Working paper

International Growth Centre

Motivating bureaucrats

Group incentives and organisational performance in local governments (Pilot project)

Matteo Bobba
Gianmarco Leon
Leonard Wantchekon

September 2017

When citing this paper, please use the title and the following reference number:
S-89331-BEN-1

DIRECTED BY LSE UNIVERSITY OF OXFORD

FUNDED BY UKaid from the British people

Motivating Bureaucrats: Group incentives and Organizational Performance in Local Governments (Pilot Project) *

Matteo Bobba† Gianmarco Leon‡ Leonard Wantchekon§

FINAL Report, September 2017

Abstract

This report looks at the development and results of the pilot project of a large-scale field experiment in Benin aimed at enhancing public service provision by means of group and individual incentives tied to on-the-job productivity of local level bureaucrats. Extensive preliminary qualitative work was conducted in the summer of 2016 to assess the legitimacy, the straightforwardness, and the usability of potential performance measures in this context. In early 2017, two types of incentive schemes were implemented among 6 town-halls (communes), as the lowest level of administration in Benin. Accordingly, three communes were assigned to a group tournament amongst the teams (services) of their local governments, in which the best performing service received monetary and non-monetary awards as measured by the degree of completion of predetermined performance targets. In the remaining three communes, the group tournament was further complemented by an individual tournament, where the best performing worker of each commune – measured by comprehensive peer assessments – was rewarded similarly. We analyze the detailed administrative and survey data collected in each commune throughout the pilot intervention to document behavioral responses to the incentive schemes, and to evaluate the extent to which the intervention affects the delivery of public works to their local communities. In sum, we find some suggestive evidence that both group and individual incentives have stronger and more consistent effects on improving outcomes.

Keywords: Personnel Economics of the State, Team incentives, Pay for performance, Public good provision, Benin.

JEL Codes: D73, J45, O12.

*We thank the staff of the Institute of Empirical Political Economy (IREEP) at the African School of Economics, and in particular Laurent Biaou, Horace Gninafon for their dedication to the fieldwork activities, Ines Lacland for outstanding research assistance, and Damase Sossou for facilitating our collaboration with the government.

†Toulouse School of Economics, University of Toulouse Capitole and IZA. E-mail: matteo.bobba@tse-fr.eu.

‡Department of Economics and Business, Universitat Pompeu Fabra, Barcelona GSE and IPEG. E-mail: gianmarco.leon@upf.edu.

§Department of Politics and Economics, Princeton University and African School of Economics. Email: lwantche@princeton.edu.

1 Introduction

The large push towards the decentralization of the public sector in the developing world during the 1990's and 2000's was aimed at improving public service provision through a better representation of voters' preferences and increased accountability. However, in most of the developing word we still observe poor quality of public services. One hypothesis behind this pattern is that the pool of candidates who can effectively manage the bureaucracy at the local level is small and poorly qualified, and usually coupled with the presence of political appointees (unproductive workers) as a result of the extensive presence of patronage networks. Another plausible and non-exclusive explanation hinges upon the widespread presence of low-powered incentives schemes in the public sector.

Our pilot project seeks to analyze the effects of monetary incentives on on-the-job performance by exploring different potential mechanisms at work. Following the theoretical work in Holmstrom [1982] and the empirical applications in Lazear [2000], a number of studies documented the different aspects of the impact of high-power monetary incentives on performance and productivity.[1] Pay for performance programs are typically based on individual performance and they use performance measures that reflect individual worker's effort, closely linked to their specific tasks. Unlike frontline service providers, jobs in a bureaucracy are highly heterogeneous, involve multiple tasks, and the final outcomes (i.e. public goods) are a product of effort from many agents, making it extremely difficult to implement high-powered incentives.

The scale up of the piloted project will be the first, to our knowledge, to look at (financial and non-financial) group-based incentives in the public sector. The potential mechanisms through which group incentives may increase productivity is through mutual monitoring, information sharing, innovation, trust, and overall job satisfaction of the workers. Moreover, individual incentives can further enhance the effect of group incentives (see Section 3) and we will assess how these two effects interact. Using a unique measure of team-level output that is comparable between teams within local governments, and very detailed individual survey-based measures, the results from this pilot project will contribute to the literature analyzing the effects of pay for performance by focusing on bureaucrats, a group of workers that (i) has

[1] Many find improved performance as a result of pay-for-performance schemes. For example, Miller and Babiarz [2013] provides a comprehensive review of evidence to highlight the positive incentive schemes in the health sector in low and middle income countries. Gertler and Vermeersch [2013] and Kazianga et al. [2014] find evidence of the positive impact of pay-for-performance in Rwanda on the productivity of public health providers and on the instances of HIV testing amongst couples, respectively. Monetary incentives also improve the quality of the pool of applicants for positions in the public sector [*Strengthening State Capabilities: The Role of Financial Incentives in the Call to Public Service*, 2013].

rarely been studied from a micro perspective, (ii) form a large share of governments' payrolls, and (iii) are key actors in determining the success of development policies. Despite the fact that the production of public goods involve interlinked tasks, effort complementarity, and present the risk of free riding, issues addressed explicitly by team based incentives, to our knowledge have only been studied in the private sector (e.g. Bandiera et al. [2013], Friebel et al. [2017]).

2 Context and Performance Measurement

The proposed scale up of the current pilot will be implemented among the universe of workers among the 77 commune governments in Benin. As the lowest level of administration, they are the entities that provide public goods and services to the Beninese population. Each local government governed by the mayor and his council under 4-year terms, and is composed of several working groups (henceforth services), with between 10 to 15 services per local government, as identified from the pilot project. Services are teams ranging from 1 to 15 workers that are led by a service chief and are responsible for a specific set of outlined tasks or activities, with little to no overlap between service functions (a complete list of the studied services is found in the appendix). The annual functions and specific tasks assigned to each service are defined in great detail at the beginning of each calendar year in the annual development plan (PAD). This programmatic document is the main source of information used to request transfers, not only from the central government but also from technical and financial partners. It outlines the main development objectives for the year, the lines of action within each objective, the specific activities that are conducive to these lines of action [2], and the service responsible for executing each specific tasks within the activities.

In 2013, the central government collaborated with German development cooperation agency (GIZ) to develop a measurement system to monitor the performance of local governments based on the PAD. In this framework, communes are requested to report back to the central government on the progress of each annually prescribed task, indicating precisely the percentage of execution in each activity. This measure provides Pareto weights to reflect the importance of each activity in accomplishing a line of action. Overall, the measurement system presents several unique features that align with our project based on group performance. First, it allows for a measure of progress and productivity at both the service- and commune-level simultaneously. Second, it focuses on procedures rather than outcomes to give a more

[2] Activities are extremely detailed as a line of action can have anywhere between 1 to 35 activities.

objective measure of public goods delivery. Third, it conveys precise and well-defined information on the composition of the teams within the commune. Fourth, the pre-existence of this procedure within the local government allows for the extraction of already available data and we therefore neither augment the workload nor disrupt the team dynamics of the local bureaucrats. Finally, as an official instrument intended for budgetary purposes, there is little incentive or scope to manipulate the measure as a result of the treatment.

It is unusual to have access to a performance measure that is comparable between groups of individuals who carry out inherently heterogeneous tasks. However, our study also requires a measure of individual performance, which is notoriously harder to assess in settings where outcomes are the result of a joint production function. A tool commonly used in the private sector to assess individual performance and to provide feedback to workers is the '360 evaluation'. The idea is to distribute a short questionnaire to all members of a service asking to evaluate their most frequented co-workers' performance (see Appendix B for the instrument used). We then average the scores received by each worker from those peers to evaluate the performance at the individual level.

3 Pilot Project and Main Hypothesis

In January 2017, we began the pilot project in six communes to explore the effectiveness of each treatment in improving the outcome of interest (i.e. public good provision of the local government); to test and fine-tune survey instruments and performance measures before the project scale-up; and to get an idea of the qualitative and quantitative (albeit descriptive) behavioral responses to the intervention that we propose to evaluate on a larger scale. We began the pilot intervention by collecting baseline information in six communes. The sample size was of 212 workers and 60 citizens. Once the baseline data collection was completed, six communes were randomly allocated to receive Treatment 1 or Treatment 2, defined as follows:

1. Treatment 1: Group Incentive Scheme Only
   In three communes all services participate in a tournament in which the best-performing service of each commune receive a monetary and non-monetary award. Specifically, each worker within the winning service is to be granted the amount of a full month's salary and acknowledgment in a public ceremony. Again, the service's performance is measured by the execution rates of the activities outlined in the PAD, as outlined in Section 2.

2. Treatment 2: Group and Individual Incentive Schemes
   The three remaining communes are subject to both a group and an individual incentive scheme. Hence, in addition to the team tournament, the best worker of each commune - as measured by the score obtained in the 360 evaluation (see Section 2) - is rewarded with the same rewards as in the group tournament. Individual incentives are hypothesized to increase even further the productivity of workers, particularly for those in services that are expected to have lower chances of being the best performers because of their ex-ante performance or the high instances of free-riding.

After collecting baseline information, all workers were invited to a public meeting between February 15 and February 28, 2017. They received detailed information on the prizes, the measurement tools, and the general program of the project, and were invited to ask clarifying questions. Additionally, we distributed flyers and posters outlining the project details to the local government (see Figure A.1 for an example). The end-line of the pilot intervention was set to June 30, 2017 and the follow-up data was collected in July 2017. The measurement and payout take place at the end of the pilot period. The end-line evaluations are further complemented by audits and focus groups for further validation of the main findings.

As previously noted, we hypothesize that group-based pay-for-performance should affect bureaucrat's productivity through several channels:

1. Mutual monitoring: Individual benefits are a function of other service member's performance, thus providing incentives for workers to monitor each other's effort.

2. Reduce negative effects of political appointees: Our qualitative work showed that political appointees are ubiquitous in local governments. Their presence demotivates workers and hurts team productivity. Team incentives reduce this negative impact by reducing free riding through increased worker motivation and monitoring.

3. Collaboration: Public good provision is the result of a production function with effort complementarities. Team incentives increase the chances that workers collaborate with each other and share relevant information.

4. Motivation and effort: There is a direct link between a higher expected pay and motivation, which should increase productivity.

5. Innovation: The presence of performance-based incentives should enhance innovation for workers seeking creative solutions to existing implementation problems.

6. Competition: Individual incentives can enhance competition between workers, and its effects can either foster innovation or undermine collaboration and trust.

It is important to note that the last two channels are also expected to be enhanced by the individual-based incentive schemes.

4 Data

The final dataset used is a combination of the following instruments and administrative data: (1) surveys with the universe of commune workers; (2) surveys with a random sample of citizens; (3) audits of selected public works at the end-line to verify the validity of our performance measures; (4) individual performance using the 360 evaluation [3]; and (5) reported evaluation of PADs.

The workers' survey forms a panel dataset that tracks government workers over a six month interval, with approximately 76.9 percent successful re-interviews. [4] We can therefore observe changes in outcomes at the individual, service, and commune level, following the treatment implementation. Face-to-face interviews and focus groups were conducted prior and upon completion of the pilot project, gathering demographic and behavioral information. Summary statistics presented in Table 1 verify proper randomization of the treatments. Workers were found to have relatively similar traits with no significant difference between treatment groups, except for some small differences in age and education.

The survey conducted amongst citizens provides census data gathering demographic and preferential traits of the population in the six targeted communities. This data allows us to analyze the consequent citizens' satisfaction level for their local government's provision of public goods. Table 2 presents the baseline summary statistics of citizens and shows no major differences between treatment groups. As census data was used for this part of the analysis, we also ensure that there are no significant difference between the baseline and end-line samples in Table 3.

[3] Before the pilot activities, we pre-tested the instrument of the 360 evaluation in two communes with 20 workers, both from the same team and across different teams.

[4] This high attrition rate can be explained by the fact that many of the workers who were not re-interviewed had been assigned to another town-hall at the time of the end-line data collection.

Table 1: Workers Baseline Characteristics by Incentive Group

Incentives	**Mean** Group	**Mean** Group and Individual	**Diff**	***p*-value**
Marital Status	1.225	1.146	-0.0784	0.3194
	(0.5983)	(0.5384)		
Age	37.618	40.780	3.163**	0.0118
	(8.4551)	(9.2759)		
Gender	0.7978	0.7561	-0.0417	0.4768
	(0.4040)	(0.4312)		
Education	13.067	11.244	-1.824***	0.0088
	(4.6093)	(5.1888)		
Number of Children	0.8989	0.9512	0.0523	0.1438
	(0.3032)	(0.2163)		
HH size	5.764	6.195	0.4311	0.3171
	(3.2510)	(2.9661)		
Job Length	4.528	4.114	-0.4143	0.5988
	(6.0661)	(5.3289)		
Wage	88851	86168	-2683.5	0.6274
	(37692)	(41041)		
Hours Worked	42.966	46.114	3.148	0.1941
	(16.6839)	(17.8350)		
N	89	123		

*$p < 0.1$; ** $p < 0.05$, *** $p < 0.01$

Note: st. dev. in parentheses.

Table 2: Citizens' Baseline Characteristics by Incentive Group

Incentive group	**Mean** Group	Group and Individual	**Diff**
Gender	0.5	0.5	0
Age	38.533	40.167	1.633
Marital Status	1.333	1.267	-0.067
Education	1.767	2.033	0.267
Work	0.333	0.133	-0.200*
N	30	30	

*$p < 0.1$; ** $p < 0.05$, *** $p < 0.01$

Note: st. dev. given in parentheses

The performance measure used to assign the awards is the sum of the real weights of the activities the service is involved in, multiplied by the physical execution rate of each activity (i.e. the percentage of completion), and divided by the sum of the real weights of the activities carried out by a given service. Using this metric, we rank services within each commune according to a weighted average of the level of performance and its growth rate over the six previous months. The validity of this measure is confirmed in the following sections using quantitative and qualitative evidence gathered from the surveys, face-to-face interviews and focus groups.

For our pilot project analysis, a simple comparison of outcomes before and after the intervention will provide us with an initial comprehensive (albeit descriptive in nature) assessment of the impact of the intervention on final and intermediate outcomes. Before we turn to the quantitative data analysis, we briefly look at the qualitative information to get an initial idea of the observed impacts, to validate the performance measures used in the subsequent quantitative analysis, and to help us in the design of the scale-up of the project.

5 Qualitative Data

First, we look at the legitimacy of our performance measure reported as the physical execution rate (in French, *le taux d'execution physique*; TEP) in the PAD by independently auditing the progress made on selected tasks and activities in our pilot communes. We then compare our assessments with the scores reported in the official government document. Out of the total 14 audited activities, 10 of them coincided exactly with those reported in the PAD.[5] We will further confirm this claim in the next section using correlations of different measures with the performance measure.

Next, we turn to qualitative findings from the workers' surveys and focus groups. Workers had identified monetary incentives as a strong motivation to increase their productivity in the workplace before the intervention was announced. They reported feeling very excited about the project during the focus groups. Furthermore, workers believed that the incentive schemes had been well explained by the research team and were perceived as being fair programs. The public meetings at baseline were identified as useful tools to increase awareness and generate credibility of the program. Importantly, our midline focus groups confirmed that all agents were aware of the on-going project and workers reported an increase in collaboration.

[5]Of those that did not match, one of the services reported a 0 percent progress despite the fact that the audit reported a 50 percent progress. The remaining three were all in the opposite direction.

All in all, the qualitative data from our pilot strongly suggest that the TEP score is a reliable measure of public good provision and that the incentive schemes may be effective in motivating workers to improve their efforts toward public good provision. This is not a surprising as workers in all communes had previously identified financial incentives as their biggest motivational force.

6 Empirical Analysis

We compare outcomes in each commune and/or group of communes under the same incentive scheme before and after the intervention. While this approach is clearly descriptive – as the two sets of communes may differ along many unobserved characteristics the are correlated with outcomes and as there may be potential confounding factors that vary over the period of the analysis – it does provide some first evidence at the effects of the incentive schemes piloted here. The two final outcomes of interest were identified as the change in citizen satisfaction level with the government's performance (as the ultimate beneficiaries) and the change in dynamics among the government workers (as the means to achieve their tasks). Through the survey design, we are also able to analyze the effect on a multitude of intermediate outcomes - namely job satisfaction, collaboration, trust, and pro-sociality tendencies - which comes to support our findings. In sum, we find that combining both individual and group incentives has a stronger and more consistent effect on ameliorating the workplace environment and citizen satisfaction.

We begin by validating the performance measure through some correlations at baseline, as previously noted in Section 5. For instance, hours of work were found to be positively correlated (0.21) with the productivity measure. In other words, we would expect an increase in productivity of the local government as total work hours increase. As seen in Table 4, we find an increase in hours worked under both incentive schemes, even though both are insignificant. [6] In the survey, respondents were also asked to identify the likelihood that co-workers in the same service complete an assigned task. Using this measure, we identify an insignificant decrease in the identification of free-riders, with a larger decrease in the Group + Individual incentive communes.

Next, we turn to the intermediate outcomes. As seen in Table 4, combining individual and group incentives had a larger and significant impact in reported satisfaction of workers with a 11.55 percent (or 7.20 percentage point) increase, compared to an insignificant 3.02

[6] These represent a 5.47 and a 6.66 percent changes for Treatment 1 and 2, respectively.

Table 3: Impact of Incentive Schemes on Worker's Environment : Intermediate Outcomes

	Group Incentives			Group and Individual Incentives		
	Mean		**Diff**	**Mean**		**Diff**
	Pre	Post		Pre	Post	
Free-riding Score	2.910	2.868	-0.0423	3.098	3.018	-0.080
Hours Worked	42.97	45.32	2.35	46.11	29.18	3.07
Job Satisfaction	0.6820	0.7026	0.0206	0.6236	0.6956	0.0720**
Collab: Helping Others	0.7921	0.8645	0.0723**	0.7683	0.8280	0.0597**
Collab: Reaching Out	0.7213	0.8237	0.1023***	0.6764	0.8070	0.1306***
Pro-Sociality	0.7315	0.7421	0.0106	0.6976	0.7719	0.0744*
N			212			190

*$p < 0.1$; ** $p < 0.05$, *** $p < 0.01$

percent (2.06 pp) increase among communes with only group incentives.[7] As seen in Table 5, a further decomposition by commune reveals that Pobe is the only commune that displays a significant increase of 13.57 percent (8.85 pp). The remaining four communes show an insignificant increase in reported satisfaction, while only Zagnanado reports a small and insignificant decrease. This is not a surprising result when compared to the changes in other intermediate outcomes in Table 4.

Indeed, we observe a more consistent increase in collaboration and pro-sociality behavior under both individual and group incentives. Using two different measures of collaboration, the analysis reveals an overall significant increase in self-reported collaboration amongst workers in both treatments. The first measure looks at the tendency of respondents to reach out to co-workers when they require help to achieve a given task in the workplace. All communes show a significant increase in collaboration rates with a 14.18 percentage (10.23 pp) increase and a 19.31 percent (13.06 pp) increase, respectively, in the reported likelihood of reaching out to another co-worker for collaboration. The second measure of collaboration studies the likelihood of the respondent to help a co-worker by collaborating with them. The analysis reveals that there is similar effect under both incentive schemes. The pro-sociality measure was constructed using the respondent's likelihood to contribute anonymously to help a coworker in need. A significant increase in pro-sociality tendencies was only found among the communes assigned to both incentive schemes.

A dictator game and a trust game were conducted to determine the level of trust amongst workers in each service. Respondents exhibit trust toward co-workers when they choose to

[7]Note that we normalize all scales hereinafter by dividing them by the total potential points based on the question so that the scale is 0-1.

Table 4: Job Satisfaction by Commune

	Mean Pre	Post	**Diff**	***p*-value**
Group Incentives				
Toucountouna	0.6409	0.7522	0.1113	0.1489
Bembereke	0.6595	0.6958	0.0364	0.5595
Zagnanado	0.7400	0.6690	-0.0710	0.2405
Group and Individual Incentives				
Allada	0.6208	0.7190	0.0982	0.1231
Grand-Popo	0.5720	0.600	0.0280	0.6420
Pobe	0.6520	0.7405	0.0885*	0.0667

*$p < 0.1$; ** $p < 0.05$, *** $p < 0.01$

give money (i.e. a value of 1 is given to the dummy) to a co-worker in the trust game but decide to keep the money (i.e. a value of 0) for themselves in the dictator game. The combination of these two answers points to trust amongst co-workers because the respondent believes that his co-worker would not choose the outcome (0, 40,000) in the trust game, an outcome that has been revealed to be dis-preferred to (10,000, 10,000) in the dictator game. By calculating the percent of respondents that provide this combination of answers in each time period, we see that there is a very similar increase in trust amongst co-workers, with a 30.17 percent increase (from 20.22 percent at baseline to 26.32 percent at end-line) in communes with only group incentives and a 29.43 percent increase (from 12.20 percent to 15.79 percent) in commune with both group an individual incentives. As reported in Table 6, only two communes, Pobe and Zagnanado, show an increase in trust that is statistically different from zero. Pobe is the only one to show an increase in trust behavior in both games, while there is only an increase in the trust game in Zagnanado.

Finally we look at the performance measure established by the Benin government, the physical execution rate (TEP). The decomposition by incentive scheme is reported in Table 6.[8] We find some significant increase in productivity under the group incentive scheme. It is important to note here that the productivity measure for the baseline takes into account a full year's work, whereas the measure of the end-line was taken mid-year and therefore only represents a lower bound of its potential.

Next, we consider the citizen survey to further explore the effects of the incentive schemes.

[8]The decomposition by commune and by service is also found in Tables 7 and 8, respectively, in the Appendix.

Table 5: Dictator v. Trust Game by Commune

	Mean		**Diff**	***p*-value**
	Pre	Post		
Group Incentives				
Toucountouna				
dictator_game	0.4091	0.3913	-0.0178	0.9058
trust_game	0.6818	0.6522	-0.0296	0.8376
Bembereke				
dictator_game	0.2703	0.4583	0.1880	0.1354
trust_game	0.4595	0.3333	-0.1261	0.3360
Zagnanado				
dictator_game	0.100	0.2069	0.1069	0.2612
trust_game	0.1667	0.4483	0.2816**	0.0185
Individual and Group Incentives				
Allada				
dictator_game	0.0851	0.1429	0.0578	0.3949
trust_game	0.1702	0.2381	0.0679	0.4318
Grand-Popo				
dictator_game	0.200	0.200	0	1
trust_game	0.240	0.433	0.193	0.1383
Pobe				
dictator_game	0.160	0.357	0.197**	0.0297
trust_game	0.24	0.5	0.26***	0.0092

*$p < 0.1$; ** $p < 0.05$, *** $p < 0.01$

Table 6: Group Performance Measure

	Mean		**Diff**	***p*-value**
	Pre	Post		
Group Incentives	0.3049	0.4183	0.1134***	0.002998
Individual and Group Incentives	0.6172	0.5456	-0.0716**	0.030738
Total	0.5028	9.4980	0.0048	0.8636

Three measures serve as proxy measures for public goods provision: the citizen's evaluation of economic conditions; their satisfaction level with different actors in the government; and the trust level in the local government's ability to manage specific tasks. All are reported by incentive scheme in Table 7. A first outcome of interest is the outlook on the current economic conditions of citizens, which was found to be positively correlated with the average TEP score of the local governments. Therefore, one would expect that an increase in local government productivity lead to an increase in the citizens' evaluation of current economic conditions. Findings show a significant increase in rating of current economic conditions under the group incentive scheme, compared to a smaller and insignificant increase under both incentive schemes. A further decomposition by reported levels of household wealth reveals that these results are mainly driven by the richer households, as seen in Table A.3. The respondents were also asked to predict the performance of the future economy of their commune. There are similar effects in satisfaction with future economic conditions. While these measures are informative, they do not directly measure the citizen's satisfaction of the role of the government and performance of their tasks but rather a combination of dynamic effects.

Table 7: Citizens' Perception of Change

	Group Incentives			Group and Individual Incentives		
	Mean		**Diff**	**Mean**		**Diff**
	Pre	Post		Pre	Post	
Outlook on Ec. Conditions						
Current	0.3867	0.4533	0.0667***	0.440	0.4667	0.0267
Future	0.6287	0.6733	0.0446***	0.6257	0.6531	0.0274
Trust Levels						
Local Government	0.6161	0.7931	0.1770**	0.6207	0.700	0.0793
Government Management of Tasks	0.5539	0.5444	-0.0095	0.5756	0.5963	0.0207
Corruption Perception	0.6400	0.5071	-0.1329*	0.5692	0.5037	-0.0655
Satisfaction Scores						
Mayor	0.5533	0.510	-0.0433	0.4733	0.630	0.1567*
Local Council	0.5333	0.4467	-0.0867	0.4433	0.5759	0.1325*
Staff Only	0.4633	0.5448	0.0815	0.4767	0.670	0.1933**
All Workers	0.5167	0.5000	-0.0167	0.4644	0.6264	0.1620**
N			120			120

*$p < 0.1$; ** $p < 0.05$, *** $p < 0.01$

This can be explained by the change in the citizens' trust in the local government. As seen in the second part of Table 8, there is a significant increase in trust in the local government and a decrease in perceived corruption. However, we do not see any significant results in trust of the government's management of local issues, although there is a decrease in the score attributed to this. While there is no significant result under both incentive schemes, the results all point to the expected directions. More convincingly, the commune-level physical execution rate is strongly correlated with citizen satisfaction of the commune government. The average satisfaction score awarded to different actors of the local government are all significantly increasing under both individual and group incentives. Citizens were asked to rate separately government officials of different rankings before and after six months of the intervention. We expect the incentive schemes to increase government productivity and its delivery of public goods, which would in turn increase satisfaction amongst the beneficiary population. Table 8 reports the first differences to capture the change in satisfaction. The data indicates that citizens in communes under both incentive schemes are consistently and significantly happier with the government. For example, there is a significant increase in satisfaction scores with all workers of 16.2 percentage points.

Overall, we see that (i) the measures of productivity at the service level correlate as expected with other measures that presumably affect productivity, and moreover, the audits conducted confirm that the self assessment of the completion of public works is mostly accurate; and (ii) both of the treatments increased productivity, reduced free riding, increased trust and job satisfaction. Even though these results are only indicative, provided he small sample size and lack of the control group, they point towards the potential gains that could be achieved by an implementation of the project at scale.

7 Scale-Up Potential

In conclusion, this project was motivated by a series of focus groups that the PIs organized in Benin in 2015 to try to understand the key constraints facing local governments for improving the quality of public service provision. Three key messages emerged from these conversations: (i) there is a lack of high powered incentives in bureaucracies, which led to low levels of motivation and high turnover, (ii) work environments are not collaborative and conducive to the generation of creative solutions, and (iii) free riding by political appointees hampers the work in the commune. Our experiment was designed to explicitly target these constraints, therefore addressing a long-standing demand of local government administra-

tors. The implementation of the pilot project in 6 communes showed that the proposed design reached wide support and enthusiasm from both workers in local governments and administrators at the local and central level. The fact that we will be conducting the experiment in the universe of communes in Benin, and that we are building the experiment on top of performance measurements that are currently in place makes it easy for the government to continue with the project, has shown positive effects and seems to be overall successful. Finally, the problems that we are addressing with this project are neither unique to local governments nor to Benin: low levels of motivation and poor public service provision are commonplace in the developing world. Development plans are used for programming purposes in local governments around the world. Therefore, the results from this experiment will provide empirical evidence to contribute to the enhancement of state effectiveness by providing a better understanding of the organization of the public sector and bureaucrat?s incentives.

For the scale-up, randomization will take place at the commune level, and given the small number of communes and the low power, we decided to only use one treatment arm rather than two. The treatment chosen is the group incentives, since theory allows sharper predictions of its effects, and better identification of the mechanisms. Given the bounded number of communes/clusters that we can work with (the universe=77), we impose this as a parameter in our power calculation. A clustered randomized evaluation with 50 workers in each commune delivers a minimum detectable effect of roughly 15 percent of the sample mean of our performance measure (with 5 percent significance, power 80 percent, 2 survey rounds, Rho=0.6, ICC=0.292). This effect is fairly small when compared with existing evaluation studies on pay for performance schemes in the public sector. Also, when we consider other outcomes at the worker level, such as the weekly hours worked (ICC=0.017), the level of satisfaction of working in the commune (ICC=0.011) and an index of collaboration with other workers within the same service (ICC=0.004), minimum detectable effects shrink to less than 5 percent of the mean. The same holds for a set of final outcomes taken from the citizens? satisfaction indices on a variety of services and activities carried out by the commune government.[9]

[9]In order to maximize the statistical power of our experiment, we plan to randomly assign the treatments or control status within groups (or strata) of communes as defined by a set of discrete covariates (e.g. geographic region, population density of the served community, total performance score in 2017, etc.). In our setting, a pairwise matched design seems to be the optimal design choice. In such a case, strata of 2 communes are constituted according to the above characteristics, and in each stratum, one unit is randomly assigned to the tournament treatment and one unit is randomly assigned to the control group. We plan to account for the relatively small number of communes in each treatment arm (37-38) using three strategies. First, we calculate statistical significance relative to the small sample t-distribution while clustering standard

errors at the commune level. We then intend to double-check the estimated p-values for the treatment effects using two econometric approaches. We will bootstrap the distribution of the test statistics using the wild-cluster bootstrap [Cameron et al., 2008]. Alternatively, we will also use randomization inference [Angrist and Imbens, 2002].

References

Angrist, J. and Imbens, G. [2002], '[covariance adjustment in randomized experiments and observational studies]: Comment', *Statistical Science* **17**(3), 304–307.

Bandiera, O., Barankay, I. and Rasul, I. [2013], 'Team incentives: Evidence from a firm level experiment', *Journal of the European Economic Association* **11**(5), 1079–1114.

Cameron, A. C., Gelbach, J. B. and Miller, D. L. [2008], 'Bootstrap-Based Improvements for Inference with Clustered Errors', *The Review of Economics and Statistics* **90**(3), 414–427.

Friebel, G., Friebel, G., Heinz, M. and Zubanov, N. [2017], 'Team incentives and performance : Evidence from a retail chain', *The American Economic Review* **107**(8), 2168–2203.

Gertler, P. J. and Vermeersch, C. [2013], Using Performance Incentives to Improve Medical Care Productivity and Health Outcomes.

Holmstrom, B. [1982], 'Moral Hazard in Teams', *The Bell Journal of Economics* **13**(2), 324.

Kazianga, H., de Walque, D. and Alderman, H. [2014], 'School feeding programs, intra-household allocation and the nutrition of siblings: Evidence from a randomized trial in rural Burkina Faso', *Journal of Development Economics* **106**, 15–34.

Lazear, E. P. [2000], 'The Power of Incentives', *The American Economic Review* **90**(2), 410–414.

Miller, G. and Babiarz, K. S. [2013], Pay-for-Performance Incentives in Low- and Middle-Income Country Health Programs, Technical report, National Bureau of Economic Research, Cambridge, MA.

Strengthening State Capabilities: The Role of Financial Incentives in the Call to Public Service [2013], *The Quarterly Journal of Economics* **128**(3), 1169.

Appendices

A Additional Tables and Figures

Figure A.1: Example of a Poster

ECOLE AFRICAINE D'ECONOMIE | IREEP Institut de Recherche Empirique en Economie Politique

L'AMELIORATION DE LA PERFORMANCE DES TRAVAILLEURS A TRAVERS LES INCITATIONS MONETAIRES ET NON MONETAIRES

OBJECTIF DU PROJET

L'Institut de Recherche Empirique en Economie Politique (IREEP) en collaboration avec certaines Mairies a initié un projet dont l'objectif général est d'améliorer votre performance (Améliorer la performance des travailleurs dans la fourniture, la gestion et l'organisation des services publics). De façon plus spécifique, il s'agira de mettre en place les incitations monétaires et non monétaires pour améliorer la productivité des travailleurs qui exercent des tâches aboutissant à des résultats multiples au sein de l'administration communale.

PERFORMANCE DES TRAVAILLEURS

SERVICE PERFORMANT

Les différents services de la mairie seront évalués sur la base du Plan de Travail Annuel (PTA) ou du Plan Annuel de Développement (PAD). Ainsi, l'IREEP en étroite collaboration avec la mairie dégagera le meilleur service après évaluation à mi-parcours. L'évaluation du PTA ou du PAD se fera sur la base des taux d'exécutions physiques des différentes activités.

MEILLEUR AGENT

En plus de dégager le meilleur service, l'IREEP en collaboration avec la mairie compte dégager le meilleur agent.

Les travailleurs qui interagissent le plus souvent ensemble vont s'évaluer sur la base d'un questionnaire qui prend en compte certains aspects à savoir le leadership du travailleur, son professionnalisme et ses compétences.

PERIODE D'EVALUATION : DATE DE DEBUT : Ce jour

PERIODE D'EVALUATION : DATE DE FIN : 30 JUIN 2017

RECOMPENSES QUI SERONT ATTRIBUEES AU MEILLEUR SERVICE

Tous les agents travaillant dans le meilleur service recevront une récompense monétaire : Le treizième du Mois.

En plus de la récompense monétaire, les agents travaillant dans le meilleur service vont recevoir publiquement les attestations de bonne performance.

RECOMPENSES QUI SERONT ATTRIBUEES AU MEILLEUR AGENT

Le meilleur travailleur recevra une récompense monétaire : Le treizième du Mois.

En plus de la récompense monétaire, le meilleur travailleur recevra publiquement une attestation de bonne performance.

Cas pratique de l'évaluation de la performance des services

SERVICE		POIDS (a)	POIDS REEL (a)	TEP (a)	POINT DU SERVICE SUR L'ACTIVITE APRES EVALUATION
SERVICE DEVELOPPEMENT LOCAL ET PLANIFICATION (SDLP)	**ACTION 1 : Projet d'Appui au Développement des Filières Agricoles et Halieutiques (PADFAH)**	$\frac{100}{100}$			
	Activité 1 : Construire un magasin de stockage des produits agricoles	$\frac{\mathbf{33}}{\mathbf{100}}$	$=\frac{33*100}{10000}=\mathbf{0.33}$	$\frac{\mathbf{80}}{\mathbf{100}}$	$=\frac{80*0.33}{100}=\mathbf{0.264}$
	Faire le suivi des activités et Réceptionner les rapports d'activité	$\frac{42}{100}$		$\frac{42}{100}$	
	Réceptionner les rapports d'activité du contrôleur de chantier	$\frac{38}{100}$		$\frac{38}{100}$	
	Produire le rapport d'exécution du projet	$\frac{20}{100}$		00	
	Activité 2 : Construire deux magasins de stockage des produits agricoles	$\frac{67}{100}$	$=\frac{67*100}{10000}=\mathbf{0.67}$	$\frac{80}{100}$	$=\frac{80*0.67}{100}=\mathbf{0.536}$
	Faire le suivi des activités et Réceptionner les rapports d'activité	$\frac{42}{100}$		$\frac{42}{100}$	
	Réceptionner les rapports d'activité du contrôleur de chantier	$\frac{38}{100}$		$\frac{38}{100}$	
	Produire le rapport d'exécution du projet	$\frac{20}{100}$		00	
	Score final du SDPL après évaluation		$=\frac{0.264+0.536}{0.33+0.67}=\mathbf{0.80}$		
SERVICE		POIDS (a)	POIDS REEL (a)	TEP (a)	POINT DU SERVICE SUR L'ACTIVITE APRES EVALUATION
SERVICE AFFAIRES GENERALES (SAG)	**ACTION 2 : Projet d'Amélioration de la Santé des444 Populations (PASP)**	$\frac{100}{100}$			
	Activité : Plaider auprès du ministère de la santé pour l'équipement des centres de santé existants et le recrutement du personnel	$\frac{\mathbf{100}}{\mathbf{100}}$	$=\frac{100*100}{10000}=\mathbf{1}$	$\frac{\mathbf{55}}{\mathbf{100}}$	$=\frac{55*1}{100}=\mathbf{0.55}$
	Identifier les centres de santé et le nombre de poste à pourvoir	$\frac{55}{100}$		$\frac{55}{100}$	
	Transmettre au SG pour validation	$\frac{10}{100}$		00	
	Transmettre au maire pour signature	$\frac{10}{100}$		00	
	Envoyer le courrier au ministère	$\frac{25}{100}$		00	
	Score final du SAG après évaluation		$=\frac{0.55}{1}=\mathbf{0.55}$		

Taux Exécution Physique (TEP) activité= Somme des poids des tâches exécutés sous cette activité.

Poids réel d'une activité=Poids de l'activité multiplié par le poids de son action

Point que recevra le service sur une activité donné après évaluation=Poids réel de l'activité multiplié par le Taux d'Exécution Physique de cette activité

Score final du service après évaluation= $\frac{1}{N}\sum_{i=1}^{n} Poids\ réel\ de\ l'activité * TEP$ (a) avec N la somme des poids réels des activités pour un service donné.

Table A.1: TEP measure by Commune

	Mean		**Diff**	***p*-value**
	Pre	Post		
Group Incentives				
Toucountouna	0.3041	0.4884	0.1843***	0.0064
Bembereke	0.2450	0.4651	0.2201***	0.0016
Zagnanado	0.4114	0.3429	-0.0685	0.2573955
Individual and Group Incentives				
Allada	0.7443	0.56141	-0.1829***	1.29e-08
Grand-Popo	0.4732	0.3793	-0.0939**	0.0190
Pobe	0.5968	0.6500	0.0532	0.4060662

Table A.2: TEP measure by Service

	Mean		**Diff**	***p*-value**
	Pre	Post		
SEC	0.2445	0.4820	0.2375***	0.0024
SAG	0.7141	0.5989	-0.1152**	0.0213
SAF	0.4997	0.4961	-0.0036	0.9654
SAD	0.5178	0.4286	-0.0892	0.2130
SAE	0.6272	0.5567	-0.0705	0.3339
SDLP	0.2394	0.5297	0.2903***	0.0010
ST	0.4614	0.3531	-0.1083**	0.0346
SEEHA	0.284	0.546	0.262*	0.0738
SIAC	0.1097	0.3886	0.2789	0.1434
SGA	0.3531	0.4632	0.1100	0.2470

Table A.3: Current v. Future Optimism for by Wealth Level and Treatment

	Group Incentives			Group and Individual Incentives		
	Mean		**Diff**	**Mean**		**Diff**
	Pre	Post		Pre	Post	
Low or No Poverty						
current_sat	3.822	4.625	0.803**	4.267	4.909	0.642**
future_sat	6.410	6.750	0.340	5.680	6.633	0.953***
Moderate Poverty						
current_sat	3.886	4.700	0.814**	4.450	4.750	0.300
future_sat	6.087	6.800	0.713***	6.764	6.613	-0.151
High Poverty						
current_sat	3.750	4.556	0.806**	4.600	4.467	-0.133
future_sat	6.354	6.689	0.335*	5.962	6.435	0.473
Extreme Poverty						
current_sat	4.067	4.080	0.013	4.343	4.240	-0.103
future_sat	6.245	6.680	0.435*	6.673	6.292	-0.381

B Individual Performance: 360 Peers' Assessments

ECOLE AFRICAINE D'ECONOMIE | IREEP Institut de Recherche Empirique en Economie Politique

IGC International Growth Centre

International Growth Centre
London School of Economics and Political Science
Houghton Street | London WC2 2AE
United Kingdom

FICHE D'EVALUATION

Dans le souci d'améliorer la performance de l'administration communale, l'Institut de Recherche Empirique en Economie Politique (IREEP) en Partenariat avec les différentes mairies dont la vôtre a initié un projet dont l'objectif général est d'améliorer la performance des travailleurs. A cet effet, nous avions mis en place un questionnaire qui permettra aux agents d'un même service et/ou les agents qui interagissent le plus souvent de s'évaluer ce qui pourra nous permettre de dégager à la fin de l'évaluation le meilleur agent qui aura une récompense. **Selon les lois qui s'appliquent, vos réponses seront strictement confidentielles**

Le sommet de l'échelle représente le sentiment le plus élevé de satisfaction tandis que le bas de l'échelle représente le sentiment le plus bas de satisfaction. Les barreaux de l'échelle correspondent aux ci-dessous, où le chiffre 10 est associé au sentiment de satisfaction le plus élevé et le chiffre 1 est associé au sentiment de satisfaction le plus faible.

Nom du candidat : AKOVOWANOU Virgile (Régisseur des dépenses)

Titre : Chef division budget au Service Affaire Financière

Définir votre relation avec le candidat : Bonne

Combien de fois vous interagissez avec le candidat ? Parfois

A. Leadership (25/100)

	Tout à fait en désaccord		En désaccord			D'accord			Tout à fait d'accord	
	1	2	3	4	5	6	7	8	9	10
Reconnait et encourage les idées des autres collègues						6				
Encourage les collègues lorsqu'ils ont de bon résultats						6				

	Très mauvaise aptitude		Mauvaise aptitude			Bonne aptitude			Très bonne aptitude	
	1	2	3	4	5	6	7	8	9	10
Aptitude à l'encadrement				4						

B. Professionnalisme (35/100)

	Très bonne implication		Bonne implication			Mauvaise implication			Très mauvaise implication	
	1	2	3	4	5	6	7	8	9	10

Implique les collègues dans les discussions/ ou aime se faire impliquer en ce qui concerne la réalisation des objectifs de l'équipe		5		

	Tout à fait en désaccord		En désaccord			D'accord			Tout à fait d'accord	
	1	2	3	4	5	6	7	8	9	10
Est réceptif aux suggestions sur l'amélioration de la productivité de l'équipe							7			
Fait des suggestions sur l'amélioration de la qualité et de la productivité du travail				5						
Accomplir les objectifs à temps				5						
Vient au service à l'heure				5						
Est-ce que les compétences du candidat constituent un atout précieux pour le service							8			
Examine attentivement les travaux réalisés avant de les soumettre/travaille avec beaucoup de soin				4						

	Jamais		Parfois			Souvent			Toujours	
	10	9	8	7	6	5	4	3	2	1
Nécessite un minimum de supervision lorsqu'il travaille				6						

	Tout à fait en désaccord		En désaccord			D'accord			Tout à fait d'accord	
	1	2	3	4	5	6	7	8	9	10
Traite les usagers de la mairie avec respect, équité et intégrité							8			

C. Compétence (40/100)

Explication des codes d'évaluation I : Insuffisant (n'atteint pas les éléments requis) II : Modéré (n'atteint pas tous les éléments requis)											III : Suffisant (correspond aux éléments requis) IV : Bon (excède les éléments requis) V : Excellent (excède considérablement les éléments requis)
Responsabilités	I		II		III		IV		V		
	1	2	3	4	5	6	7	8	9	10	Courte explication ou motivation

Responsabilité 1 : Inventaires des actifs et passifs							7				
Responsabilité 2 : Payement des agents								8			
Responsabilité 3 Production des pièces justificatives							7				

Nom de l'évaluateur : AVOHOU Guillaume (Chef du Service Affaire Financière)

The International Growth Centre (IGC) aims to promote sustainable growth in developing countries by providing demand-led policy advice based on frontier research.

Find out more about
our work on our website
www.theigc.org

For media or communications
enquiries, please contact
mail@theigc.org

Subscribe to our newsletter
and topic updates
www.theigc.org/newsletter

Follow us on Twitter
@the_igc

Contact us
International Growth Centre,
London School of Economic
and Political Science,
Houghton Street,
London WC2A 2AE

IGC
International
Growth Centre

DIRECTED BY
LSE
UNIVERSITY OF OXFORD

FUNDED BY
UKaid
from the British people

Designed by soapbox.co.uk