted by the number of observations in that office. See discussion in Section III.B.

Table D9—: Heterogeneous Effects on Bribe Payments, Controlling for Variables Interacted With Treatment

	Typical bribe payment										
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Scorecard × Overperform baseline	2,169	2,227	2,090*	2,221*	2,317*	1,886	2,044	2,044	2,088	2,181	1,814
	(1,252)	(1,247)	(1,002)	(1,042)	(986)	(1,281)	(1,291)	(1,289)	(1,280)	(1,214)	(1,001)
Scorecard × Population (100k)		-265									-441
		(409)									(261)
Scorecard × Area (1k km^2)			-8,444								-2,436
			(4,277)								(5,118)
Scorecard × Population density				-63							-112
				(62)							(69)
Scorecard × House quality index					754						972
					(751)						(814)
Scorecard × Time used e-gov system						271					245
						(189)					(136)
Scorecard × Apps. received (1k)							-1,253				1,555
							(4,537)				(3,622)
Scorecard × Female baseline ACL								775			-764
								(1,508)			(1,132)
Scorecard × No baseline ACL									442		-1,355
									(1,708)		(1,363)
Scorecard × Land size (st. dev.)										113	-118
										(379)	(345)
Treatment & Baseline control	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Start-month & Stratum FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1,802	1,802	1,802	1,802	1,802	1,802	1,802	1,802	1,802	1,726	1,726
Clusters	112	112	112	112	112	112	112	112	112	112	112

Note: This table reports the heterogeneous effects on bribe payments by office baseline performance, controlling for other variables interacted with the treatment. The first row of estimates shows the difference in the effect between under- and overperforming offices. Column (1) shows the main specification. Columns (2)-(5) control for the treatment interacted with the sub-district's population, areas, population density and, as a proxy for wealth, a house quality index constructed using the share of houses at least partly built with concrete or bricks, that has a sanitary toilet, and that has electricity. Data on population and house quality from BBS (2014). Columns (6) and (7) control for the number of months the office had used the e-governance system at baseline and the number of applications the office received in the baseline month. Columns (8)-(10) control for the ACL being female at baseline, no ACL being appointed at baseline, and the size of the land the application was for. Column (11) includes all the controls. Standard errors are clustered at the office level. Observations are inversely weighted by the number of observations in that office. See discussion in Section III.B.

Table D10—: Effect on Processing Times: Alternative Specifications

	Time index					
Panel A: Overall effect	(1)	(2)	(3)	(4)	(5)	(6)
Scorecard	0.089	0.125	0.140	0.097	0.137	
	(0.089)	(0.070)	(0.061)	(0.076)	(0.057)	
Scorecard × Post						0.156
						(0.080)
Panel B: Heterogeneous effects						
Scorecard × Overperform	-0.009	0.001	0.005	0.031	0.032	
	(0.116)	(0.089)	(0.081)	(0.101)	(0.077)	
Scorecard × Underperform	0.197	0.264	0.280	0.169	0.250	
	(0.118)	(0.094)	(0.089)	(0.110)	(0.084)	
Overperform baseline	0.537	0.572	0.341	0.487	0.261	
	(0.110)	(0.088)	(0.103)	(0.130)	(0.116)	
Scorecard × Post						
× Overperform						0.040
						(0.074)
Scorecard × Post						
× Underperform						0.283
						(0.114)
P-value: subgroup diff.	0.21	0.04	0.02	0.37	0.06	0.13
Start-month FE	No	No	No	Yes	Yes	Yes
Stratum FE	No	No	No	Yes	Yes	No
Weighted by office	No	Yes	Yes	No	Yes	Yes
Baseline controls	No	No	Yes	No	Yes	No
Office FE	No	No	No	No	No	Yes
Observations	1,034,688	1,034,688	1,034,688	1,034,688	1,034,688	1,188,351
Clusters	311	311	311	311	311	306

Note: This table shows the robustness of the estimated effect of the scorecards on the time index to different regression specifications. The time index is constructed from two variables: whether the application was processed on time and the log of the overall processing time. Panel A shows the estimates of the overall effect and Panel B shows the estimates of the heterogeneous effects, as in Table 2. The specifications differ from the ones used in those tables in the following ways: column (1) shows the estimate from uniformly weighted regressions with no controls. Column (2) uses no controls. Column (3) uses no strata or month fixed effects, but it controls for baseline month values for the number of applications processed within 45 working days, the number of applications pending beyond 45 working days, the number of applications received, and the percentage of applications received in the month two months before the baseline that were processed within the time limit. Column (4) shows the estimate from a uniformly weighted regression. Column (5) controls for the baseline month controls. Column (6) shows the estimate from a regression, including applications made 45 working days before the start of the experiment (for the 306 offices that have such data) and using office fixed effects. Standard errors are clustered at the office level. See discussion in Section III.D.

Table D11—: Scorecards' Effect on Bribes: Alternative Specifications

	Typical payments						Reported payments						To gov. off.
Panel A: Overall effect	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
Scorecard	1,194	1,164	875	688	1,890	482	388	386	307	238	610	194	245
	(717)	(698)	(623)	(522)	(853)	(449)	(196)	(195)	(172)	(179)	(264)	(127)	(141)
Panel B: Heterogeneous effects													
Scorecard x Overperform	1,808	1,602	2,267	1,325	2,811	1,576	647	648	659	519	1,011	468	462
	(920)	(864)	(827)	(612)	(1,067)	(585)	(234)	(219)	(233)	(207)	(418)	(168)	(173)
Scorecard x Underperform	666	811	-393	-215	1,085	-534	175	175	28	75	255	-21	82
	(1,068)	(1,069)	(923)	(813)	(1,320)	(646)	(303)	(300)	(241)	(235)	(348)	(172)	(208)
Overperform baseline	-1,219	-1,272	-2,138	-1,916	-2,055	-1,667	-455	-634	-759	-1,005	-634	-636	-565
	(837)	(919)	(982)	(843)	(1,149)	(754)	(244)	(255)	(280)	(250)	(404)	(202)	(237)
P-value: subgroup diff.	0.42	0.57	0.04	0.06	0.31	0.02	0.22	0.21	0.06	0.10	0.19	0.05	0.16
Start-month FE	No	No	Yes	Yes	Yes	Yes	No	No	Yes	Yes	Yes	Yes	Yes
Stratum FE	No	No	Yes	Yes	Yes	Yes	No	No	Yes	Yes	Yes	Yes	Yes
Weighted by office	No	Yes	No	Yes	Yes	Yes	No	Yes	No	Yes	Yes	Yes	Yes
PDS controls	No	No	No	Yes	No	No	No	No	No	Yes	No	No	No
Percentile Winsorized	99	99	99	99	No	95	99	99	99	99	No	95	99
Observations	1,802	1,802	1,802	1,802	1,802	1,802	2,869	2,869	2,869	2,869	2,869	2,869	2,869
Clusters	112	112	112	112	112	112	112	112	112	112	112	112	112

Note: This table shows the robustness of the estimated effect of the scorecards on bribes. Panel A shows the estimates of the overall effect and Panel B of the heterogeneous effects. Columns (1)-(6) show the effect on the response to the question of how much it is "normal for a person like yourself to pay" in bribes. Columns (7)-(12) show the effect on reported bribe payments to government officials or agents. The specifications differ from the specification in Table 3 in the following ways: columns (1) and (7) use uniformly weighted regressions with no fixed effects. Columns (2) and (8) use regressions with no fixed effects. Columns (3) and (9) use uniformly weighted regressions. Columns (4) and (10) use post-double-selection (PDS) to select controls among 154 variables generated from the baseline administrative and survey variables unlikely to have been affected by the experiment (e.g. applicant characteristics; land value, size, and usage; whether e-governance system was used; enumerator conducting the survey; and indicators for each quartile of continuous control variables). Columns (5) and (11) use unwinsorized outcome variables. Columns (6) and (12) use outcome variables winsorized at the 95th percentile. Column (13) uses payments made directly to government officials as the outcome variable, excluding payments made to agents. Standard errors are clustered at the office level. See discussion in Section III.D.

Table D12—: Effect on Processing Times: Controlling for Applications Received by Office

	≤45 working days	ln(working days)	Time index
Panel A: Overall effect	(1)	(2)	(3)
Scorecard	0.059	-0.122	0.128
	(0.027)	(0.059)	(0.059)
ln(Applications Received)	-0.021	0.066	-0.057
	(0.018)	(0.038)	(0.038)
Panel B: Heterogeneous effects			
Scorecard × Overperform	0.009	-0.038	0.029
	(0.037)	(0.079)	(0.079)
Scorecard × Underperform	0.113	-0.214	0.234
	(0.040)	(0.088)	(0.088)
Overperform baseline	0.197	-0.324	0.384
	(0.050)	(0.107)	(0.108)
Scorecard × ln(Applications Received)	-0.006	0.007	-0.010
	(0.029)	(0.059)	(0.061)
ln(Applications Received)	-0.023	0.070	-0.061
	(0.021)	(0.046)	(0.046)
P-value: subgroup diff.	0.059	0.140	0.086
Start-month & Stratum FEs	Yes	Yes	Yes
Observations	1,034,688	1,034,688	1,034,688
Clusters	311	311	311
Overperformers: control mean	0.72	45.91	0.32
Underperformers: control mean	0.51	68.94	-0.11

Note: This table shows the robustness of the estimated effect of the scorecards on processing time to controlling for the number of applications the office received in the same calendar month as the application. The analysis is identical to that of Table 2 with the control for the natural logarithm of the number of applications received added in Panel A, and that variable controlled for and interacted with the treatment in Panel B. Standard errors are clustered at the office level. See discussion in Section III.D.

Table D13—: Effect on Bribes: Controlling for Applications Received by Office

	Amount		Any bribe		Amount if > 0	
Panel A: Overall Effect	(1)	(2)	(3)	(4)	(5)	(6)
Scorecard	963.77	305.08	-0.02	-0.00	1,531.37	1,183.58
	(618.75)	(183.29)	(0.02)	(0.02)	(764.42)	(457.83)
ln(Applications Received)	335.23	108.67	-0.01	0.01	496.81	159.15
	(289.53)	(104.05)	(0.01)	(0.01)	(336.36)	(243.55)
Panel B: Heterogeneous effects						
Scorecard × Overperform	641.28	2,076.89	0.04	0.03	1,734.61	2,483.68
	(227.40)	(751.84)	(0.03)	(0.03)	(615.85)	(929.56)
Scorecard × Underperform	42.31	-40.30	-0.04	-0.06	747.87	745.91
	(253.61)	(994.03)	(0.03)	(0.03)	(633.71)	(1,237.69)
Overperform baseline	-846.34	-1,829.70	-0.09	-0.09	-1,388.73	-1,358.73
	(279.42)	(1,027.27)	(0.03)	(0.04)	(694.22)	(1,232.02)
Scorecard × ln(Applications Received)	-37.33	84.49	0.01	-0.01	-277.39	255.43
	(188.71)	(552.24)	(0.02)	(0.02)	(424.67)	(621.24)
ln(Applications Received)	146.37	285.59	0.01	-0.00	321.35	338.08
	(167.00)	(433.29)	(0.02)	(0.02)	(398.01)	(472.38)
P-value: subgroup diff.	0.08	0.09	0.07	0.07	0.27	0.26
Start-month & Stratum FEs	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2,869	1,802	2,869	1,802	779	1,324
Clusters	112	112	112	112	111	112
Overperformers control mean	944.4	5455	.2286	.7171	4131	7606
Underperformers control mean	1578	6726	.3161	.7881	4992	8535
Bribe measure	Reported	Typical	Reported	Typical	Reported	Typical

Note: This table shows the robustness of the estimated effect of the scorecards on bribes to controlling for the number of applications the office received in the period from March 2019 - July 2019, which is the period when most applications of the surveyed applicants were processed. The analysis is identical to that of Table 3 with the control for the natural logarithm of the number of applications received added in Panel A, and that variable controlled for and interacted with the treatment in Panel B. Standard errors are clustered at the office level. See discussion in Section III.D.

Table D14—: Scorecards' Effects on Office-by-Month Level Outcomes

Panel A: Overall effect	(1) IHS proc.≤45	(2) IHS pen.¿45	(3) Rank proc.	(4) Rank pen.	(5) Index
Scorecard	0.211	-0.104	2.096	2.021	0.090
	(0.121)	(0.143)	(1.684)	(1.771)	(0.061)
Panel B: Heterogeneous effects					
Scorecard × Overperform	-0.022	0.171	-0.708	-0.847	-0.066
	(0.147)	(0.219)	(2.187)	(2.619)	(0.082)
Scorecard × Underperform	0.466	-0.412	5.173	5.231	0.263
	(0.195)	(0.186)	(2.614)	(2.413)	(0.091)
Overperform baseline	0.393	-0.422	6.373	3.846	0.210
	(0.226)	(0.280)	(3.321)	(3.539)	(0.125)
P-value: subgroup diff.	0.048	0.049	0.089	0.095	0.008
Month & Stratum FEs	Yes	Yes	Yes	Yes	Yes
Baseline controls	Yes	Yes	Yes	Yes	Yes
Observations	4,516	4,516	4,516	4,516	4,516
Clusters	311	311	311	311	311

Note: This table reports the effect of the scorecards on office-by-month level outcomes. Panel A shows the estimates of the overall effect. Panel B shows the estimates of the heterogeneous effects. Column (1) shows the effect on the inverse hyperbolic sine (IHS) transformation of the number of applications processed within 45 working days. Column (2) shows the effect on the IHS of the number of applications pending beyond 45 working days. Column (3) shows the effect on the percentile ranking in terms of the number of applications processed within 45 working days. Column (4) shows the effect on the percentile ranking in terms of the number of applications processed within 45 working days; a higher number of pending applications leads to a lower ranking. Column (5) shows the result on an ICW index created with the outcome variables of columns (1)-(4); a higher index value indicates a better performance. For the control group, the index has a mean of zero and a standard deviation of one. Standard errors are clustered at the office level. See discussion in Section III.D.

Table D15—: Scorecards' Effect on Processing Times: Alternative Functional Forms and Imputation Techniques

	≤ 45 w. days	Working days		ln(w. days)	
	(1)	(2)	(3)	(4)	(5)
Scorecard	0.240*	-6.847*	-0.101*	-0.124**	-0.123**
	(0.134)	(4.120)	(0.0589)	(0.0581)	(0.0542)
Start-month FE	No	Yes	Yes	Yes	Yes
Stratum FE	No	Yes	Yes	Yes	Yes
Observations	1,034,688	1,034,688	1,034,688	1,034,688	1,006,272
Clusters	311	311	311	311	311
Specification	Logit	OLS	Neg. Binomial		
Imputation				Office mean	Drop obs.

Note: This table shows the robustness of the effect of the scorecards on processing times from Table 2 to different assumptions regarding the functional form of the relationship between the treatment and outcome variable and to the imputation procedure used to assign a processing time to the 2% of applications that are not yet processed. Column (1) uses a logit model to estimate the effect of the scorecards on the probability of an application being processed on time. Column (2) uses an OLS regression to estimate the effect on the untransformed number of working days. Column (3) uses a negative binomial regression. Column (4) uses the mean of processing times for applications *in that office* that were processed after the number of days that the application I am imputing the processing time for has been pending. Column (5) drops all applications that are not yet processed from the sample. Standard errors are clustered at the office level. Observations are inversely weighted by the number of observations in that land office. See discussion in Section III.D.

Table D16—: Scorecards' Effect on Applications Not Affected by Survey

Panel A: Overall effect	(1)	(2)	(3)
	≤45 w. days	ln(w. days)	Time index
Scorecard	0.068	-0.172	0.140
	(0.029)	(0.067)	(0.059)
Panel B: Heterogeneous effects			
Scorecard × Overperform	0.015	-0.043	0.031
	(0.041)	(0.088)	(0.082)
Scorecard × Underperform	0.126	-0.313	0.260
	(0.040)	(0.098)	(0.082)
Overperform baseline	0.234	-0.531	0.478
	(0.055)	(0.136)	(0.112)
P-value: subgroup diff.	0.055	0.042	0.051
Start-month & Stratum FEs	Yes	Yes	Yes
Observations	541,681	541,681	541,681
Clusters	310	310	310
Control mean	0.52	78.33	0.00

Note: This table reports the effect of the scorecards when restricting the sample to applications in offices where the survey did not take place and applications made one month or more before the start of the survey. Hence, these results are unlikely to have been affected by the survey activities. Standard errors are clustered at the office level. Observations are inversely weighted by the number of observations in that land office. See discussion in Section III.D.

Table D17—: Effect of Peer Performance List

	≤45 working days		ln(working days)	
	(1)	(2)	(3)	(4)
Peer Performance List	0.002	-0.007	-0.030	0.037
	(0.045)	(0.048)	(0.094)	(0.105)
Post × Peer Performance List		0.008		-0.060
		(0.045)		(0.095)
Scorecard		0.048		-0.124
		(0.038)		(0.090)
Post × Scorecard		0.020		-0.001
		(0.038)		(0.083)
Start-month & Stratum FEs	Yes	Yes	Yes	Yes
Observations	280,002	1,034,688	280,002	1,034,688
Clusters	155	311	155	311
Control mean	0.68	0.59	49.47	63.87

Note: This table reports the effect of the peer performance list. Columns (1)-(2) show the effect on the percentage of applications processed on time. Columns (3)-(4) show the effect on the log of processing time. Columns (1) and (3) show the effect using data from offices receiving the scorecards only. Columns (2) and (4) show the effect of both the peer performance and the scorecards using data from all offices, $Post$ refers to the application being made after the peer performance list intervention started. Standard errors are clustered at the office level. Observations are inversely weighted by the number of observations in that land office. See discussion in Section C.6.

Table D18—: Effects of Information Treatment and Scorecards on Bribes

	Typical payment			Reported payment		
	(1)	(2)	(3)	(4)	(5)	(6)
Information treatment	212	-326	-311	-31	-56	-63
	(490)	(767)	(760)	(150)	(184)	(188)
Scorecard		462			285	
		(719)			(218)	
Scorecard x Information		987			29	
		(1,105)			(278)	
Info. x Scorecard x Overperform			1,644			740
			(926)			(280)
No info. x Scorecard x Overperform			2,427			533
			(956)			(280)
Info. x Scorecard x Underperform			1,343			-26
			(1,411)			(334)
No info. x Scorecard x Underperform			-1,519			122
			(896)			(310)
Overperform baseline			-1,712			-824
			(983)			(294)
Start-month & Stratum FEs	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1,802	1,802	1,802	2,869	2,869	2,869
Clusters	539	112	112	570	112	112
Control mean	6,553	6,258		1,508	1,328	

Note: This table reports the effect of the scorecards and information treatments on bribe payments. Columns (1)-(3) show the effect on the response to the question of how much it is "normal for a person like yourself to pay" beyond the official fee. Columns (4)-(6) show the effect on reported payments to government officials or agents. All outcome variables are winsorized at the 99th percentile. The outcome variables are in BDT. USD/BDT≈84.3. In columns (1) and (4), standard errors are clustered at the office-by-day level. In columns (2)-(3) and (5)-(6), standard errors are clustered at the office level. Observations are inversely weighted by the number of observations in that land office. See discussion in Section V.C and Appendix C.7.

Table D19—: Scorecards' Effect on Timing of Bureaucrats' Transfers

Panel A: Overall effect	(1) Transfer	(2) Duration	(3) No ACL
Scorecard	0.003	0.641	0.002
	(0.006)	(0.707)	(0.023)
Panel B: Heterogeneous effects			
Scorecard × Overperform	0.007	0.393	0.008
	(0.008)	(0.970)	(0.032)
Scorecard × Underperform	-0.002	0.928	-0.006
	(0.008)	(1.057)	(0.035)
Overperform baseline	-0.001	-0.285	0.013
	(0.010)	(1.182)	(0.033)
P-value: subgroup diff.	0.464	0.712	0.784
Month FE	Yes	No	Yes
Stratum FE	Yes	Yes	Yes
Observations	4,516	306	4,516
Clusters	311	306	311
Control mean	0.07	12.54	0.12

Note: This table reports the effect of the scorecards on the timing of bureaucrats' transfers. Column (1) shows the effects on the percentage of ACLs transferred away from the office in a particular office-month, using data for each office month after the start of the experiment until the last month of the experiment (March 2020). Column (2) shows the effects on the duration of the posting in months for the first bureaucrat to hold the position as ACL in each of the offices in the experiment. Column (3) shows the effects on not having any ACL in a particular office-month. The data is administrative data from the e-governance system. Standard errors are clustered at the office level, except for in column (2), where heteroskedasticity robust standard errors are used. See discussion in Appendix C.7.

Table D20—: Treatment Effects on Expected Processing Time

	ln(Expected processing time)				
	(1)	(2)	(3)	(4)	(5)
Scorecard	-0.006		-0.000		
	(0.035)		(0.041)		
Information treatment		-0.049	-0.044		
		(0.022)	(0.024)		
Scorecard x Information			-0.009		
			(0.042)		
Scorecard × Overperform baseline				0.046	
				(0.050)	
Scorecard × Underperform baseline				-0.041	
				(0.049)	
Info. x Scorecard x Overperform					0.018
					(0.056)
No info. x Scorecard x Overperform					0.073
					(0.057)
Info. x Scorecard x Underperform					-0.067
					(0.051)
No info. x Scorecard x Underperform					-0.014
					(0.057)
Overperform baseline				-0.129	-0.129
				(0.058)	(0.058)
Start-month & Stratum FEs	Yes	Yes	Yes	Yes	Yes
Observations	2,467	2,467	2,467	2,467	2,467
Clusters	112	561	112	112	112
Control mean	61	68	62		

Note: This table reports the effect of the scorecards and information treatments on expected processing times at the time of the in-person interview. The outcomes variable is the log transformation of the sum of the expected future processing time and the processing time already incurred at the time of the in-person interview, winsorized at the 99th percentile. Standard errors are clustered at the office level, except for in column (2), where standard errors are clustered at the office-by-survey-day level, as that is the level of randomization for the information treatment. Observations are inversely weighted by the number of observations in that land office. See discussions in Section V.C and Appendix C.7.

Table D21—: Associations between Performance, Corruption, Tenure, and Promotions

Panel A	Promotion					
	(1)	(2)	(3)	(4)	(5)	(6)
Underperformance	-0.0021			0.0050	0.0047	
	(0.0019)			(0.0066)	(0.0050)	
Ave. Typical Bribe (Thousands)		-0.0098		0.0274		
		(0.0111)		(0.0504)		
Ave. Reported Bribe (Thousands)			-0.0302		0.1217	
			(0.0372)		(0.1552)	
Underperformance × Typical bribe				-0.0008		
				(0.0010)		
Underperformance × Reported Bribe					-0.0030	
					(0.0029)	
Year Joined Civil Service						-0.2436
						(0.0239)
Observations	302	111	111	111	111	295
Dep. Var. Mean	0.68	0.79	0.79	0.79	0.79	0.67
Panel B	Suspention or OSD posting					
	(1)	(2)	(3)	(4)	(5)	(6)
Underperformance	0.0029			0.0017	-0.0007	
	(0.0010)			(0.0033)	(0.0024)	
Ave. Typical Bribe (Thousands)		0.0059		-0.0126		
		(0.0057)		(0.0249)		
Ave. Reported Bribe (Thousands)			0.0462		-0.1076	
			(0.0186)		(0.0743)	
Underperformance × Typical bribe				0.0003		
				(0.0005)		
Underperformance × Reported Bribe					0.0028	
					(0.0014)	
Year Joined Civil Service						0.0076
						(0.0142)
Observations	302	111	111	111	111	295
Dep. Var. Mean	0.060	0.045	0.045	0.045	0.045	0.061

Note: This table shows associations between three bureaucrat characteristics: performance, corruption, and tenure, on the one hand, and bureaucrats career outcomes on the other hand. The first career outcome is promotions (Panel A), the second career outcome is being suspended or posted as a OSD, which are considered to be negative outcomes for the bureaucrats. Column (1) shows the associations between the average performance, as measured by the performance scorecards, across all months the bureaucrat was posted as ACL. Columns (2) and (3) shows the associations between the average bribe payments, as measured in the survey data. Columns (4) and (5) show the associations between performance, bribe payments, and the interaction between the two. Column (6) shows the associations with the year that the bureaucrat joined the civil service. Heteroskedasticity robust standard errors. See discussion in Appendix B.1.

Table D22—: Scorecards' Effects on Visits, Processing Time, and Satisfaction

Panel A: Overall effect	(1) Visits	(2) ≤45 w. days	(3) ln(w. days)	(4) Satisfaction
Scorecard	-1.266	0.040	-0.059	-0.038
	(0.422)	(0.030)	(0.049)	(0.060)
Panel B: Heterogeneous effects				
Scorecard × Overperform baseline	-1.054	0.012	-0.005	-0.080
	(0.606)	(0.049)	(0.078)	(0.081)
Scorecard × Underperform baseline	-1.272	0.055	-0.088	-0.016
	(0.636)	(0.041)	(0.066)	(0.086)
Overperform baseline	-1.676	0.113	-0.216	0.181
	(0.784)	(0.051)	(0.084)	(0.097)
P-value: subgroup diff.	0.818	0.523	0.437	0.598
Start-month & Stratum FEs	Yes	Yes	Yes	Yes
Observations	2,869	2,869	2,869	2,869
Clusters	112	112	112	112
Control mean	10.73	0.56	59.81	

Note: This table reports the effect of the scorecards on visits to land offices, processing times, and stated satisfaction measured in the survey data. Satisfaction was measured on a five-point scale, transformed into an index with a mean of zero and a standard deviation of one in the control group. Standard errors are clustered at the office level. Observations are inversely weighted by the number of observations in that land office. See discussion in Section C.2.

Table D23—: Spillovers on Applications Outside E-Governance System

Panel A: Overall effect	(1) Visits	(2) ≤45 w. days	(3) ln(w. days)	(4) Satisfaction
Scorecard	-1.11	0.00	-0.13	0.08
	(0.79)	(0.06)	(0.10)	(0.12)
Panel B: Heterogeneous effects				
Scorecard × Overperform baseline	1.10	-0.07	-0.00	0.06
	(1.59)	(0.12)	(0.22)	(0.18)
Scorecard × Underperform baseline	-1.47	0.07	-0.19	0.08
	(1.38)	(0.06)	(0.09)	(0.15)
Overperform baseline	-2.13	0.02	-0.23	0.08
	(2.25)	(0.12)	(0.22)	(0.26)
Start-month FE	Yes	Yes	Yes	Yes
Stratum FE	Yes	Yes	Yes	Yes
Observations	644	644	644	644
Clusters	58	58	58	58

Note: This table reports spillover effects of the scorecards on applications made outside the e-governance system. 58 of the 112 surveyed land offices had at least one application made outside the e-governance system. Standard errors are clustered at the office level. Observations are inversely weighted by the number of observations in that land office. See discussion in Appendix C.4.

Table D24—: Scorecards' Effect on Processing Times, First Randomization

	(1) ≤45 w. days	(2) ln(w. days)	(3) Time index
Scorecard	0.099	-0.169	0.196
	(0.048)	(0.109)	(0.106)
Start-month FE	Yes	Yes	Yes
Stratum FE	Yes	Yes	Yes
Observations	617,139	617,139	617,139
Clusters	112	112	112
Control mean	0.57	64.27	0.01

Note: The table shows results using data from the 112 land offices in the initial wave of randomization. Column (1) shows the effect on applications processed within the time limit. Column (2) shows the effect on the log of processing time. Column (3) shows the effect on an index combining the two outcome variables. The data contains all applications made between one month before the start of the experiment and 45 working days before the experiment ended. Observations are inversely weighted by the number of observations in that land office. Standard errors are clustered at the office level. See discussion in Appendix B.6.

Table D25—: Treatment Effects on Survey Attrition

	Attrition: Survey		Attrition: Typical bribe		No Response	
Panel A: Overall effect	(1)	(2)	(3)	(4)	(5)	(6)
Scorecard	0.027		-0.006		-0.025	
	(0.015)		(0.020)		(0.023)	
Information treatment		-0.002		0.020		0.023
		(0.012)		(0.024)		(0.025)
Panel B: Heterogeneous effects						
Scorecard × Overperform baseline	0.034		0.005		-0.017	
	(0.017)		(0.029)		(0.033)	
Scorecard × Underperform baseline	0.022		-0.017		-0.034	
	(0.026)		(0.030)		(0.033)	
Overperform baseline	-0.017		-0.014		-0.004	
	(0.022)		(0.031)		(0.036)	
Start-month & Stratum FEs	Yes	Yes	Yes	Yes	Yes	Yes
Control mean	0.09	0.11	0.44	0.43	0.38	0.36
Observations	3,213	3,213	3,213	3,213	2,869	2,869
Clusters	112	112	112	112	112	112

Note: This table reports the effect of the scorecards and information treatment on attrition. In columns (1)-(2), attrition is measured from the in-person survey to the phone follow-up survey. In columns (3)-(4), attrition is measured from the in-person survey to the answer about the typical bribe. In columns (5)-(6), the outcome is an indicator variable for not answering the typical bribe question, conditional on participating in the follow-up survey. Standard errors clustered at the office level. Observations inversely weighted by the number of observations in that land office. See discussion in Appendix C.1.

Table D26—: Lower Lee Bounds for Scorecards' Effects on Bribes

	Typical payment		Reported payment	
	(1)	(2)	(3)	(4)
Scorecard	864		238	
	(620)		(180)	
Scorecard × Overperform baseline		1,719		578
		(749)		(228)
Scorecard × Underperform baseline		-112		-21
		(965)		(256)
Overperform baseline		-1,802		-809
		(974)		(294)
Start-month & Stratum FEs	Yes	Yes	Yes	Yes
Observations	1,808	1,829	2,922	2,922
Clusters	112	112	112	112
Control mean	6,127		1,284	
Overperformers: Control mean		5,455		944
Underperformers: Control mean		6,726		1,578

Note: This table reports the lower Lee bounds for the effects on bribe payments. If the estimated effect of the scorecards on attrition was positive (negative), randomly selected observations equal to the differential attrition are added back into the treatment (control) group data and are assigned a value of zero (the largest reported bribe) for bribe payments. Outcome variables are winsorized at the 99th percentile. Standard errors are clustered at the office level. Observations are inversely weighted by the number of observations in that land office. See discussion in Appendix C.1.

Table D27—: Scorecards' Effects on Applications Received and Land Size

	ln(Applications received)	ln(Land size)
Panel A: Overall effects	(1)	(2)
Scorecard	-0.053	-0.028
	(0.073)	(0.063)
Panel B: Heterogeneous effects		
Scorecard × Overperform	-0.062	0.025
	(0.107)	(0.092)
Scorecard × Underperform	-0.044	-0.087
	(0.106)	(0.091)
Overperform baseline	0.147	-0.088
	(0.123)	(0.110)
Start-month FE	No	Yes
Stratum FE	Yes	Yes
Observations	311	1,007,014
Clusters		311

Note: This table reports the effects of the scorecards on the number of applications received and the land size of those applications. In column (1), observations are at the office level. In column (2), observations are at the application level. Panel A shows the estimates of the overall effect and Panel B shows the estimates of the heterogeneous effects. Data contains all applications made between one month before the experiment started and 45 working days before the experiment ended (August 13, 2018 to January 20, 2020). Standard errors are clustered at the office level. Observations in column (2) are inversely weighted by the number of observations in that land office. See discussion in Appendix C.4.

Table D28—: Scorecards' Effects on Rejection Rates

Panel A: Overall effect	(1) Rejection	(2) Previously rejected
Scorecard	-0.000	0.021
	(0.021)	(0.019)
Panel B: Heterogeneous effects		
Scorecard × Overperform baseline	0.014	0.056
	(0.029)	(0.038)
Scorecard × Underperform baseline	-0.016	-0.018
	(0.031)	(0.013)
Overperform baseline	-0.068	0.000
	(0.038)	(0.029)
P-value: subgroup diff.	0.49	0.08
Start-month & Stratum FEs	Yes	Yes
Observations	1,034,688	2,760
Clusters	311	112
Control mean	0.32	0.05

Note: This table reports the effect of the scorecards on rejections of applications. Column (1) shows the effect on the percentage of applications rejected in the administrative data. Column (2) shows the effect on applicants surveyed who returned after having had their application rejected, which is a proxy for incorrect rejections. Standard errors are clustered at the office level. Observations are inversely weighted by the number of observations in that land office. See discussion in Appendix C.4.

Table D29—: Information Treatment Effect: Alternative Treatment Variables

Panel A:	Typical bribe payment			
	(1)	(2)	(3)	(4)
Information treatment	212	120	211	221
	(490)	(493)	(501)	(548)
Observations	1,802	1,733	1,802	1,448
Clusters	539	536	539	437
Panel B:	ln(Expected processing time)			
Information treatment	-0.049	-0.044	-0.007	-0.029
	(0.022)	(0.022)	(0.022)	(0.024)
Observations	2,467	2,467	2,467	2,064
Clusters	561	561	561	462
Start-month & Stratum FEs	Yes	Yes	Yes	Yes
Sample	Full	Full	Full	Correct

Note: This table shows the estimates of the effects of the information intervention using alternative treatment variables. The results investigated is the overall effect on bribe payments, the effects on bribe payments among offices receiving the scorecard treatment, and the effect on the expected processing time at the time of the in-person survey. Column (1) uses my preferred treatment variable based on the median treatment delivered in a land office survey day. Column (2) uses the actual treatment delivered for each applicant. Column (3) uses the assigned treatment. Column (4) uses the assigned treatment, restricting the sample to applicants who received the assigned treatment. Standard errors clustered at the office level. Observations are inversely weighted by the number of observations in land office. See discussion in Appendix C.8.

*

References

Bai, Jie, Seema Jayachandran, Edmund J Malesky, and Benjamin A Olken. 2019. "Firm growth and corruption: Empirical evidence from Vietnam." *The Economic Journal*, 129(618): 651–677.

Bandiera, Oriana, Michael Carlos Best, Adnan Qadir Khan, and Andrea Prat. 2021. "The allocation of authority in organizations: A field experiment with bureaucrats." *The Quarterly Journal of Economics*, 136(4): 2195–2242.

BBS. 2014. "Bangladesh Population and Housing Census 2011." Bangladesh Bureau of Statistics (BBS), Government of Bangladesh Dataset.

Becker, Gary S, and George J Stigler. 1974. "Law enforcement, malfeasance, and compensation of enforcers." *The Journal of Legal Studies*, 3(1): 1–18.

de la Sierra, Raúl Sanchez, Kristof Titeca, Haoyang Stan Xie, Albert Malukisa Nkuku, and Aimable Amani Lameke. 2022. "The Real State: Inside the Congo's Traffic Police Agency." National Bureau of Economic Research Working Paper 30258.

Do, Quoc-Anh, Trang Van Nguyen, and Anh Tran. 2021. "Do People Pay Higher Bribes for Urgent Services? Evidence from Informal Payments to Doctors in Vietnam." Working Paper.

Huntington, Samuel P. 1968. *Political Order in Changing Societies.* New Haven:Yale University Press.

Kaufmann, Daniel, and Shang-Jin Wei. 1999. "Does "grease money" speed up the wheels of commerce?" National Bureau of Economic Research Working Paper 7093.

Lee, David S. 2009. "Training, wages, and sample selection: Estimating sharp bounds on treatment effects." *The Review of Economic Studies*, 76(3): 1071–1102.

Leff, Nathaniel H. 1964. "Economic development through bureaucratic corruption." *American Behavioral Scientist*, 8(3): 8–14.

Malesky, Edmund J, Dimitar D Gueorguiev, and Nathan M Jensen. 2015. "Monopoly money: Foreign investment and bribery in Vietnam, a survey experiment." *American Journal of Political Science*, 59(2): 419–439.

Mattsson, Martin. 2019. "Schedule for Information Provision." Spreadsheet. Available upon request.

Mattsson, Martin. 2023. "When Does Corruption Cause Red Tape? Bribe Discrimination Under Asymmetric Information." SSRN Working Paper 4395094.

Mussa, Michael, and Sherwin Rosen. 1978. "Monopoly and product quality." *Journal of Economic Theory*, 18(2): 301–317.

Niehaus, Paul, and Sandip Sukhtankar. 2013*a*. "Corruption dynamics: The golden goose effect." *American Economic Journal: Economic Policy*, 5(4): 230–69.

Niehaus, Paul, and Sandip Sukhtankar. 2013*b*. "The marginal rate of corruption in public programs: Evidence from India." *Journal of Public Economics*, 104: 52–64.

Olken, Benjamin A. 2007. "Monitoring corruption: Evidence from a field experiment in Indonesia." *Journal of Political Economy*, 115(2): 200–249.

Olken, Benjamin A, and Rohini Pande. 2012. "Corruption in developing countries." *Annu. Rev. Econ.*, 4(1): 479–509.

Reid, Otis, Jean Freddy Tshimanga Kabasababu, and Jonathan Weigel. 2019. "The supply of bribes: Evidence from roadway tolls in the D.R. Congo." Working paper.

Rose-Ackerman, Susan. 1978. *Corruption: A study in political economy.* New York:Academic Press.

Svensson, Jakob. 2003. "Who Must Pay Bribes and How Much? Evidence from a Cross Section of Firms." *The Quarterly Journal of Economics*, 118(1): 207–230.

Time and Date. 2021. "Public Holidays in Bangladesh." https://www.timeanddate.com/holidays/bangladesh/, Accessed: 22 November 2021.

Transparency International Bangladesh. 2016. "Corruption In Service Sectors, National Household Survey 2015." Research Report.

Weaver, Jeffrey. 2021. "Jobs for Sale: Corruption and Misallocation in Hiring." *American Economic Review*, 111(10): 3093–3122.