

Bureaucrat Time-Use and Productivity: Evidence from a Survey Experiment

Jozefina Kalaj, Daniel Rogger, and Ravi Somani*

March 2020

Abstract

Bureaucratic effectiveness is an important input into state capacity. The tasks public officials choose to spend their time on determines how their human capital impacts national development. Yet empirical evidence on how to effectively measure public officials' time use, what determines their allocation decisions, and how this feeds into their productivity is scarce. We contribute on all three of these margins through a survey experiment with Ethiopian bureaucrats. We randomly test alternative measures of bureaucratic time use by varying recall period, enumeration methodology and the degree of task detail in recall surveys. Benchmarking these modes to time use diaries, we identify the relative inaccuracy of requesting task detail and the survey time and data entry costs of using graphical methods. Measuring time use in the public administration precisely rests on the resolution of a tension between the relatively high level of education of public officials and the homogeneity, but varying intensity, of their tasks. We then describe the nature of time use of public officials across Ethiopia's government, and show correlational evidence that the structure of time use matters for service delivery outcomes.

*Kalaj: Pepperdine University, jozefina.kalaj@pepperdine.edu; Rogger: World Bank, drogger@worldbank.org; Somani: World Bank, rsomani@worldbank.org. We thank the Ministry of Public Service and Human Resource Development of the Federal Government of Ethiopia for their leadership and support throughout design, implementation and data collection. We are very grateful to Elsa Araya, Verena Fritz and Kerenssa Kay of the World Bank's Governance Global Practice for collaboration on implementation of the survey that underlines this work. We also thank Margarida Soares Rodrigues for excellent research assistance during this project. The study received ethics approval from Oxford University Blavatnik School of Government's Departmental Research Ethics Committee (protocol SSD/CUREC1A/BSG C1A 16-010). We gratefully acknowledge financial support from the Centre for the Study of African Economies [grant number CUR00090], the UK's Department for International Development, the International Growth Centre [grant number 32407], and the World Bank's Development Impact Evaluation i2i fund, Bureaucracy Lab and Governance Global Practice.

1 Introduction

The nature of public administration implies that official’s time use is fundamental in the production of public services. Public administrations are characterized as having multiple tasks, multiple principals and a frequent need to coordinate their activities with other public agents and agencies (Dixit, 2002). The distribution of their time across these activities determines which tasks are completed most successfully, and thus is a key input in the delivery of government services.

While there have been recent developments in the measurement of individual bureaucrat performance, there has been very little micro-level empirical evidence on how bureaucrats allocate their time across tasks within the bureaucracy (Finan et al., 2017). This paper contributes to this gap by developing, and experimentally testing, different measures of bureaucrat time use in the field. We study the correlates of these measures of time use to provide the first evidence on the nature of time use in the public administration and its impacts on service delivery.

Specifically, we work closely with the Government of Ethiopia to develop four different survey modules of bureaucrat time use based on standard approaches to measuring time use enumerated in other settings. We enumerate them across a representative sample of 1,776 civil servants working in 380 organizations of the three tiers of the Ethiopian Government. The time use measures are embedded in a broader survey that aims to capture officials’ characteristics and the incentive environment in which they operate, based on the World Management Survey adapted to the public-sector setting (Bloom and Van Reenen, 2007; Rasul and Rogger, 2018). The exercise is one of the most comprehensive surveys of public officials ever undertaken. We simultaneously collect detailed time-use diaries for a subset of these civil servants in the capital city as a benchmark measure of time use.

The first part of this paper empirically tests which survey measure is closest to the data in the time-use diaries, based on the sample of civil servants for whom we have both measures of time use. We therefore present the first evidence we know of on the comparative efficacy of methods for collecting time use data in the public administration. These modules vary in the time recall period (week or month); the length of the list of activities to distribute time across; and the use of a graphical or text-based module.

We find significant evidence that the length of the activities listed in the survey module influences estimates of time use. All three of the survey modules that limit the set of activities to five perform equally well in terms of accuracy relative to the time-use diary. The module that asks respondents to fill in their time use across a long list of activities (fourteen) performs significantly worse. Nevertheless, the best-performing survey modules differ from the time-use diaries by roughly a third of working time, on average, across all of the activities. The module with a long list of activities performs twice as poorly, relative to the diary. These findings are consistent with the notion that the homogeneity of tasks in public administration distort respondent recollection of their activities towards the most salient. These findings are consistent with the

existing literature, which indicates that stylized measures of time use perform well in the case of scheduled and structured activities but not in other cases (Bianchi et al., 2000; Juster et al., 2003; Kan and Pudney, 2008).

We use the most accurate of our measures to investigate the correlations between time use and individual characteristics of the bureaucrat and the incentive environment in which they operate. Managers spend significantly more time in meetings and central government officials spend significantly more time working alone. Amongst other characteristics, we measure the public-service motivation and locus of control of civil servants. We find that those exhibiting higher public-service motivation spend less time in meetings and without work and those with higher levels of internal locus of control spend significantly more time working alone. Their influence on time use may be one factor through which the characteristics of public official's impacts the quality of service delivery.

In terms of the incentive environment, we find that those operating under multiple principals spend significantly more time in meetings and traveling, and less time interacting with clients. Those operating under better management practices, as measured by the World Management Survey, spend less time interfacing with clients. However, we find that the nature of management practice an official works under matters for the distribution of their time. Our aggregate result on management is driven by those civil servants operating under management practices with strong monitoring, targeting and incentives. To our knowledge, this is the first breakdown of bureaucrat time use by the management conditions under which they operate, providing a basis for further micro-level research to understand the mechanisms through which management practices filter through to bureaucratic productivity.

We then combine our survey data with government monitoring data on organizational performance. We take the latest service delivery data from administrative and monitoring sources for each of the sectors in which our respondents operate (agriculture, education, health, revenue, and trade). Guided by a simple production function, we empirically explore how the allocation of time across the activities correlates with service delivery outcomes. We find that civil servants that spend more time on coordination activities (meetings and interfacing with clients) work in organizations with significantly better service delivery outcomes. An additional hour spent on coordination activities is associated with a 0.007 standard deviation increase in service delivery. We find that time spent working alone is not significantly correlated with service delivery. To our knowledge, this is the first empirical evidence that links bureaucrat time-use to service delivery outcomes across a multi-sector representative sample of the bureaucracy.

This paper contributes to two main strands of literature. Existing studies have focused on the allocation of time between market and non-market activities (Becker, 1965; Juster et al., 2003; Bonke, 2005; Otterbach and Sousa-Poza, 2010) and distribution of time over domestic activities (Bianchi et al., 2000; Kan and Pudney, 2008; Schulz and Grunow, 2012). Yet, the allocation of time while at market work has received little attention (Hurst, 2015). To our knowledge, this is the first large-scale survey experiment in the workplace and the first

large-scale survey of workplace time use across multiple sectors within the public sector. Our analysis is also related to recent studies focusing on the relationship between the time use of leaders and organizational outcomes. Bandiera et al. (2012) show a positive association between a firm’s productivity and the time the CEO allocates to planned meetings with firm employees, with unplanned time impacting the firm negatively. In the public sector context, Dasgupta and Kapur (2017) suggest that service delivery depends on the officials’ ability to dedicate more time towards planning and managerial activities and less on handling micro-transactions with constituents.

This paper also contributes to the growing literature on bureaucrat productivity (see Finan et al. (2017) for an overview). While there have been recent advances in the measurement of bureaucrat productivity, there is still little empirical evidence on the individual activities and tasks that bureaucrats spend their time on.¹ This paper provides the first representative picture of bureaucrat time-use across activities in any setting, but with attention on a development setting where bureaucratic functioning is both varied and inefficient (Lewis-Faupel et al., 2016; Rasul and Rogger, 2018). As the literature on public administrators grows, it becomes increasingly important to have validated measures of key characteristics of bureaucratic functioning like that presented here.

The rest of this paper proceeds as follows: Section 2 details the data, our measures of time use and the experiment; Section 3 investigates the correlates of bureaucrat time use, with a focus on individual characteristics and management practices. Section 4 explores the correlation between time-use and service-delivery outcomes. Section 5 concludes.

2 Data and Measurement: A Survey Experiment

2.1 Approaches to Measuring Time Use

Time-use data has typically been collected through two main methods: using diaries to track activities in close to real-time, and using questioning in surveys.² In the case of diaries, the respondent is left with the survey form and asked to document activities as they take place over the course of the next day or specified time period. Surveys either request the respondent detail their recent activities

¹For example, researchers have investigated the productivity of public officials using project-level data (Rasul and Rogger, 2018; Rasul et al., 2019), service-delivery data (Ashraf et al., 2018; Gulzar and Pasquale, 2017); sales figures (Ashraf, Bandiera and Jack, 2014); attendance (Callen et al., 2018); tax collection figures (Khan et al., 2019); subjective assessments of bureaucrat performance by civil servants, politicians, media professionals, members of business and professional associations, and civil society and think-tanks (Bertrand et al., 2018).

²A third method for measuring time use, experiential sampling, is also noted in the literature. Paging devices, such as cellphones, are used to alert the respondent to report the specific activity that is underway. Experiential sampling and diary methods share similarities and evidence suggests they generate similar estimates of time use. For more information on this method see Juster et al. (2003).

for recording by the enumerator or present the respondent with a specific list of stylized activities that they are asked to recall the amount of time devoted to over a particular period – generally a week, month, or year.

Surveys have the advantage of being simpler and cheaper to enumerate, and are thus the most common source of time-use data (Seymour et al., 2017). However, their reliability varies across settings and rigorous evaluations of different approaches to measuring time use are scarce. This is particularly true in developing countries, despite time use data collection growing considerably in recent years. This paper provides a first assessment of the efficacy of survey-based time use measures in a public administration setting.

Diaries provide a benchmark against which survey efforts can be judged (Marini and Shelton, 1993). Making use of technology, recent studies have corroborated the limited bias for diary estimates. For instance, Kelly et al. (2015) used images gathered from cameras worn by respondents to test the validity of diaries. Results show no significant difference between the diary and camera data in terms of average time spent on ten separately coded daily activities. The intention of any survey-based method is therefore to replicate the results of a diary-based method. It is important to know when stylized measures yield results comparable to diaries for specific settings (Juster et al., 2003).

Studies comparing the two methods indicate that the nature of the underlying task, respondent characteristics and contextual factors are significant predictors of deviations between diaries and stylized measures. Activities which follow a repeated schedule, such as total market work hours and specific tasks performed regularly while on the job, have a lower risk of recall bias (Juster et al., 2003). Studies of the time dedicated to market work imply that respondents are accurately able to recall total work hours (Bonke, 2005; Otterback and Sousa-Poza, 2010). Average errors in estimates of market work time are between two and four percent of total time reported in diaries.³ This finding is partially mediated by workload, with stylized instruments yielding 28% higher estimates than diaries for those reporting 56-60 weekly work hours (Otterback and Sousa-Poza, 2010). However, as will be seen such high, and potentially more variable, workloads are rare in the public service. What has not been studied is the extent to which respondents are able to accurately report their distribution of activities within the workplace. This paper presents the first evidence on the accuracy of that reporting and how it is mediated by the nature of measurement.

If tasks vary in intensity across time, such as an official having many meetings one week but few the next, respondents will tend to bias their estimates. An example of this phenomenon from a study that is closely related to the current one in approach is Beegle et al. (2012). The authors randomly assign eight alternative consumption questionnaires, including three diary designs, to house-

³Average error is defined as relative deviation from diary reported hours. For example, Otterback and Sousa-Poza (2010) estimate 38 and 37 work hours according to stylized and diary instruments, respectively; an overestimation of 2.7%. Niemi (1993) finds no significant differences between stylized and diary data for salary earners.

holds in Tanzania. For a given recall period, they find that stylized consumption estimates are lower than diary ones despite a diet consisting of a relatively limited range of items. Similar results arise from studying housework. Using data from two U.S. surveys, each utilizing distinct time use methods, Bianchi et al. (2000) find that stylized estimates of weekly household work are about 50% higher than the diary reported hours. Other research implies errors in stylized estimates of time undertaking housework of around 35% (Bonke, 2005; Kitterod and Lyngstad, 2005; Schulz and Grunow, 2012).

Measurement error is particularly large for poorly educated respondents, who find it more difficult to accurately recall their activities (Beegle et al., 2012). The quality of education of officials varies substantially across public service in the developing world, with schooling quality varying across regions and districts. Due to rapidly changing education systems, this variation may also present itself across ages and layers of hierarchy, highlighting an area of investigation in our later analysis. However, since most public administrators will be screened for basic literacy, such concerns may be of less validity in the public service setting. Gender is unlikely to mediate time use measurement substantially in a public administration setting given the commonality of the work environment and task space across the sexes, though it is an area on which we can provide empirical evidence.

Finally, contextual factors have been shown to play an important part in collecting accurate time use. For example, Beegle et al. (2012) find that respondents residing in urban areas provide more accurate consumption estimates for longer recall periods, whilst Wambile et al. (2016) suggests recall estimates of per capita consumption to be closer to diaries in rural areas. This is linked to these respondents facing different environmental features, such as the practice of shopping at markets providing salient points of reference for consumption. Robinson and Gershuny (2011) find that the nature of work (such as whether individuals undertake legal, educational or security tasks) effects time estimates.

Public officials *de jure* institutional features - working in a hierarchical office environment on administrative tasks - vary little across contexts, but their *de facto* environment - the incentive regimes and management practices they work under - varies significantly across divisions (Rasul and Rogger, 2018; Rasul et al, 2019). The breadth of our data collection on such contextual features allows us to study whether either of these mediate the quality of measurement. Error also arises in settings where activities interact with multiple individuals (Gibson and Kim, 2007), with the discordancy in measurement increasing with the number of members of an activity (such as members of a household). Bureaucratic activity is frequently undertaken with other officials within a division, and often with a sizable number.

Thus, many of the classic characteristics of public administration have been identified in the wider literature as a potential source of recall errors that require investigation. The relatively unique institutional environment of officialdom - common tasks undertaken with varying intensity, across a range of *de facto* institutional environments - has a *a priori* ambiguous impacts on measurement quality. This provides a solid rationale for empirical investigation.

2.2 Surveys of Bureaucrat Time Use

Between June and September 2016 we held face-to-face interviews with 1,776 public administrators in 380 organizations across the three tiers of Ethiopia’s government. Our analysis covers the agriculture, education, health, revenue, and trade sectors. For each of these five sectors, we sample the main ministry, all the corresponding regional offices, and a geographically representative sample of corresponding district offices from across the country. Table A1 lists the organizations included in our study, and Figure A1 provides a map of the district governments we sampled. Within each organization, we interviewed senior managers and a representative sample of their staff. We limit our scope to the professional grades of technical and administrative officers, excluding grades that cover cleaners, drivers, secretaries, etc. Altogether, the survey is one of the most comprehensive enumeration exercises ever undertaken in the public administration.

To develop and enumerate a questionnaire that was relevant for the Ethiopian setting, we worked closely with the Ministry of Public Service and Human Resource Development and employed ex-civil servants within our enumeration teams to facilitate navigation of the public service. The survey captured the demographic and other individual characteristics of public officials; the management practices under which they operate; and measures of time-use. The implementation of the survey was successful across the organizations we visited, with 99.5% of public officials sampled agreeing to be interviewed and 98.2% of interviews being classified by the enumerator as having gone ‘somewhat well’ (26.4%) or ‘very well’ (71.7%).

Table A2 provides descriptives for the 380 organizations and 1,776 individuals that we study. All officials work within a relatively standard structure, with a manager overseeing levels of hierarchy below him within a clearly defined organizational structure. The average age of the officials we study is 35, 80% are male and the majority have a degree (82%). Civil servants, on average, have been in their position for just under 3 years, in their organization for over 7 years, and in the civil service for 13 years. 25% of our sample is at the director level. Civil servants are contracted to work the mandated 40-hour week.

2.2.1 Measuring Bureaucrat Time Use: Time-Use Diaries

During the enumeration period, we provided civil servants in the capital city with time-use diaries to capture their allocation of time across five main activities over the course of a working week (five days). The categories of time-use were developed in collaboration with senior and junior civil servants in the government and included: Meetings within the service; Working alone/on administration; Interfacing with people outside of the service; Traveling for work; and Without work to do. The format of the diary is presented in Appendix: Time-Use Diary. During the following weeks, centrally-based enumerators returned to the respondents to collect the completed diaries.

2.2.2 Measuring Bureaucrat Time Use: Time-Use Surveys

Informed by the existing measurement literature, our time use survey experiment consists of four randomly assigned distinct stylized question modules. We assess the relative accuracy with which different approaches elicit the time use of respondents by benchmarking the stylized survey modules against time use diaries completed by respondents. The dimensions we seek to gain further insight on include length of recall period, level of detail in the activity list, and reporting of time proportions with the help of visuals.

We randomized public officials into one of four possible survey modules of time use. The first approach (Short|Week) echoes the categorisation in the diaries by asking civil servants to state the number of hours that they spend, during a typical week, into the following categories, consistent with the time-use diary: Meetings within the service; Working alone/on administration; Interfacing with people outside of the service; Traveling for work; and, Without work to do/waiting for others to input.⁴ The concern with this approach is that any single week is not representative of the official’s wider time use and induces variance in aggregated estimates (Crossley and Winter, 2015).

The second approach (Short|Month) aims to gain a more representative distribution of their tasks by asking civil servants to state the proportion of time that they spend on the same list of activities in a typical month. The concern with this approach is that memory declines with the length of the recall period (see Sudman et al. (1996) for a review of this issue in the consumption literature). As Crossley and Winter (2015) state, “The situation is complicated by the fact that forgetting does not occur at random but might be differential across respondents and types of questions.”

The third approach (Long|Week) presents a longer list of activities for the respondent to specify how many hours s/he spends on during a typical week. Table A3 in the appendix presents the long list of options. The ambition of this approach is to increase the granularity of detail on the respondent’s time use. The concern, as echoed in the wider literature on time use measurement, is that longer lists of activities lead to double-counting of hours. Respondents find it increasingly difficult to accurately reflect their time use across more disaggregated categories.

The final approach (Graphical|Week) asks civil servants to draw the distribution of their time, across the same categories, during a typical week using a pie graph. Where there is concern that a grid-like design for enumeration is inappropriate, some researchers have proposed the use of graphical methods of identifying time use (Masuda et al., 2014). Though officials may be amongst the most likely in a society to interact with tables of categories, their ability to

⁴The options, ‘Interfacing with people outside of the civil service (e.g. clients)’ and ‘Interacting with frontline workers’, are aggregated to an option, ‘Interfacing with people outside of the service’, throughout the analysis. There is a related literature on whether such questions should target ‘regular behaviors’ or those ‘from the last week’. Chang and Krosnik (2003), for example, study news media consumption and find that ‘typical week’ questions perform better than ‘last week’ questions in that context. However, they also conclude that more systematic research is needed on how to frame the conception of the reference period.

interact with different modes of data collection is an empirical question worthy of investigation here.

To enable direct comparison across our stylized modules, we group and aggregate the long list activities to match those on the short list. For example, hours spent in meetings with direct supervisor, meetings with case team members and mentors (under a scheme named the ‘Change Army’), organization level meetings, awareness raising meetings on policies and strategies, and other meetings, are added together to correspond to ‘Meetings within the service’ from the short list of activities. Table A3 presents the aggregation of long list options. Next, we convert hours to proportion of time for each activity on the short list.

Table A4 provides evidence of the successful implementation of the survey experiment. The table shows that recipients of the different survey modules are balanced across individual characteristics (Table A4A); organizational characteristics (Table A4B); and survey characteristics (Table A4C).⁵ Across each of the comparisons between experiment groups, the joint F-test across all individual, organizational, and survey characteristics is insignificant at the 10-percent level. Table A5 provides evidence towards the same conclusion when restricting the sample to those only in the capital city, Addis Ababa, where we were able to distribute and collect the time-use diaries. Hence, for all survey respondents and for the smaller set of respondents for whom we have both survey and benchmark diary data, the experiment was implemented with success.

2.3 Descriptive Evidence on Bureaucrat Time Use

Before examining how well each of the survey modules compare to the time-use diary, we provide an aggregate picture of bureaucrat time use in Figure 1. Figure 1A compares the distributions in total working hours between our national survey, the set of full-time public-sector workers working in public administration based in Addis Ababa and the set of full-time private-sector workers based in Addis Ababa from the 2013 Labor Force Survey (LFS). We restrict the sample of the LFS to Addis Ababa to focus our analysis on public administrators rather than frontline workers, since the LFS does not explicitly differentiate between different types of public employment. The figure shows that the distribution of total working hours stated by public-sector workers in the Ethiopian Civil Servants Survey and the set of public administrators in the

⁵Individual characteristics include: age, an indicator for whether the respondent is male, an indicator for the highest education qualification of the respondent; years in the current position, years in the current organization, years in the civil service; an indicator for the sector of work. Work environment characteristics include the organizational access to electricity, phone networks, internet, computers, vehicles, and skilled staff, and management practices (both aggregate and individual topics) as measured by the World Management Survey. Survey environment characteristics include: an indicator for whether the interview was conducted in complete privacy; indicators for whether the respondent was knowledgeable about their own environment; an indicator for whether the respondent was willing to reveal basic and confidential information; an indicator for whether the respondent seemed very patient; an indicator for whether the interview was perceived to go very well by the enumerator.

Labor Force Survey follow similar patterns, corroborating the two approaches. There is a large mass around the 40-hour-per-week mark, suggesting that most public administrators work exactly the mandated number of hours per week. The median number of hours in both data-sets is 40 and the means are 41 and 43 hours respectively. As might be expected, the distribution of working hours for full-time employees in the private sector exhibits much more variation. The median is 48 and the mean 52.

The LFS public sector data seems to miss a set of workers who do not work their mandated hours, with the only deviation between the two surveys being the proportion of public sector workers who work between 25 and 40 hours a week. Granular time use data, like that collected by our survey, provides the opportunity to better understand the realities of the public sector labor market, with officials in this bracket making up approximately 31% of all public administrators in our sample. Together with our wider survey data, we can assess the characteristics of individuals in this ‘bump’. They are older than their colleagues, work in organizations with lower levels of performance management, and have a greater number of principals that they report to. These characteristics are often correlated with low effort in the contracting literature (Lucas, 1978; Baker et al., 1988; Holmström and Milgrom, 1988; Dixit, 2002).

Figure 1B shows the distribution of civil servant working hours across the five major categories used in this paper. Averaged across all survey options and respondents, the figure suggests that bureaucrats spend almost 40% of their time working alone, the most common category. After this, bureaucrats spend a quarter of their time in meetings, 15% of their time interfacing with clients, 15% of their time traveling, and 5% of their time without work. The distribution of time use is very similar across employees and senior managers in our setting. To our knowledge, this is the first representative picture of how bureaucrats allocate their time across such categories. The best available comparison for the private sector is with senior managers in the private sector as reported in Bandiera et al. (2017). In their data, senior managers spend 70% of their time interacting with others and the rest of the time working alone (25%) or traveling (5%). To the extent that comparisons can be made amongst these studies, private sector managers are far more likely to be in meetings and far less likely to work alone, in contrast to prevailing stereotypes of the two sectors.

We find no substantial differences between how public officials in different sectors use their time. The aggregate distributions of time use are almost identical across the five sectors we assess, implying that the structures of public administration are more of a determining factor in how public officials use their time than the specific sector in which they work.

2.4 Experimental Assessments of Time-Use Surveys

In Figures 2 and 3 and Table 1 we show the results of our experiment, comparing the different survey modules to the time-use diary. Figure 2 displays raw differences between the time use diaries and respondent reports. Figure 3 and Table 1 presents the results of regressions of the following form:

$$y_{i,j} = \alpha + \beta_1 LongWeek_{i,j} + \beta_2 ShortMonth_{i,j} + \beta_3 GraphicalWeek_{i,j} \quad (1)$$

$$+ \delta' X_{i,j} + \gamma' Z_j + \epsilon_{i,j} \quad (2)$$

Where i refers to the individual and j to the organization. $y_{i,j}$ refers to the *absolute* difference between the stated time use and the entry in the time-use diary. The Short|Week option is the omitted category, hence the coefficients β_k are to be interpreted as relative to the Short|Week for $k = 1, 2, 3$. $X_{i,j}$ represent individual characteristics and survey characteristics. Z_j represent work environment characteristics at the organization level.

Figure 2 illustrates that *on average* the different modes of surveying do surprisingly well in terms of their raw differences from the diary. The deviations from zero represent an aggregation of under- and overestimates across individuals. This implies that survey methods perform rather well in estimating general features of time use for the public service, as presented above.

Once we also average across the five main categories we study, the raw differences are zero, implying that officials are providing reports of their time use that add up to their total hours. Given the standardized work hours of public administration, this may be expected in a range of public sector settings. This finding is consistent with results from the private sector and indicates that most people accurately recall the time they are in paid work.

Figure 2 indicates that there is significant variation in the quality of measurement of the different categories of activity we recorded, and in the performance of the Long|Week module. Figure 3 graphically displays that when we take the absolute difference between responses and diaries, the scale of errors is far more significant. Across our survey methods, individuals under- and overestimate the time they spend on particular tasks, though as seen above the raw errors sum to zero. The best performing category, the Short|Month module, still has an absolute error of 55%. This result is some of the first evidence on the quality of reporting on the distribution of activities in the workplace, and is consistent with error rates reported in the wider literature on time use methods.

To investigate this formally, we turn to Table 1. The dependent variable in columns (1) to (4) of Table 1 is the *absolute* difference between the survey response and the time-use diary, aggregated across all categories of time use. Using the absolute difference allows us to focus on how accurately respondents report their distribution of activities whilst at work.⁶ The dependent variable for columns (5) to (9) is the absolute difference between the time-use diary and survey response for each individual category of time use, noted in the column heading. Figure 3A and Column (1) of Table 1 show the results of specification (1) unconditional on any of the controls. Column (2) adds individual controls to the specification; Column (3) adds work environment characteristics; Column (4) adds survey controls. Figure 3B provides a graphical illustration of the re-

⁶We find that a joint test of enumerator fixed effects is significant at the 1% level, with an F-statistic of 6.12. We therefore condition on them in all individual-level specifications.

sults in Column (4). The sample throughout is the set of officials we interviewed in the capital city that are our focus for the diary comparison.

The results show that the Long|Week option performs significantly worse than the other options, leading to an additional 50% difference in the absolute difference relative to the diary, across all categories of time use, significant at the 1-percent level. The longer list of categories seems to reduce a respondent's ability to accurately reflect the distribution of their activities, similar to results in the consumption and expenditure literatures (Pradham, 2001; Beegle et al., 2010). The results presented in columns (5) to (9) suggest that the differences driven by the Long|Week option come from estimates over time spent in meetings and time working alone (the major components of time use, as presented in Figure 1).⁷

Figure 2 indicates that the Long|Week option overestimates the proportion of time spent in meetings and underestimates the time working alone. The differences in these two categories is substantial, significantly skewing the implied distribution of time use towards meetings. In the Long|Week case, bureaucrats are more accurate in estimates of time use when estimating over activities that they perform less regularly, such as interfacing, traveling and without work. These findings are consistent with the existing literature, which highlights the importance of salience to the accuracy of stylized measures of time use (Bianchi et al., 2000; Kan and Pudney, 2008).

Errors from the Short|Month and Graphical|Week options are statistically indistinguishable from the Short|Week, suggesting that the recall period (week versus month) and the format of the survey (graphical versus list) do not generate systematically different responses within the public administration. The coefficients on these modes are insignificant across all columns. In contrast, consumption and expenditure studies have found significant differences between modules with varying recall periods (Scott and Amenuvegbe, 1990; Beegle et al., 2012; Gaddis et al., 2020). The Short|Week does slightly better on accuracy than the Graphical|Week option, with an unconditional p-value rejecting equality of these options of 0.08, rising to 0.10 for specifications with a full set of controls. However, the significance of the difference between these options lessens when we break down the analysis to categories of activity (see Columns 5 to 9 of Table 1), and become insignificant at the usual levels.

The absolute differences represent a significant deviation from the time use diaries. As can be seen from Figure 3A, the absolute difference between the survey response and the time-use diary is 75% across our sample. Excluding the Long|Week option, the absolute difference between the survey response and the time-use diary is still 62%. This is the sum of absolute differences in each of the categories, with the absolute differences amongst categories ranging from 6% to 21%. Since a single unit of incorrectly recorded time has impacts on both the category it was and should have been recorded under, this implies roughly a third of total time is misreported. Thus, within the highly accurate report

⁷Respondents are using the full distribution of task options we provide, with no option counting for zero percent of time use on average.

of total time in paid work, bias persists in public officials report's of their time use across activities. This implies that survey methods perform rather less well in estimating individual patters of time use than for the service as a whole, but the level of errors is in line with those found in other sectors, such as housework (Bianchi et al., 2000; Schulz and Grunow, 2012).

Looking at the control means in Table 1, which shows the difference between the time-use diary and the survey response for the Short|Week category, it is evident that estimates of time spent working alone perform relatively worse than estimates over other categories of time use across the Short|Week, Short|Month and Graphical|Week options. This pattern of errors echoes findings from the wider literature. The errors arise from the nature of public administration, with some aspects of administration more salient than others. The regularity of the hours worked in a week in the public sector implies that respondents statements over their total hours worked is relatively accurate. The overall raw difference is zero, with overestimates in one category being compensated by underestimates in other categories. However, individual official's estimates of their task distribution suffer from the salience of meetings, travel and interacting with citizens/clients.

2.4.1 Impact of Respondent Characteristics

A key concern in investigations of time use measurement is the impact of respondent characteristics on the relative efficacy of measurement modalities. We can explore how the characteristics of bureaucrats correlate with the absolute distance between the survey estimate of time use and the time-use diary. We assess this by studying the coefficients on $X_{i,j}$ in equation 1 adding interactions between $X_{i,j}$ and the survey option in Table A6. The results suggest that there is no significant evidence that basic demographic and hierarchical characteristics are correlated with the distance between the survey response and the diary entry. Civil servants that are older, more educated, and more senior do not systematically provide differential survey responses relative to their time-use diary entries.⁸ Public administrators are all sufficiently above the 'literacy threshold' to effectively engage with the time use survey, but as noted above suffer in a common way from reporting given the nature of their work. We find no evidence in our data that gender determines errors in time use reporting.

Similarly, contextual factors can mediate the effectiveness of modes of measurement, as discussed above. In line with the fact that time use is very similar across sectors, sector fixed effects are not jointly significant in the experiment at the usual levels, with an F-statistic of 1.31 and joint p-value of 0.28. Thus, our modules perform similarly across sectors. As will be described in more detail below, we have measures of the de facto incentive environment of the

⁸Evidence on the impact of education on the accuracy of recall is mixed. Frazis and Stewart (2004) find educated respondents overestimate total work hours relative to the diary. Bonke (2005) finds those more educated are more likely to understate paid work. Kitterod and Lyngstad (2005) and Otterback and Sousa-Poza (2010) find no significant differences between educational groups.

organizations we study. Due to power limitations stemming from the number of organizations we have time use diaries for, we cannot include the full set of organization-level controls in equation (1) if studying management practices. However, a regression without organizational-controls implies that a public official's broad de facto incentive environment does not impact on the quality of their reporting, with a p-value on the corresponding coefficient being 0.76. A joint test of the interactions between management and our module types has an F-statistic of 1.35 and a p-value of 0.26. Finally, we investigate the extent to which officials who report that a greater percentage of their tasks and projects requires engagement with other stakeholders. Once again, we find no evidence that the quality of time use reporting is mediated by individuals who have a greater range of collaborators in their work.

This leaves us with the conclusion that public administrators across the service and under different institutional environments respond in a similar way to the distinct modes of measurement we test. This may be expected given their common familiarity with administrative tasks.

2.4.2 Time, Quality and Cost Considerations

On from accuracy, we assess the extent to which different modules take more time to complete than others. Perhaps due to its novel nature, we find that the Graphical Week mode takes 4 minutes longer on average ($p=0.00$), on a baseline of 13 minutes (the module included a range of questions beyond simply apportioning time into different categories). This is a substantial proportionate increase, and there is not clear evidence that to gain accuracy a graphical approach is required in high literacy environments such as the public administration. However, that it provides a similar quality of response implies that it is a feasible option in environments in which graphics may be more easily engaged with than other methods.⁹

We do not find differences in the time to enumerate the other three options, nor in the impacts of the different modes on the enumerators perception of the quality of the survey. We ask the enumerator to assess whether the respondent was knowledgeable about their own environment, whether the respondent seemed very patient, and broadly whether the interview was perceived to go very well. Along all of these margins, we do not see differential effects of being enumerated one or another of the time-use modules. However, given that 98% of the interviews were assessed by the interviewer as going somewhat or very well, we have limited variation on which to assess these margins.

Finally, differential impacts on survey costs were driven by the need to undertake careful data entry of the graphical module. Since measurement required assessing angles in a pie chart (see Appendix for enumeration of the graphical mode), the work was done by a research assistant rather than a data entry firm, increasing per unit costs.

⁹We also find no differences in the quality of measurement in government offices with differing levels of access to electricity, phone networks, internet, computers and vehicles.

3 Investigating Bureaucrat Time Use

In this section, we take advantage of our rich data, conditional on our findings above, and explore the correlates of bureaucrat time use. We exclude all time use data collected through the Long|Week module and condition all analysis on time use module and enumerator fixed effects. To provide structure to this descriptive exercise we assess characteristics of the bureaucracy typically invoked in a public-sector production function (Delfgaauw and Dur, 2007; Bandiera et al., 2017; Link and van Hasselt, 2019). Specifically, we focus on measures of the quality of human capital and the nature of management.

3.1 Time Use and Individual Characteristics

In this sub-section, we explore the extent to which individual characteristics correlate with bureaucrat time use. We follow the theoretical and empirical literature on public-sector contracting to focus on measures of hierarchy, as well as on education and non-cognitive skills (Dixit, 2002; Mookherjee, 2006).

Figure 4 provides descriptive evidence on how individual characteristics correlate with the distribution of time across the five main categories used in this paper. Figure 4A shows that managers spend 2% more of their time in meetings relative to employees (significant at the 5-percent level) and 1% less of their time without work (significant at the 1-percent level). Figure 4B compares bureaucrats working at central and local levels of government and shows that central officials spend 2% less time in meetings (significant at the 5-percent level), 9% more of their time working alone (significant at the 1-percent level), and 7% less time traveling (significant at the 1-percent level). Figure 4C shows that bureaucrats with at least an undergraduate education spend 2% less of their time in meetings (significant at the 5-percent level), 4% more time working alone ($p = 0.001$), 3% less time traveling ($p = 0.013$) and 1% less time without work ($p = 0.015$). Figure 4D shows that those with more experience spend 2% more of their time working alone (significant at the 5-percent level) and 2% less time interfacing with clients ($p = 0.009$).¹⁰

We also explore whether individuals that exhibit higher levels of public-service motivation, as measured by the Perry (1996) scale, distribute their time differently in the bottom-left figure and similarly for those that exhibit higher internal locus of control, as measured by the Levenson (1981) scale, in Figures 4E and 4F respectively.¹¹ We find that those with higher-than-median public-service motivation spend 2% less time in meetings ($p = 0.01$) and 1% less time without work ($p = 0.018$). These may be some of the margins through which officials with higher public service motivation impact service delivery. We also find that those with higher-than-median internal locus of control spend 2% more

¹⁰Note that this finding is not purely a result of more senior job roles, as we find no differential time use by civil service grade level.

¹¹Internal locus of control is a measure of the extent to which individuals base success on their own actions and behavior, rather than external factors, such as chance, luck, or powerful others.

of their time working alone on average ($p = 0.09$).

Relative to the trade sector, we find that those working in health and revenue spend 2% more time in meetings (significant at the 5- and 10-percent level); those working in the education sector spend 8% more time working alone (significant at the 1-percent level); those working in agriculture, education, health and revenue spend significantly less time interfacing with clients; and those working in agriculture and health spend significantly more time traveling (4% and 3%, respectively, significant at the 1-percent and 5-percent level). We find that those working in more rural areas (higher than median levels of rurality) spend significantly less time traveling (4%, $p = 0.000$) and more time working alone (4%, $p = 0.001$). Gender has little impact on reported time use, though men spend 2% more time traveling than women ($p = 0.01$).

These results suggest that the individual characteristics of bureaucrats are predictive of the way that employees allocate their time across activities. These findings have potential implications for the allocation of public-sector personnel to tasks.

3.2 Time Use, Principals and Management Practices

In the following we combine the data that we have on bureaucrat time-use with rich survey data about the incentive environment in which bureaucrats operate. Following the theoretical and empirical literature on the contracting environment of public-sector organizations, we center our analysis around management practices related to incentives and autonomy (Rasul and Rogger, 2018; Rasul et al., 2019; Bandiera et al., 2020).

3.2.1 Measuring Management Practices

To understand the incentive environments in which public officials in the organizations we are studying make decisions on time allocation, we also collected data on management practices. Following recent efforts to collect data on the management practices of public administrations (Rasul and Rogger, 2018; Rasul et al., 2019) we used a public sector version of the World Management Survey (WMS; Bloom and Van Reenen, 2007; Bloom et al., 2012) to elicit measures of the management practices under which public officials operate.

The WMS evaluation tool elicits management practices through a semi-structured interview covering 7 topics: flexibility, incentives, monitoring, roles, staff involvement, staffing, and targeting. Table A7 details each of the 19 management-related questions, by topic, as well as the 1-5 scoring grid used by our enumerators for each question. These questions provide a holistic assessment of practices related to the topic.¹² The index of aggregate management

¹²To provide a sense of the holistic (rather than specific) nature of these questions, we go through one example: a question relating to management practices relating to monitoring was, “In what kind of ways does your Directorate track how well it is delivering services? Can you give me an example?” Enumerators could then score responses on a continuous 1-5 scale, where for indication the scoring grid described a score of one as corresponding to circumstances where

practices is constructed as the mean across the z-score of each of the items described in Table A7 and hence provides an assessment of the strength of management practices in the organization over the 7 topics.

3.2.2 Incentives and Autonomy

Following Rasul and Rogger (2018) and Rasul et al. (2019), for the analysis below we group the management practices into those focused on ‘incentive-based’ management practices and those centered on ‘autonomy-focused’ management practices. This categorization follows the broad trend in the literature on public administrators, which compares interventions that differentially increase incentives and autonomy (Finan et al., 2017; Bandiera et al., 2020). The Management: Incentives aggregate index combines the monitoring, targeting, incentives and staffing components. These components can be considered those that focus on organizational processes around objective performance targets that are incentivized, tracked, and that feed into staff promotion and retention decisions. The Management: Autonomy aggregate index combines the roles, flexibility and staff involvement components. These components are those that focus on the extent to which the organization provides discretion to staff, adapts to the needs of the client and local environment, and where staff can play an active role in organization decisions. Further details are provided in an Appendix.

3.2.3 Time Use and Management Practices

In Figure 5, we find evidence that the incentive environment of public officials is significantly correlated with decisions over time use. In Figure 5A, we see that those operating under multiple principals spend 2% more time in meetings ($p = 0.013$); 3% less time working alone ($p = 0.004$); 2% less time interfacing with clients ($p = 0.002$); and 4% more time traveling ($p = 0.000$). This finding is consistent with the theoretical prediction that multiple principals necessitate more coordination, potentially at the expense of other productive activities. This also provides a potential micro-level mechanism through which public officials perform better under single principals (Gulzar and Pasquale, 2017).

In Figure 5B, we find that a one-standard-deviation increase in aggregate management practices is associated with a 4% increase in time spent working alone ($p = 0.000$); a 3% decrease in time interfacing ($p = 0.000$); and a 1% decrease in time without work ($p = 0.000$). In Figure 5C and 5D, we focus on particular components of management practices. Figure 5C shows that a one-standard-deviation increase in Management: Incentives, conditional on Management: Autonomy, is associated with a 3% increase in time spent working alone ($p = 0.088$); a 3% decrease in time spent interfacing ($p = 0.002$); and a 1% decrease in time spent without work ($p = 0.003$). On the other

the “Directorate does not track performance”; a score of three corresponded to, “Directorate tracks a number of performance indicators. These are seen and reviewed by senior management only”; and a score of five corresponded to, “Full set of indicators are tracked formally and continuously. Reviews are conducted regularly and involve representatives of all directorate staff groups. The results of the review are formally communicated to all staff.”

hand, we find no significant correlation between category-specific time-use and Management: Autonomy, conditional on Management: Incentives, as shown in Figure 5D. These results provide evidence that management practices and working conditions significantly interact with the distribution of bureaucrat time-use across activities. These results point in the direction of further research into the micro-level determinants of how bureaucrats allocate resources across individual activities as a potential factor in understanding organizational differences in service delivery outcomes.

4 Bureaucrat Time Use and Service Delivery

In this section, we extend the analysis to consider the contribution of time-use for service-delivery production in the public sector. We use organization-averages of the distribution of time, and exclude results collected through the Long|Week module. We explore how the distribution of time across activities correlates with service delivery outcomes and how these correlations interact with the management practices demonstrated within the organization.

4.1 Time Use and Service Delivery

In Table 2, we study how the organization-average distribution of time across activities correlates with the service delivery outcomes of the organization. We merge the survey data with the latest available administrative data on service delivery outcomes at the organization level. For education officials, we look at the number of primary school enrollees, the pupil-teacher ratio (inverse), the pupil-class ratio (inverse), and the pupils-per-school rate (inverse). For health officials, we collect data on the antenatal care rate, the contraceptive acceptance rate, and the rate of full immunization among infants in the district. For agriculture officials, we collect data on the share of households relying on subsistence farming (inverse) and the agriculture income per year per household. For tax officials, we collect data on the number of tax identification numbers issued in the district, the proportion of total income that the office generates from own sources and the proportion that is received from transfers (inverse). For trade officials, we collect data on the number of business licenses issued and the revenue collected from business licenses. Each of the indicators is converted into a z-score by subtracting the mean and dividing by the standard deviation, to create a unit-less measure, allowing the aggregation across such different scales of indices.¹³ We then take an average across each of these unit-less indicators for each sector office to capture an organization-specific measure of performance.

To structure our analysis, we posit the following production function:

$$Y = f(A, h_1, h_2)$$

¹³Those indicators that are increasing in ‘bad’ outcomes, those followed by an ‘(inverse)’ in the above, are first inverted such that a higher value corresponds to better service delivery outcomes.

Where Y refers to the service delivery outcome; A to the productivity of the organization, proxied by organizational management practices and work environment characteristics; h_1 to the hours spent working alone; h_2 to the hours spent coordinating, either within the office (in meetings) or with clients (interfacing). We combine the interfacing and meetings categories of time use into a measure of time spent ‘coordinating’ due to their relative comparability. We then combine the traveling and without work components as time spent not working. Our objects of interests are $\frac{dY}{dh_k}$, $k = \{1, 2\}$. That is, the marginal product of an hour spent working alone or coordinating, relative to not working. To accord with the above production function, we create a measure of hours of work spent on each activity. This multiplies the proportion of time spent on the stated activity with the total number of hours worked per week.¹⁴ Following Bandiera et al. (2017), we assume a straightforward, linear, production function and run the following regression:

$$y_j = \alpha + \beta_1 h_{1,j} + \beta_2 h_{2,j} + \lambda X_j + \gamma_1 M_{1,j} + \gamma_2 M_{2,j} + u_j$$

Where j represents the organization; y_j represents the service delivery indicator, $h_{k,j}$ represents the organization-average number of hours spent on activity k , $k = \{1, 2\}$. X_j represents the work environment characteristics of organization j , including sector fixed effects, to capture capital investments in the office and service delivery production constraints specific to the sector. $M_{1,j}$ represents the z-score of Management: Incentives in organization j , and $M_{2,j}$ represents the z-score of Management: Autonomy in organization j .

The results from this regression are presented in Table 2 columns (1) to (4). Column (1) presents the unconditional regression of the service delivery index on the distribution of time-use in the organization. Column (2) adds controls for management practices. Column (3) controls for sector fixed effects and column (4) adds the full set of work environment characteristics and the number of principals in the organization. The results provide evidence that time spent coordinating contributes significantly to better service delivery outcomes, whereas time spent working alone does not. This is some of the first micro-level evidence on how differential bureaucrat activities correlate with service delivery outcomes across multiple sectors in the bureaucracy (Bertrand et al., 2015). These findings are consistent with evidence highlighting the importance of coordination activities for service delivery outcomes, especially given the multi-tasking nature of public official’s assignments (Dixit, 2002; Dasgupta and Kapur, 2017). The results are also consistent with other studies that document the positive effects of incentives on performance in the public sector (Muralidharan and Sundararaman, 2011; Finan et al., 2017; Leaver et al, 2019).

¹⁴For clarity, we are not including data collected using the Long|Week module in results in this section, but show this set of results with the Long|Week option in Table A8.

4.2 Management, Time Use and Service Delivery

In columns (5) to (10), we begin to explore the interactive role of management practices and time use on service delivery. We run regressions of the following form:

$$y_j = \alpha + \beta_1 h_{1,j} + \beta_2 h_{2,j} + \gamma_1 M_{1,j} + \gamma_2 M_{2,j} + \delta_{k,l} h_{k,j} M_{l,j} + \lambda X_j + u_j$$

For $k, l = \{1, 2\}$. Column (7) includes an interaction between Management: Incentives and working alone, which is negative and significant at the five-percent level, suggesting that working alone reduces the positive contribution of incentive-based management practices to service delivery. Column (8) suggests that there is no significant interaction between Management: Incentives and time spent coordinating. Column (9) provides evidence that time spent working alone is also less productive, for service delivery, under greater levels of autonomy. The relevant coefficient is negative and significant at the five-percent level. Column (10) shows that there is no significant interaction between time spent coordinating and Management: Autonomy. These results provide suggestive evidence that time spent working alone is worse for service delivery when agents are provided with greater performance incentives, in line with the findings from Rasul and Rogger (2018) and Rasul et al. (2019). However, we also find that time spent working alone is also worse for service delivery when agents operate in an environment with greater levels of discretion. Together, these results reinforce the notion that public administration is a multi-dimensional setting, requiring effective coordination between agents for efficient production.

In Table A9, we explore whether the average individual characteristics in the organization significantly interact with time-use in the production of public services. We find evidence that coordination is more productive for public-service delivery when employees are more educated but not when they are more experienced. Overall, while these results are purely associative and not causal, they provide a basis for further research into what types of activities may be better suited to certain types of bureaucrats and how these activities interact with the management environment to impact service delivery.

5 Conclusion

This paper provides evidence on the time use of public administrators in the developing world along three margins. First, it provides experimental evidence on the optimal modes for surveying public officials based on standard approaches to time use enumerated in other settings. Second, it provides baseline descriptive evidence on the nature of work undertaken by public officials across government. Third, it presents significant correlational evidence that official's time use matters for service delivery outcomes.

Measuring time use in the public administration rests on the resolution of a tension between the relatively high level of education of public officials and

the homogeneity, but varying intensity, of their tasks. Our results imply that, consistent with existing studies, the salience of a task critically determines the accuracy of recall on its time investment. For some aspects of public administration, such as interfacing with citizens or travelling for work, this works in favor of accurate recall. However, in tasks that make up the majority of a public officials time - working alone and in meetings - this distorts their ability to accurately reflect their time use. Baseline characteristics important in other studies of time use, such as the level of education, are not predictive of these errors. Rather, they seem to be prevalent across the 1,776 public administrators and 380 public organizations of the Ethiopian Civil Service that we study.

We experiment with 4 modes of stylized questions in a survey of public officials across all three tiers of Ethiopia’s government, and benchmark these against time use diaries enumerated to a subset of our respondents. In terms of accuracy, the Short|Week, Short|Month and Graphical|Week modules perform similarly relative to each other but the Long|Week option significantly overestimates the proportion of time spent in meetings and underestimates the time working alone. We find limited impacts of the different modes on the perceived quality of the survey process, with almost all surveys being rated as successful along all margins we assess. The proximate results of the former three methods gives researchers flexibility in the structure of stylized survey questions in terms of recall period and enumeration method. Given that the Graphical|Week module takes 4 minutes longer to enumerate than the other modes and has associated data entry costs, there is a rationale to restrict the use of this mode to environments where a graphical interface would be easier for respondents to engage with than a list.

The survey modules accurately capture the number of hours a public official undertakes paid work. This ability to accurately capture paid work time is in line with the wider literature on time use surveys. The granularity of our data identifies a ‘bump’ of roughly a third of officials working less than their mandated hours, which we do not observe in comparable labor force survey data.

Even then, all the survey modules distort reporting away from less salient tasks such as working alone and towards more salient tasks such as meetings and travel. The absolute difference between the survey response and the time-use diary excluding the Long|Week option is 62%. This is the sum of absolute differences in each of the categories, with the absolute differences amongst categories ranging from 6% to 21%. Given the double counting of these errors across categories, roughly a third of public administrator’s time is mis-reported in our survey data. This is towards the lower bound of existing estimates for other sectors such as housework, but still substantial. The proximate nature of some public officials, such as those working in ministries in a capital city, may make the enumeration and collection of diaries financially viable, and our study would suggest benefits to that approach. However, once such diaries must be collected from disparate entities across say a range of local governments, this may no longer be a feasible option.

To improve the quality of time use measurement in surveys of public admin-

istrators, future studies could experiment with markers to help officials better estimate the time that they are using on repeated and homogenous tasks in response to stylized survey questions. For example, by recovering details of what an official works alone on, and then assessing the time for each of those tasks might improve the accuracy of recall. A promising avenue for time use studies in public administration is the application of Kahneman et al. (2004)'s 'Day Reconstruction Method (DRM)'. By anchoring the previous day's sequence of events in their location, what specifically they were working on, who were they with, and the feelings they experienced, DRM generates greater texture in the description of homogenous tasks. Similarly, technological innovations in which a specially-equipped smartphone regularly prompts users to complete a self-administered survey that includes questions about their current activity may reduce the costs and increase the feasibility of diary methods.

The survey data we collect from across Ethiopia's public service provides the most comprehensive overview of how public administrators use their time to date. Almost half of their time is working alone, a quarter is in meetings, and the rest is made up by interactions with clients or travelling. The largest difference from the private sector is that senior bureaucrats spend less time in meetings and more time working alone, perhaps in contrast to popular conceptions of the two sectors.

Granular studies of time use like this one allow us to assess undocumented features of state capability through the lens of what determines a bureaucrat's use of time. Individual characteristics, including public service motivation and their sense of control over their work lives, mediates the use of a public officials time, as do the management practices they work under. The results we find are consistent with theoretical predictions relating to the public sector, such as multiple principal settings leading to more time invested in coordination. However, much of the variation we observe in time use is unexplained by the relatively rich data we have on individual characteristics and institutional environment. Specifically what determines time use in bureaucracy is still very much an open question.

Finally, we present correlations between the time use data we collect and the quality of services affected by the administrators we study. We show that investments in coordination in the public administration are substantial predictors of service quality. Time working alone mitigates the positive impacts of performance incentives on service delivery and negatively impacts any positive impacts of autonomy.

Together, our results imply the importance of human capital and its application through time allocation, to public sector outcomes. Strengthening state capabilities at the intensive margin will be critically about supporting the effective measurement and use of public official's time.

References

- [1] Ashraf, N., Bandiera, O. and Jack, B. K. (2014). No Margin, No Mission? A Field Experiment on Incentives for Public Service Delivery. *Journal of Public Economics*, 120: 1-17.
- [2] Ashraf, N., Bandiera, O. and Lee, S. (2018). Losing Prosociality in the Quest for Talent? Sorting, Selection, and Productivity in the Delivery of Public Services, Mimeo.
- [3] Bandiera, O., Prat, A., and Sadun, R. (2012). Managerial capital at the top: Evidence on CEOs time use and firm performance in India. Working Paper.
- [4] Bandiera, O., Hansen, S., Prat, A., and Sadun, R. (2017). CEO Behavior and Firm Performance. National Bureau of Economic Research, Working Paper 23248.
- [5] Bandiera, O., Best, M. C., Khan, A. Q., and Prat, A. (2020). The Allocation of Authority in Organizations: A Field Experiment with Bureaucrats. National Bureau of Economic Research, Working Paper 26733.
- [6] Becker, G. S. (1965). A Theory of the Allocation of Time. *The Economic Journal*, 493-517.
- [7] Beegle, K., De Weerd, J., Friedman, J., and Gibson, J. (2012). Methods of Household Consumption through Surveys: Experimental Results from Tanzania. *Journal of Development Economics*, 98(1), 3-18.
- [8] Bertrand, M., Burgess, R., Chawla, A., and Xu, G. (2015). Determinants and Consequences of Bureaucrat Effectiveness: Evidence from the Indian Administrative Service, Mimeo.
- [9] Bertrand, M., Burgess, R., Chawla, A., and Xu, G. (2018). The glittering prizes: Career incentives and bureaucrat performance. *The Review of Economic Studies*.
- [10] Bianchi, S. M., Milkie, M. A., Sayer, L. C., and Robinson, J. P. (2000). Is anyone doing the housework? Trends in the gender division of household labor. *Social Forces*, 79(1), 191-228.
- [11] Bloom, N. and Van Reenan, J. (2007). Measuring and Explaining Management Practices Across Firms and Countries. *Quarterly Journal of Economics* 122 (4): 1351-1408.
- [12] Bloom, N. Sadun, R. and Van Reenan, J. (2012). The Organization of Firms Across Countries. *Quarterly Journal of Economics*, 127 (4): 1663-1705.
- [13] Bonke, J. (2005). Paid Work and Unpaid Work: Diary Information Versus Questionnaire Information. *Social Indicators Research*, 70: 349-368.

- [14] Callen, M., Gulzar, S., Hasanain, A., Khan, M. Y. and Rezaee, A. (2018). Data and policy decisions: Experimental evidence from Pakistan. Stanford Institute of Economic Policy Research (SIEPR).
- [15] Chang, L. and Krosnick, J.A. (2003). Measuring the Frequency of Regular Behaviors: Comparing the “Typical Week” to the “Past Week”. *Sociological Methodology*, 33: 55-80.
- [16] Crossley, T. F., Winter, J. K. (2015). Asking Households About Expenditures: What Have We Learned?. Chapter 1 of Carroll, C. D., Crossley, T. F. and Sabelhaus, J. (eds). (2015). *Improving the Measurement of Consumer Expenditures*. Boston: NBER.
- [17] Dasgupta, A., and Kapur, D. (2017). The Political Economy of Bureaucratic Overload: Evidence from Rural Development Officials in India. Working Paper.
- [18] Delfgaauw, J., and Dur, R. (2007). Incentives and Workers’ Motivation in the Public Sector. *The Economic Journal*, 118 (525): 171-191.
- [19] Dixit, A. (2002). Incentives and Organizations in the Public Sector: An Interpretative Review. *Journal of Human Resources*, 37(4): 696-727.
- [20] Finan, F., Olken, B. A., and Pande, R. (2017). The Personnel Economics of the Developing State. *Handbook of Economic Field Experiments*, 2: 467-514.
- [21] Frazis, H., and Stewart, J. (2004). What can time-use data tell us about hours of work? *Monthly Labor Review*, 127(12): 3-9.
- [22] Gaddis, I., Oseni, G., Palacios-Lopez, A., & Pieters, J. (2020). Measuring Farm Labor: Survey Experimental Evidence from Ghana. *The World Bank Economic Review*.
- [23] Gibson, J. and Kim, B. (2007). Measurement Error in Recall Surveys and the Relationship between Household Size and Food Demand. *American Journal of Agricultural Economics*, 89(2): 473-489.
- [24] Gulzar, S. and Pasquale, B. J. (2017). Politicians, Bureaucrats, and Development: Evidence from India. *American Political Science Review*, 111(1): 162-183.
- [25] Hurst, E. (2015). Measuring time use in household surveys. *Journal of Economic and Social Measurement*, 40(1-4), 177-196.
- [26] Juster, F. T., Ono, H., and Stafford, F. P. (2003). An assessment of alternative measures of time use. *Sociological Methodology*, 33(1), 19-54.
- [27] Kan, M. Y., and Pudney, S. (2008). 2. Measurement Error in Stylized and Diary Data on Time Use. *Sociological Methodology*, 38(1), 101-132.

- [28] Kahneman, D., Krueger, A. B., Schkade, D. A., Schwarz, N., and Stone, A. A. (2004). A Survey Method for Characterizing Daily Life Experience: The Day Reconstruction Method. *Science*, 306(5702): 1776-1780.
- [29] Kelly, P., Thomas, E., Doherty, A., Harms, T., Burke, Ó., Gershuny, J., and Foster, C. (2015). Developing a method to test the validity of 24 hour time use diaries using wearable cameras: a feasibility pilot. *PLoS One*, 10(12), e0142198.
- [30] Khan, A. Q., Khwaja, A. I., and Olken, B. A. (2019). Making moves matter: Experimental evidence on incentivizing bureaucrats through performance-based postings. *American Economic Review*, 109 (1): 237-70.
- [31] Kitterod, R. H., and Lyngstad, T. H. (2005). Diary versus questionnaire information on time spent on housework - The case of Norway. *Electronic International Journal of Time Use Research*, 2: 13-32.
- [32] Leaver, C., Ozier, O., Serneels, P., and Zeitlin, A. (2019). Recruitment, Effort, and Retention Effects of Performance Contracts for Civil Servants: Experimental Evidence from Rwandan Primary Schools. Mimeo. Washington: World Bank.
- [33] Levenson, H. (1981). Differentiating among internality, powerful others, and chance. In H. Lefcourt (Ed.), *Research with the Locus of Control Construct* (Vol. 1), New York: Academic Press, 15-63.
- [34] Lewis-Faupel, S., Y. Neggers, B. A. Olken, and R. Pande. (2016). Can Electronic Procurement Improve Infrastructure Provision? Evidence from Public Works in India and Indonesia. *American Economic Journal: Economic Policy*, 8 (3): 258-83.
- [35] Link, A. N., and van Hasselt, M. (2019). A Public Sector Knowledge Production Function. *Economics Letters*, 175: 64-66.
- [36] Marini, M. and Shelton, B. A. (1993). Measuring household work: Recent experience in the United States. *Social Science Research*, 4: 361-385.
- [37] Masuda, Y. J., Fortmann, L., Gugerty, M. K., Smith-Nilson, M. and Cook, J. 2014. Pictorial Approaches for Measuring Time Use in Rural Ethiopia. *Social Indicators Research*, 115(1): 467–82.
- [38] Mookherjee, D. (2006). Decentralization, Hierarchies, and Incentives: A Mechanism Design Perspective. *Journal of Economic Literature*, 44(2): 367-390.
- [39] Muralidharan, K., and Sundararaman, V. (2011). Teacher Performance Pay: Experimental Evidence from India. *Journal of Political Economy*, 119(1): 39-77.

- [40] Niemi, I. (1993). Systematic error in behavioural measurement: Comparing results from interview and time budget studies. *Social Indicators Research*, 30: 229-244.
- [41] Otterbach, S. and Sousa-Poza, A. (2010). How Accurate are German Work-time Data? A Comparison of Time-diary Reports and Stylized Estimates. *Social Indicators Research*, 97: 325-339.
- [42] Perry, J. L. (1996). Measuring public service motivation: An assessment of construct reliability and validity. *Journal of Public Administration Research and Theory*, 6(1), 5-22.
- [43] Pradhan, M. (2001). Welfare Analysis with a Proxy Consumption Measure - Evidence from a Repeated Experiment in Indonesia. Tinbergen Institute, Tinbergen Institute Discussion Papers.
- [44] Rasul, I. and Rogger, D. (2018). Management of Bureaucrats and Public Service Delivery: Evidence from the Nigerian Civil Service. *Economic Journal*, 128: 413-446.
- [45] Rasul, I., Rogger, D. & Williams, M. (2019). Management and Bureaucratic Effectiveness: Evidence from the Ghanaian Civil Service. Mimeo, University College London.
- [46] Robinson, J., and Gershuny, J. (2012). Occupational Differences in Estimates of Time at Work. Maryland Population Research Center. Working Paper.
- [47] Schulz, F., and Grunow, D. (2012). Comparing Diary and Survey Estimates on Time Use. *European Sociological Review*, 28(5): 622-632.
- [48] Scott, C., and Amenuvegbe, B. (1990). Effect of Recall Duration on Reporting of Household Expenditures: An Experimental Study in Ghana. The World Bank. Working Paper.
- [49] Seymour, G., Malapit, H., & Quisumbing, A. (2017). Measuring time use in development settings. The World Bank. Working Paper.
- [50] Sudman, S., Bradburn, N., & Schwarz, N. (1996). Thinking about answers: The Application of cognitive processes to survey methodology. San Francisco, CA: Jossey-Bass.

Appendix: Time-Use Diary

2016 ETHIOPIA CIVIL SERVANTS' SURVEY – TIME USE DIARY

INTRODUCTION

This part of the survey will be self-administered over the next five working days. The goal of these questions is to understand how you use your time over a typical work week.

Once again, we would like to remind you that this survey is being undertaken in the strictest confidence. In particular:

- a) This study is being undertaken by independent researchers who will keep your answers completely secure; and,
- b) No one outside of the research team will be able to associate the individual answers you provide with you, so please feel free to answer honestly.

UNDERSTANDING YOUR TIME USE

In the following exercise, we want to understand how you use your time over a typical week. You will see five types of activity. **For the next five working days, we would like you to estimate the number of hours you spend on each of the five activities per day.** In one week, an enumerator will come to your office to collect your completed form.

WORKING DAY 1		
	TODAY, how many hours did you spend on the following activities:	Number of hours spent on activity
(a)	Meetings within the service	
(b)	Working alone/on administration	
(c)	Interfacing with people outside of the civil service	
(d)	Travelling for work	
(e)	Without work to do	

WORKING DAY 2		
	TODAY, how many hours did you spend on the following activities:	Number of hours spent on activity
(a)	Meetings within the service	
(b)	Working alone/on administration	
(c)	Interfacing with people outside of the civil service	
(d)	Travelling for work	
(e)	Without work to do	

WORKING DAY 3		
	TODAY, how many hours did you spend on the following activities:	Number of hours

Appendix: Time-Use Survey Modules

Short Week

Q3.1	Out of those <answer from Q1> hours in a week, how many hours do you spend on the following activities:	Number of hours spent on activity
(a)	Meetings within the service	
(b)	Working alone/on administration	
(c)	Interfacing with people outside of the civil service (e.g. clients)	
(d)	Travelling for work	
(e)	Interacting with frontline workers	
(f)	Without work to do/waiting for others to input	

Long Week

Q3.2	Out of those <answer from Q1> hours, how many hours do you spend on the following activities:	Number of hours spent on activity
(a)	Meetings with your direct supervisor	
(b)	Meetings with case team members and change army team members	
(c)	Organisation level meetings (with all staff)	
(d)	Awareness raising meetings on policies, strategies, and others	
(e)	Other meetings	
(f)	Interacting with clients outside of government (e.g. users)	
(g)	Activities and meetings outside of your mission	
(h)	Sending and responding to emails or making phone calls	
(i)	Travelling for work/field visits	
(j)	Personal time (Breaks, down time, coffee, rests, internet browsing)	
(k)	Administrative duties (e.g. preparing a budget)	
(l)	Preparing work plans / planning / capacity building	
(m)	Evaluations or appraisals (staff or activities)	
(n)	Without any work to do/waiting for others to input	

Short Month

Q3.3	Of the roughly <[answer from Q1*4.38(number of weeks in a month)] rounded to nearest 10 hours> in a typical month, what proportion of time do you spend on the following activities:	Proportion of hours spent on activity
(a)	Meetings within the service	
(b)	Working alone/on administration	
(c)	Interfacing with people outside of the civil service	
(d)	Travelling for work	
(e)	Without work to do/waiting for others to input	

Graphical Week

NOTE TO ENUMERATOR: Provide respondent with 'Your Work Week' form and read the following: "In the following exercise, we want to understand how you use your time in a typical week. **Please draw in lines to represent the proportion of time you spend on the following activities:** (a) Meetings within the service; (b) Working alone/on administration; (c) Interfacing with people outside of the civil service; (d) Travelling for work; (e) Without work to do/waiting for others to input"

Figure 1: The Distribution of Civil Servant Time Use

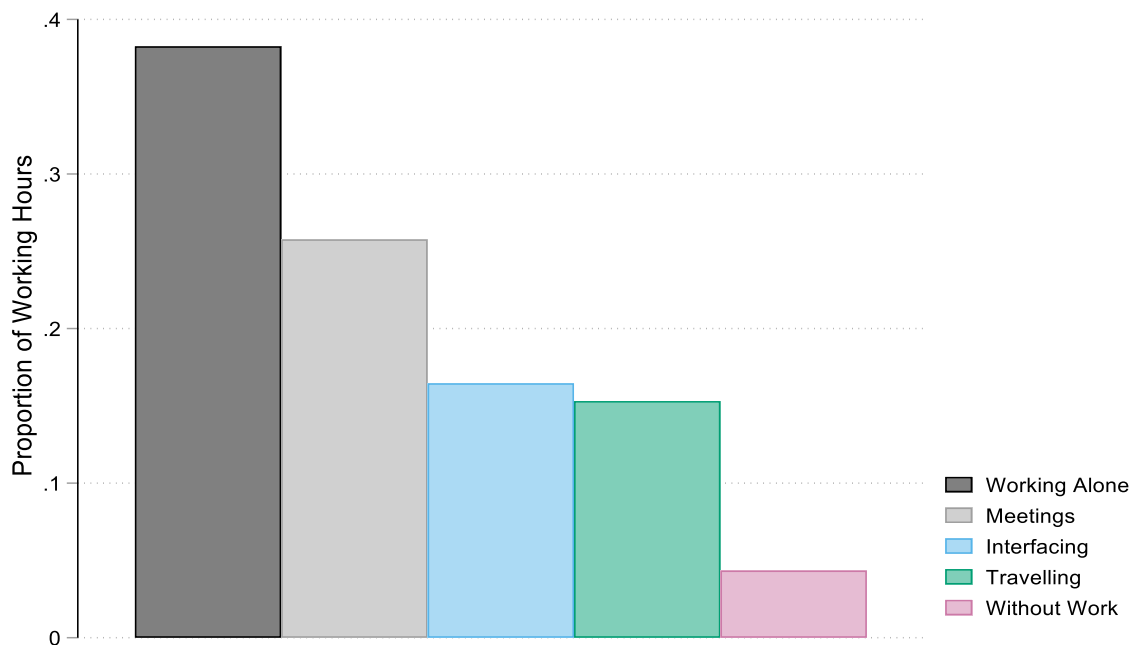
1A: Hours per Week: Public and Private Sectors



Public Sector, LFS includes only public administrators in Addis Ababa

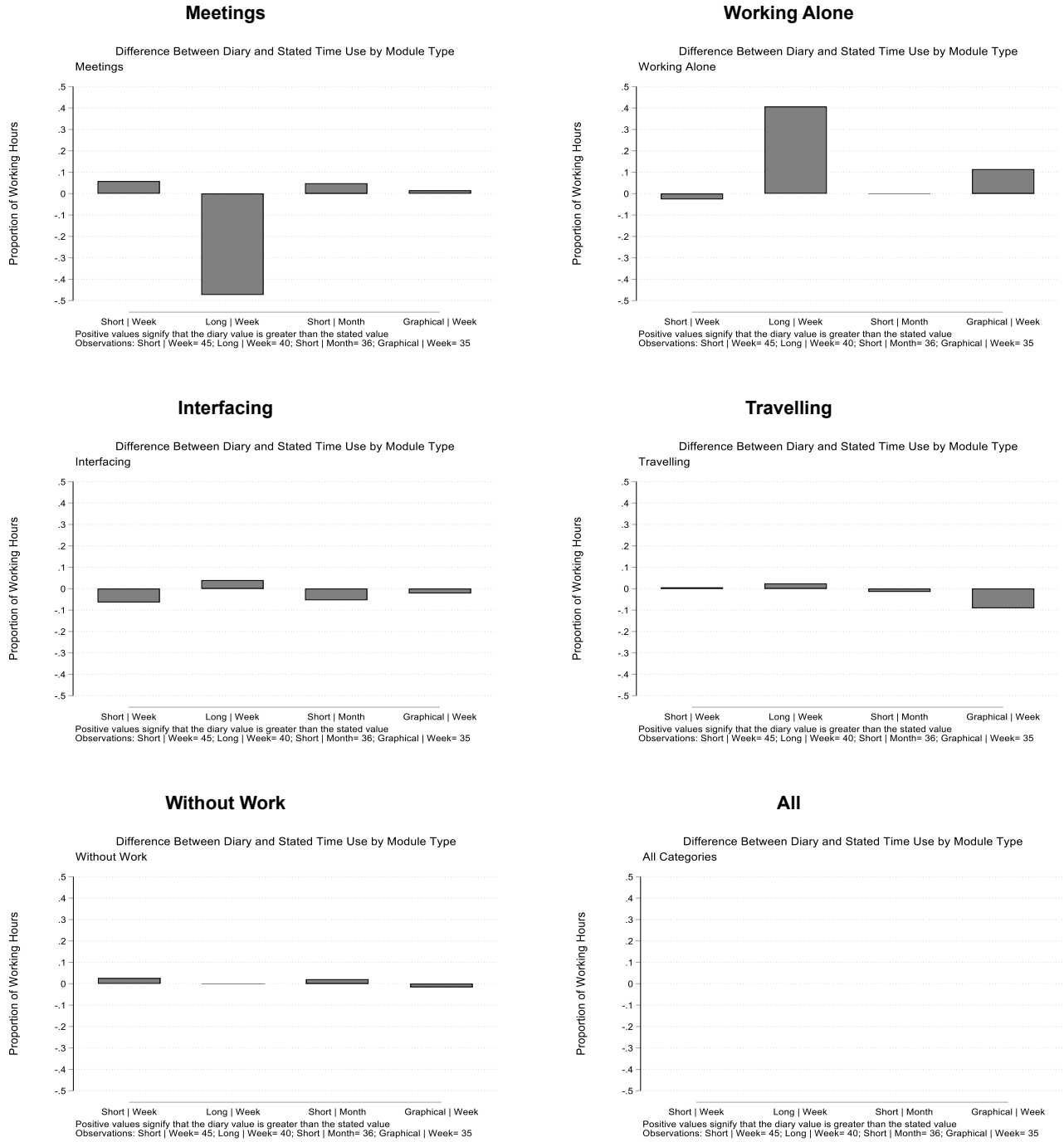
1B: The Distribution of Time Across Activities

Distribution of Time Use Across Activities, All Survey Options, All Respondents



Notes: The top figure shows the kernel density distributions of stated hours worked per week for private-sector employees and public administrators in Addis Ababa using the Labor Force Survey (LFS) and the Ethiopian Civil Servants Survey (ECSS). The bottom figure shows the average proportion of time spent on each activity category across all respondents and all survey methods.

Figure 2: Raw Differences Between Stylized Modules and Time-Use Diaries

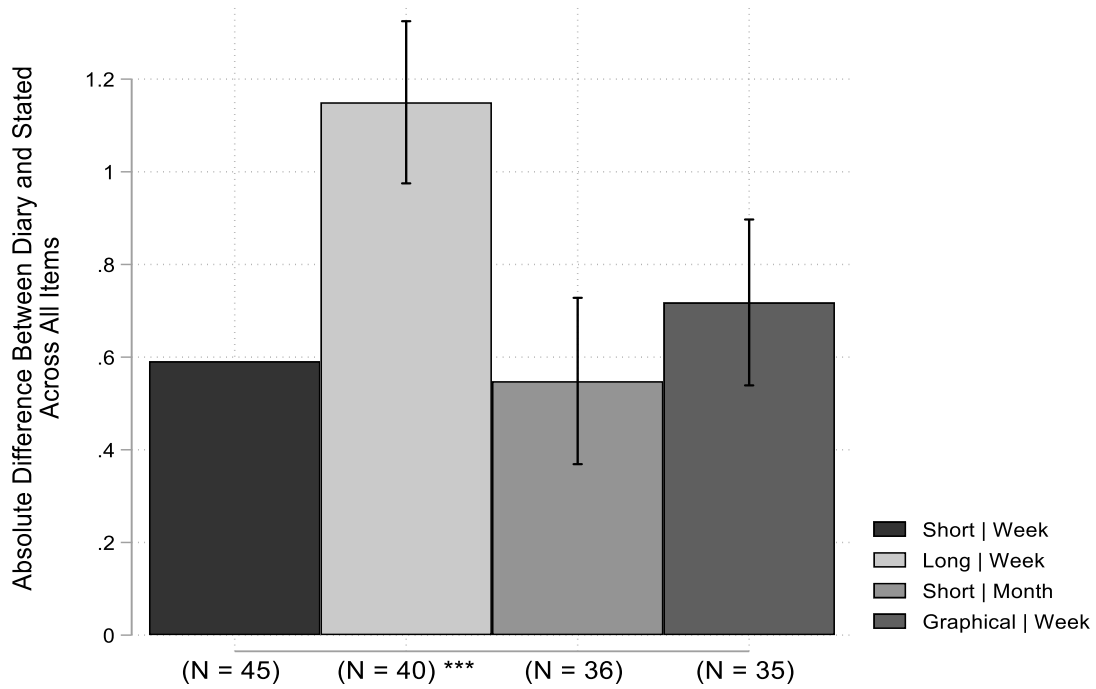


Notes: The figure shows the raw difference between the proportion of working hours spent in the specified activity category as stated in the time-use diary minus the time spent as stated in the stylized module survey. Each sub-figure represents the activity category, as titled. Each column represents the stylized survey module. Positive values mean that the time-use diary value is larger than the stylized survey module value.

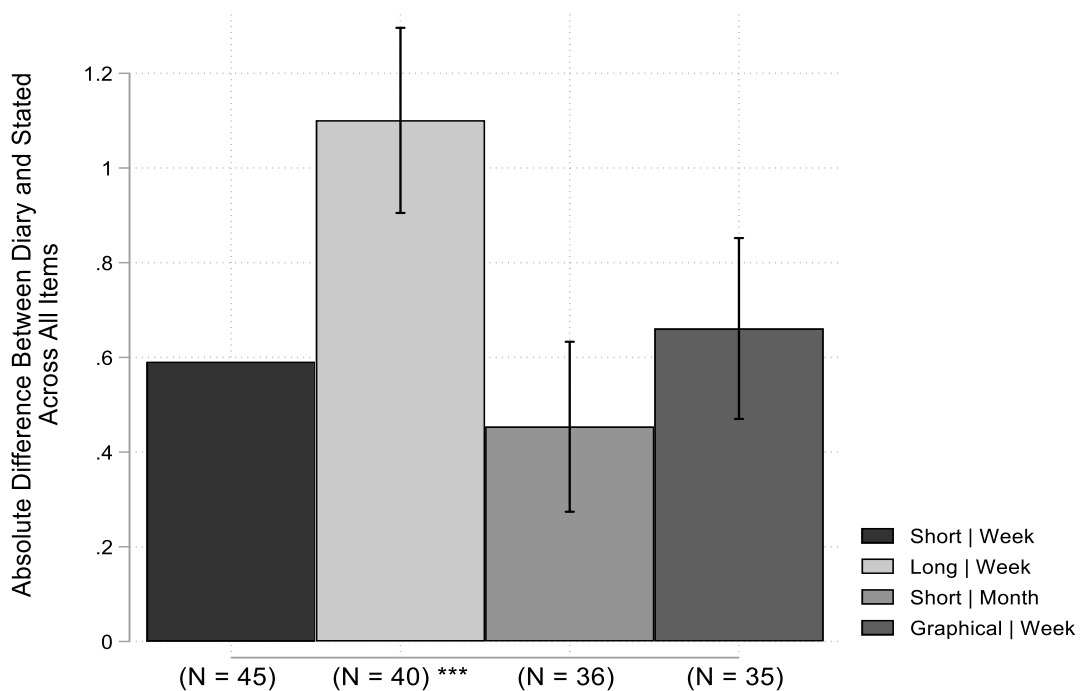
Figure 3: Treatment Effects, Central Respondents

Dependent Variable: Absolute Difference Between Diary Entry and Stated Time Use By Randomized Survey Method

3A: Unconditional Treatment Effects



3B: Conditional Treatment Effects



Notes: Confidence intervals based on robust standard errors. Conditional treatment effects are conditional on individual characteristics, work characteristics, and survey characteristics. Individual characteristics include: age, an indicator for whether the respondent is male, an indicator for the highest education qualification of the respondent; years in the current position, years in the current organization, years in the civil service; an indicator for the sector of work. Work environment characteristics include the organizational access to electricity, phone networks, internet, computers, vehicles, and skilled staff, and aggregate management practices as measured by the World Management Survey. Survey environment characteristics include: enumerator fixed effects; an indicator for whether the interview was conducted in complete privacy; an indicator for whether the respondent was knowledgeable about own environment; an indicator for whether the respondent was knowledgeable about own environment and organization; an indicator for whether the respondent was willing to reveal basic and confidential information; an indicator for whether the respondent seemed very patient; an indicator for whether the interview was perceived to go very well by enumerator. The sample includes civil servants in the capital city, Addis Ababa, for whom time-use diaries were provided and tracked.

Table 1: Treatment Effects: Absolute Differences Between Diary and Stated Time Use

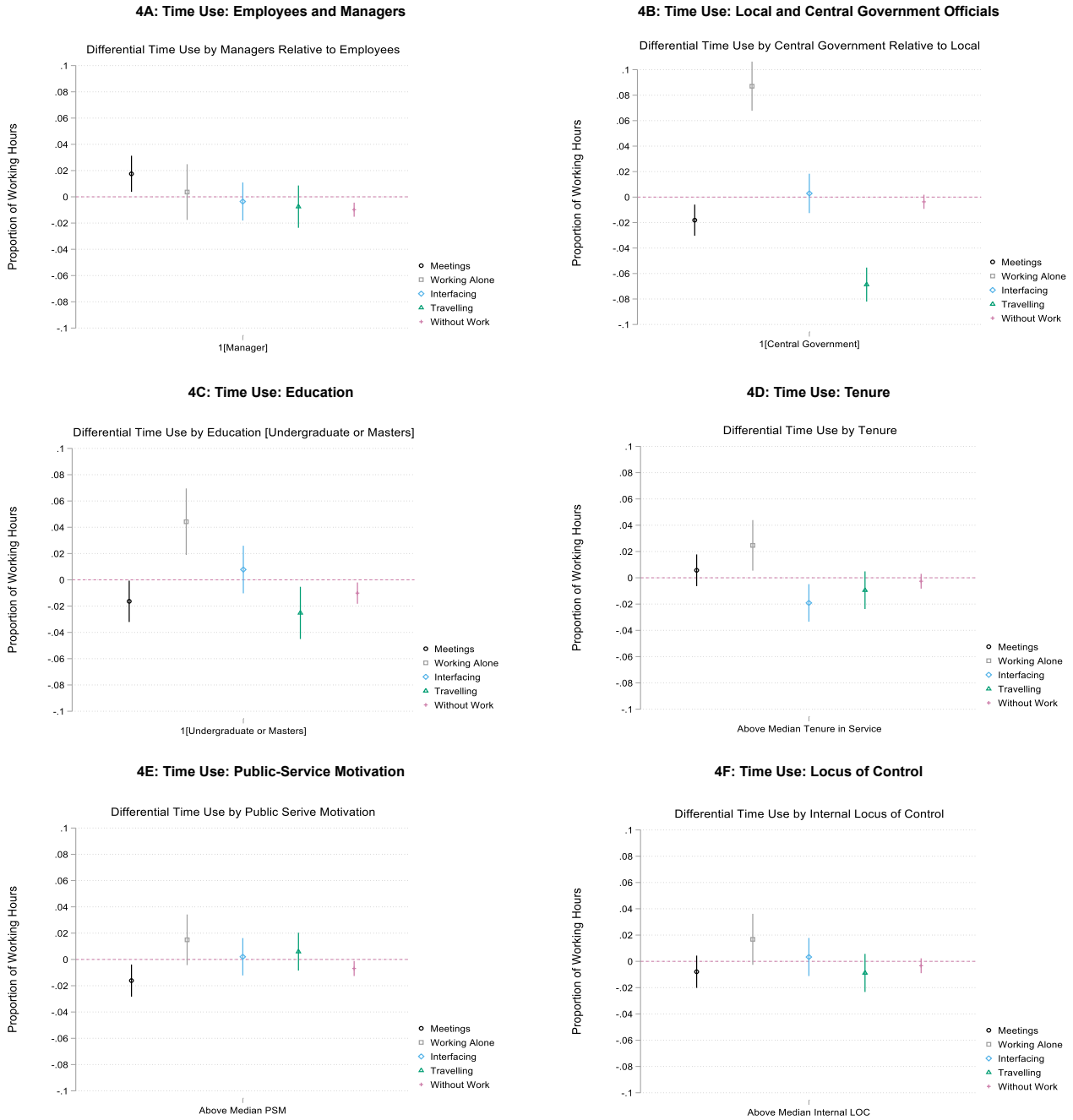
OLS Estimates and Standard Errors

Robust Standard Errors Reported

	(1) Overall Absolute Difference	(2) Overall Absolute Difference	(3) Overall Absolute Difference	(4) Overall Absolute Difference	(5) Meetings Absolute Difference	(6) Working Alone Absolute Difference	(7) Interfacing Absolute Difference	(8) Travelling Absolute Difference	(9) Without Work Absolute Difference
Long Week [2]	0.56*** [0.089]	0.54*** [0.089]	0.53*** [0.095]	0.53*** [0.11]	0.36*** [0.054]	0.22*** [0.060]	-0.0026 [0.041]	-0.039 [0.038]	-0.011 [0.020]
Short Month [3]	-0.043 [0.091]	-0.12 [0.091]	-0.11 [0.089]	-0.045 [0.089]	-0.030 [0.039]	0.0037 [0.043]	-0.010 [0.034]	-0.0049 [0.036]	-0.0034 [0.017]
Graphical Week [4]	0.13 [0.091]	0.076 [0.091]	0.060 [0.094]	0.12 [0.11]	0.012 [0.045]	0.061 [0.055]	0.020 [0.027]	0.040 [0.041]	-0.010 [0.022]
<i>p-value [2]=[3]</i>	0.00	0.00	0.00	0.00	0.00	0.00	0.41	0.27	0.50
<i>p-value [2]=[4]</i>	0.00	0.00	0.00	0.00	0.00	0.00	0.88	0.03	0.73
<i>p-value [3]=[4]</i>	0.08	0.04	0.08	0.10	0.38	0.28	0.39	0.21	0.75
Individual controls	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Work environment controls	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Survey controls	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes
Control Mean		0.59			0.11	0.21	0.11	0.10	0.06
Adjusted R-squared	0.25	0.30	0.29	0.35	0.34	0.32	0.025	0.053	0.13
Observations	156	156	156	156	156	156	156	156	156

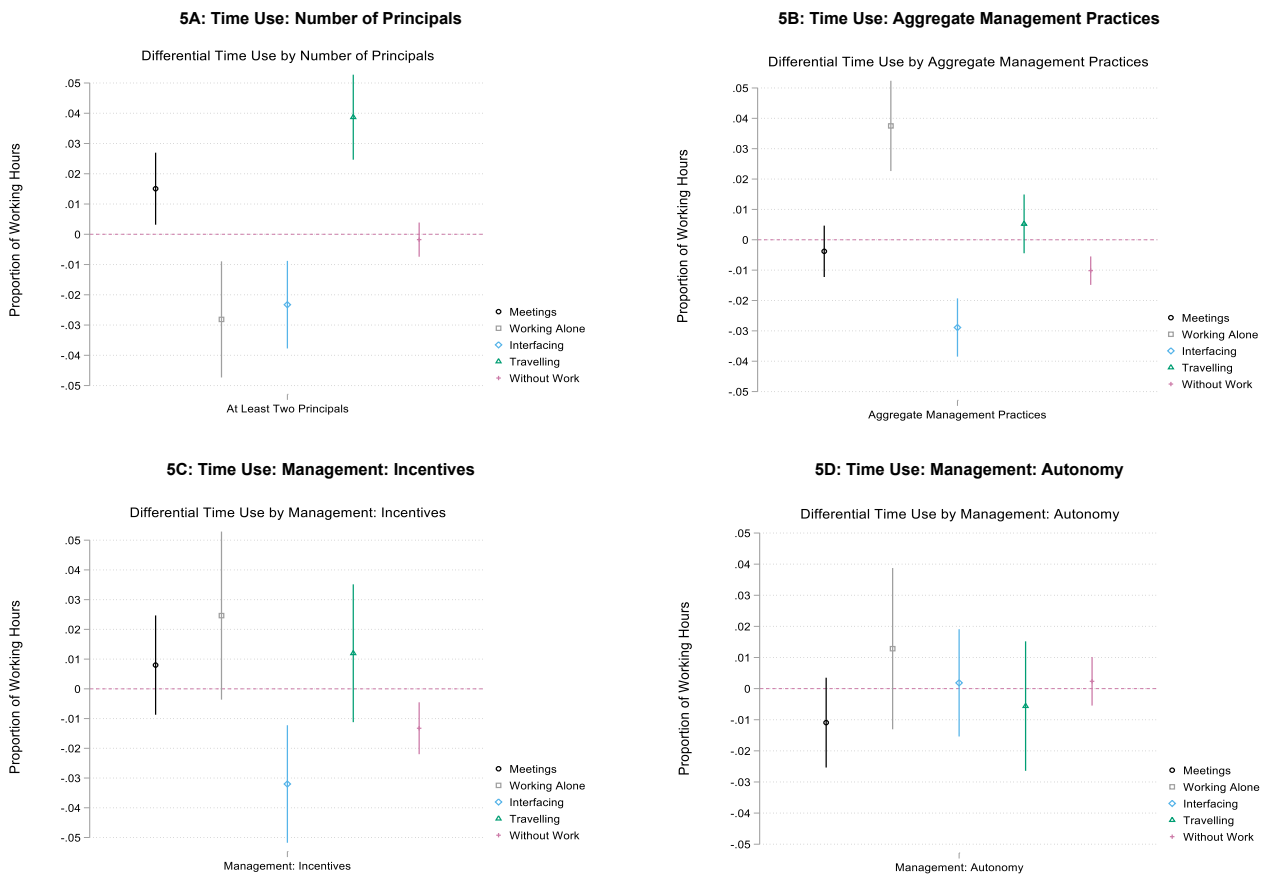
Notes: * p<0.1; ** p<0.05; *** p<0.01. Robust standard errors in brackets. The dependent variable in columns (1) to (4) is the absolute difference, across each time-use category, between the time-use diary and the survey response; the dependent variable in columns (5) to (9) is the absolute difference between the time-use diary and the survey response for each of the categories of time use noted in the column heading. Individual characteristics include: age, an indicator for whether the respondent is male, an indicator for the highest education qualification of the respondent; years in the current position, years in the current organization, years in the civil service; an indicator for the sector of work. Work environment characteristics include the organizational access to electricity, phone networks, internet, computers, vehicles, and skilled staff, and aggregate management practices as measured by the World Management Survey. Survey environment characteristics include: an indicator for whether the interview was conducted in complete privacy; an indicator for whether the respondent was knowledgeable about own environment; an indicator for whether the respondent was knowledgeable about own environment and organization; an indicator for whether the respondent was willing to reveal basic and confidential information; an indicator for whether the respondent seemed very patient; an indicator for whether the interview was perceived to go very well by enumerator; and enumerator fixed effects. Figures rounded to two significant figures

Figure 4: The Distribution of Time Use Across Activities by Individual Characteristics



Notes: The figure shows the OLS estimate and confidence interval of management factors on the proportion of time spent on each activity. The sample is the full sample of civil servants. The presented coefficient is an indicator for whether the respondent is a manager (top-left); an indicator for whether the respondent works in the central government (top-right); an indicator for whether the respondent has an undergraduate degree or higher (second-row-left); an indicator for whether the respondent has above-median experience in the civil service (second-row-right); an indicator for whether the respondent exhibits higher-than-median public-service motivation as measured by the Perry (1996) scale (bottom-left); and an indicator for whether the respondent exhibits higher-than-median internal locus of control, as measured by the Levenson (1981) scale. The results are conditional on the indicator for the randomized survey method given to the respondent.

Figure 5: The Distribution of Time Use Across Activities by Management



Notes: The figure shows the OLS estimate and confidence interval of management factors on the proportion of time spent on each activity. The sample is the full sample of civil servants. The presented coefficient is an indicator for whether the respondent operates under more than two principals (top-left); Aggregate Management Practices (Z-Score), as measured by the World Management Survey scale (top-right); Management: Incentives (Z-Score) (bottom-row-left); Management: Autonomy (Z-Score) (bottom-right). Management: Incentives is the average z-score across incentives, targeting, monitoring, and staffing components of management; Management: Autonomy is the average z-score across the roles, flexibility, and staff involvement components of management. All results are conditional on the indicator for the randomized survey method given to the respondent. The results for Management: Incentives and Management: Autonomy condition on both indices of management practices.

Table 2: Time Use, Management, and Service Delivery

OLS Estimates and Standard Errors

Robust Standard Errors Reported

Dependent Variable: Z-Score Across All Organization-Level Service Delivery Indicators (District-Level Only)

	(1) Time Use Only	(2) Management practices	(3) Sector FE	(4) District Characteristics	(5) Lagged Service Delivery Index	(6) Work environment characteristics and number of principals	(7) Working alone and incentives	(8) Coordination and incentives	(9) Working alone and autonomy	(10) Coordination and autonomy
Hours working alone (organization average)	0.0037 [0.0028]	0.0012 [0.0031]	0.0020 [0.0032]	0.0034 [0.0029]	0.0014 [0.0026]	0.0013 [0.0028]	0.0019 [0.0027]	0.0012 [0.0028]	0.00098 [0.0028]	0.00096 [0.0028]
Hours coordinating (interfacing and in meetings, organization average)	0.0089** [0.0036]	0.0072* [0.0037]	0.0074* [0.0038]	0.0091** [0.0037]	0.0069** [0.0032]	0.0067* [0.0035]	0.0073** [0.0035]	0.0068* [0.0035]	0.0072** [0.0035]	0.0074** [0.0036]
Hours other (organization average: travelling and without work, organization average)	Omitted Category									
Management: Incentives		0.15*** [0.050]	0.16*** [0.053]	0.070 [0.049]	0.046 [0.043]	0.054 [0.044]	0.19** [0.088]	-0.0067 [0.069]	0.055 [0.044]	0.046 [0.044]
Management: Autonomy		-0.066 [0.047]	-0.078 [0.049]	0.00040 [0.045]	-0.0055 [0.039]	-0.022 [0.043]	-0.032 [0.044]	-0.026 [0.044]	0.096 [0.074]	-0.11 [0.074]
Number of principals						0.020 [0.025]	0.013 [0.025]	0.020 [0.025]	0.017 [0.025]	0.020 [0.025]
Hours working alone x Management: Incentives							-0.0070** [0.0033]			
Hours coordinating x Management: Incentives								0.0045 [0.0044]		
Hours working alone x Management: Autonomy									-0.0068** [0.0033]	
Hours coordinating x Management: Autonomy										0.0068 [0.0042]
Hours working alone x Number of principals										
Hours coordinating x Number of principals										
Sector fixed effects	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
District characteristics	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Lagged service delivery index	No	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes
Work environment characteristics	No	No	No	No	No	Yes	Yes	Yes	Yes	Yes
Long week option included						No				
Adjusted R-squared	0.015	0.041	0.031	0.19	0.35	0.34	0.35	0.34	0.35	0.35
Observations	276	276	276	276	276	276	276	276	276	276

Notes: * p<0.1; ** p<0.05; *** p<0.01. Robust standard errors in brackets. The unit of observation is the district organization. The dependent variable is the z-score of the service delivery index across all sectors in the sample. The sample is restricted to district-level organizations only. Time use measured by the long-week survey option are excluded. Management: Incentives is the average z-score across incentives, targeting, monitoring, and staffing components of management; Management: Autonomy is the average z-score across the roles, flexibility, and staff involvement components of management. Work environment characteristics include the organizational access to electricity, phone networks, internet, computers, vehicles, and skilled staff. District characteristics include population, the percentage of rural inhabitants, ethnic fractionalization, remoteness (travel time to nearest urban area). Lagged service delivery index is the service delivery index for the district from the most-recent previous data. Figures are rounded to two significant figures.

TA1: List of Surveyed Organizations

Tier of Governance	Region	Organization Name
Federal	-	Federal Ministries of Agriculture; Education; Health; Revenue; and, Trade
Regional	Addis Ababa	Addis Ababa City Administration Bureaus of Education; Health; Revenue; and, Trade
Regional	Afar	Afar Regional Bureaus of Agriculture; Education; Health; Revenue; and, Trade
Regional	Amhara	Amhara Regional Bureaus of Agriculture; Education; Health; Revenue; and, Trade
Regional	Benishangul Gumuz	Benishangul Gumuz Regional Bureaus of Agriculture; Education; Health; Revenue; and, Trade
Regional	Dire Dawa	Dire Dawa City Administration Bureaus of Agriculture; Education; Health; Revenue; and, Trade
Regional	Gambella	Gambella Regional Bureaus of Agriculture; Education; Health; Revenue; and, Trade
Regional	Harar	Harar Regional Bureaus of Agriculture; Education; Health; Revenue; and, Trade
Regional	Oromia	Oromia Regional Bureaus of Agriculture; Education; Health; Revenue; and, Trade
Regional	SNNPR	SNNPR Regional Bureaus of Agriculture; Education; Health; Revenue; and, Trade
Regional	Somali	Somali Regional Bureaus of Agriculture; Education; Health; Revenue; and, Trade
Regional	Tigray	Tigray Regional Bureaus of Agriculture; Education; Health; Revenue; and, Trade
District	Afar	Afar Awash Fentale Agriculture, Education, Health, Revenue, and Trade Offices
District	Afar	Afar Telalak Agriculture, Education, Health, Revenue, and Trade Offices
District	Afar	Afar Teru Agriculture, Education, Health, Revenue, and Trade Offices
District	Amhara	Amhara Awabel Agriculture, Education, Health, Revenue, and Trade Offices
District	Amhara	Amhara Basona Worana Agriculture, Education, Health, Revenue, and Trade Offices
District	Amhara	Amhara Borena (Former Debrešina) Agriculture, Education, Health, Revenue, and Trade Offices
District	Amhara	Amhara Chefa Gula Agriculture, Education, Health, Revenue, and Trade Offices
District	Amhara	Amhara Dejen Agriculture, Education, Health, Revenue, and Trade Offices
District	Amhara	Amhara Enarj Enawaga Agriculture, Education, Health, Revenue, and Trade Offices
District	Amhara	Amhara Gidane Agriculture, Education, Health, Revenue, and Trade Offices
District	Amhara	Amhara Jabitahnan Agriculture, Education, Health, Revenue, and Trade Offices
District	Amhara	Amhara Jile Timuga Agriculture, Education, Health, Revenue, and Trade Offices
District	Amhara	Amhara Kutaber Agriculture, Education, Health, Revenue, and Trade Offices
District	Amhara	Amhara Simada Agriculture, Education, Health, Revenue, and Trade Offices
District	Benishangul Gumuz	Benishangul Gumuz Dibate Agriculture, Education, Health, Revenue, and Trade Offices
District	Benishangul Gumuz	Benishangul Gumuz Yasso Agriculture, Education, Health, Revenue, and Trade Offices
District	Gambella	Gambella Gambella Zuria Agriculture, Education, Health, Revenue, and Trade Offices
District	Gambella	Gambella Abobo Agriculture, Education, Health, Revenue, and Trade Offices
District	Oromia	Oromia Ale Agriculture, Education, Health, Revenue, and Trade Offices
District	Oromia	Oromia Amigna Agriculture, Education, Health, Revenue, and Trade Offices
District	Oromia	Oromia Arsi Negelle District government Agriculture, Education, Health, Revenue, and Trade Offices
District	Oromia	Oromia Babile Agriculture, Education, Health, Revenue, and Trade Offices
District	Oromia	Oromia Bako Tibe Agriculture, Education, Health, Revenue, and Trade Offices
District	Oromia	Oromia Begi Agriculture, Education, Health, Revenue, and Trade Offices
District	Oromia	Oromia Dedessa Agriculture, Education, Health, Revenue, and Trade Offices
District	Oromia	Oromia Digluna Tijo Agriculture, Education, Health, Revenue, and Trade Offices
District	Oromia	Oromia Gida Ayana (Gida Kiremu) Agriculture, Education, Health, Revenue, and Trade Offices
District	Oromia	Oromia Goro Gutu (Goro) Agriculture, Education, Health, Revenue, and Trade Offices
District	Oromia	Oromia Guduru Agriculture, Education, Health, Revenue, and Trade Offices
District	Oromia	Oromia Haro Maya Agriculture, Education, Health, Revenue, and Trade Offices
District	Oromia	Oromia Hitosa Agriculture, Education, Health, Revenue, and Trade Offices
District	Oromia	Oromia Jardega Jarte Agriculture, Education, Health, Revenue, and Trade Offices
District	Oromia	Oromia Jeldu Agriculture, Education, Health, Revenue, and Trade Offices
District	Oromia	Oromia Kofale Agriculture, Education, Health, Revenue, and Trade Offices
District	Oromia	Oromia Mesela Agriculture, Education, Health, Revenue, and Trade Offices
District	Oromia	Oromia Midaga Tola Agriculture, Education, Health, Revenue, and Trade Offices
District	Oromia	Oromia Nono Agriculture, Education, Health, Revenue, and Trade Offices
District	Oromia	Oromia Seru Agriculture, Education, Health, Revenue, and Trade Offices
District	Oromia	Oromia Siraro Agriculture, Education, Health, Revenue, and Trade Offices
District	Oromia	Oromia Tikur Enchini Agriculture, Education, Health, Revenue, and Trade Offices
District	Oromia	Oromia Wadera Agriculture, Education, Health, Revenue, and Trade Offices
District	Oromia	Oromia Were Jarso Agriculture, Education, Health, Revenue, and Trade Offices
District	SNNPR	SNNPR Amaro Agriculture, Education, Health, Revenue, and Trade Offices
District	SNNPR	SNNPR Analimo Agriculture, Education, Health, Revenue, and Trade Offices
District	SNNPR	SNNPR Basketo Agriculture, Education, Health, Revenue, and Trade Offices
District	SNNPR	SNNPR Benatsemay Agriculture, Education, Health, Revenue, and Trade Offices
District	SNNPR	SNNPR Bona Agriculture, Education, Health, Revenue, and Trade Offices
District	SNNPR	SNNPR Chere Agriculture, Education, Health, Revenue, and Trade Offices
District	SNNPR	SNNPR Dale Agriculture, Education, Health, Revenue, and Trade Offices
District	SNNPR	SNNPR Decha Agriculture, Education, Health, Revenue, and Trade Offices
District	SNNPR	SNNPR Doyo Gena Agriculture, Education, Health, Revenue, and Trade Offices
District	SNNPR	SNNPR Gomibora Agriculture, Education, Health, Revenue, and Trade Offices
District	SNNPR	SNNPR Hawassa Zuriya Agriculture, Education, Health, Revenue, and Trade Offices
District	SNNPR	SNNPR Kucha Agriculture, Education, Health, Revenue, and Trade Offices
District	SNNPR	SNNPR Shebedino Agriculture, Education, Health, Revenue, and Trade Offices
District	SNNPR	SNNPR Wenago Agriculture, Education, Health, Revenue, and Trade Offices
District	Somali	Somali Afdem Agriculture, Education, Health, Revenue, and Trade Offices
District	Somali	Somali Erer District government Agriculture, Education, Health, Revenue, and Trade Offices
District	Somali	Somali Harshin Agriculture, Education, Health, Revenue, and Trade Offices
District	Somali	Somali Jijiga Zuria District government Agriculture, Education, Health, Revenue, and Trade Offices
District	Somali	Somali Kebri Beyah Agriculture, Education, Health, Revenue, and Trade Offices
District	Tigray	Tigray Erob Agriculture, Education, Health, Revenue, and Trade Offices
District	Tigray	Tigray Gulo Mekeda Agriculture, Education, Health, Revenue, and Trade Offices
District	Tigray	Tigray Hintalo Wajerat Agriculture, Education, Health, Revenue, and Trade Offices
District	Tigray	Tigray Tahtay Koraro Agriculture, Education, Health, Revenue, and Trade Offices
District	Tigray	Tigray Wereilehi Agriculture, Education, Health, Revenue, and Trade Offices

Table A2: Summary Statistics

Means and Standard Deviations

	All
Individual characteristics	
Age	35.34 [8.85]
Male	0.80 [0.40]
Diploma/TVET/Post-High-School	0.17 [0.38]
Undergraduate	0.71 [0.45]
Masters	0.11 [0.31]
Agriculture	0.21 [0.41]
Education	0.20 [0.40]
Health	0.20 [0.40]
Revenue	0.20 [0.40]
Trade	0.20 [0.40]
Years in current position	2.75 [2.64]
Years in current organization	7.32 [7.16]
Years in civil service	13.18 [8.92]
Number of organizations worked in civil service	2.10 [1.81]
Grade	6.10 [2.02]
Director or acting director	0.25 [0.43]
Hours of work in a typical week	41.17 [10.46]
Observations	1776
Organizational characteristics	
Electricity access [Max=8]	4.91 [1.89]
Phone network access [Max=5]	3.50 [1.29]
Internet access [Max=5]	1.55 [1.63]
Computer access [Max=10]	3.40 [2.96]
Vehicle access [Max=10]	0.15 [0.55]
Staff capacity to produce memos [Max=10]	4.99 [3.16]
Staff capacity to produce presentations [Max=10]	3.00 [2.55]
Staff capacity to produce spreadsheets [Max=10]	3.08 [2.67]
Aggregate management practices (WMS) [Max=5]	2.60 [0.55]
Management practice: incentives [Max=5]	2.39 [0.76]
Management practice: targeting [Max=5]	2.94 [0.78]
Management practice: monitoring [Max=5]	3.55 [0.89]
Management practice: roles [Max=5]	2.41 [0.57]
Management practice: flexibility [Max=5]	2.38 [0.70]
Management practice: staff involvement [Max=5]	2.64 [0.66]
Management practice: staffing [Max=5]	2.01 [0.57]
Number of principals	2.08 [0.74]
Observations	380

Notes: Electricity access is the organization average estimate of the number of working hours for which there is

Table A3: Aggregation of Activities

Aggregation	Activity in Survey Module			
	Short Week	Long Week	Short Month	Graphical Week
Meetings	Meetings within the service	Meetings with your direct supervisor Meetings with case team members and change army team members Organisation level meetings (with all staff) Awareness raising meetings on policies, strategies, and others Other meetings	Meetings within the service	Meetings within the service
Working alone	Working alone/on administration	Sending and responding to emails or making phone calls Administrative duties (e.g. preparing a budget) Preparing work plans / planning / capacity building Evaluations or appraisals (staff or activities)	Working alone/on administration	Working alone/on administration
Interfacing	Interfacing with people outside of the civil service (e.g. clients) Interacting with frontline workers	Interacting with clients outside of government (e.g. users) Activities and meetings outside of your mission	Interfacing with people outside of the civil service (e.g. clients)	Interfacing with people outside of the civil service (e.g. clients)
Travelling	Travelling for work	Travelling for work/field visits	Travelling for work	Travelling for work
Without work	Without work to do/waiting for others to input	Personal time (Breaks, down time, coffee, rests, internet browsing) Without any work to do/waiting for others to input	Without work to do/waiting for others to input	Without work to do/waiting for others to input

Table A4A: Randomization Balance Across Individual Characteristics
Means, Standard Deviations, and P-Values from T-Tests of Means

	(1) Short Week	(2) Long Week	(3) Short Month	(4) Graphical Week	P-Values from T-Tests of Means					
	Mean/SE	Mean/SE	Mean/SE	Mean/SE	(1)=(2)	(1)=(3)	(1)=(4)	(2)=(3)	(2)=(4)	(3)=(4)
Age	35.73 [0.43]	35.39 [0.41]	35.20 [0.41]	35.00 [0.43]	0.56	0.37	0.23	0.74	0.51	0.74
Male	0.75 [0.02]	0.82 [0.02]	0.81 [0.02]	0.81 [0.02]	0.01**	0.03**	0.04**	0.79	0.73	0.93
Diploma/TVET/Post-High-School	0.18 [0.02]	0.17 [0.02]	0.19 [0.02]	0.14 [0.02]	0.85	0.56	0.17	0.44	0.23	0.05*
Undergraduate	0.70 [0.02]	0.71 [0.02]	0.71 [0.02]	0.73 [0.02]	0.77	0.82	0.38	0.95	0.54	0.51
Masters	0.11 [0.01]	0.11 [0.01]	0.10 [0.01]	0.12 [0.02]	0.98	0.52	0.55	0.50	0.56	0.22
Agriculture	0.23 [0.02]	0.21 [0.02]	0.20 [0.02]	0.18 [0.02]	0.55	0.43	0.12	0.84	0.33	0.44
Education	0.22 [0.02]	0.20 [0.02]	0.20 [0.02]	0.19 [0.02]	0.49	0.69	0.38	0.78	0.83	0.63
Health	0.19 [0.02]	0.20 [0.02]	0.19 [0.02]	0.22 [0.02]	0.47	0.94	0.19	0.52	0.53	0.21
Revenue	0.18 [0.02]	0.19 [0.02]	0.21 [0.02]	0.20 [0.02]	0.57	0.21	0.39	0.49	0.76	0.72
Trade	0.19 [0.02]	0.20 [0.02]	0.19 [0.02]	0.20 [0.02]	0.96	0.93	0.78	0.89	0.82	0.72
Years in current position	2.65 [0.11]	2.84 [0.13]	2.82 [0.14]	2.69 [0.12]	0.26	0.36	0.80	0.90	0.39	0.51
Years in current organization	7.42 [0.33]	7.59 [0.34]	7.25 [0.34]	6.99 [0.34]	0.73	0.72	0.36	0.48	0.21	0.58
Years in civil service	13.43 [0.43]	13.17 [0.41]	13.19 [0.42]	12.91 [0.44]	0.65	0.68	0.39	0.97	0.67	0.65
Number of organizations worked in civil service	2.20 [0.11]	2.08 [0.07]	2.11 [0.08]	2.00 [0.06]	0.35	0.53	0.13	0.74	0.42	0.27
Grade	6.13 [0.09]	6.15 [0.09]	6.07 [0.10]	6.02 [0.11]	0.91	0.63	0.42	0.55	0.36	0.73
Director or acting director	0.25 [0.02]	0.26 [0.02]	0.26 [0.02]	0.24 [0.02]	0.70	0.79	0.63	0.91	0.39	0.46
Hours of work in a typical week	41.94 [0.51]	41.36 [0.46]	40.70 [0.51]	40.60 [0.50]	0.40	0.09*	0.07*	0.34	0.27	0.89
Observations	459	466	449	402						
F-test of joint significance (p-value)					0.62	0.35	0.10	0.99	0.81	0.82

Notes: * p<0.1; ** p<0.05; *** p<0.01

Table A4B: Randomization Balance Across Work Environment Characteristics

Means, Standard Deviations, and P-Values from T-Tests of Means

	(1) Short Week	(2) Long Week	(3) Short Month	(4) Graphical Week	P-Values from T-Tests of Means					
	Mean/SE	Mean/SE	Mean/SE	Mean/SE	(1)=(2)	(1)=(3)	(1)=(4)	(2)=(3)	(2)=(4)	(3)=(4)
Access to electricity	5.20 [0.09]	5.21 [0.09]	5.25 [0.09]	5.08 [0.10]	0.96	0.71	0.38	0.75	0.35	0.22
Access to phone network	3.62 [0.06]	3.61 [0.06]	3.69 [0.06]	3.73 [0.06]	0.85	0.47	0.19	0.36	0.13	0.58
Access to internet	2.07 [0.08]	2.24 [0.08]	2.03 [0.08]	2.10 [0.09]	0.15	0.69	0.84	0.07*	0.25	0.56
Access to computers	4.81 [0.16]	4.97 [0.17]	4.80 [0.17]	4.83 [0.18]	0.47	0.99	0.90	0.46	0.57	0.89
Access to vehicles	0.20 [0.03]	0.22 [0.03]	0.23 [0.03]	0.28 [0.04]	0.57	0.34	0.08*	0.69	0.20	0.36
Staff capacity to produce memos	5.82 [0.15]	6.11 [0.15]	6.08 [0.16]	6.15 [0.17]	0.18	0.25	0.16	0.87	0.88	0.76
Staff capacity to produce presentations	3.82 [0.13]	4.08 [0.13]	3.94 [0.14]	4.06 [0.15]	0.18	0.56	0.24	0.46	0.92	0.54
Staff capacity to produce presentations	3.99 [0.14]	4.16 [0.14]	4.05 [0.15]	4.32 [0.16]	0.40	0.77	0.12	0.60	0.44	0.22
Aggregate management practices (WMS)	2.73 [0.03]	2.76 [0.03]	2.71 [0.03]	2.75 [0.03]	0.38	0.73	0.54	0.24	0.82	0.37
Management practice: Incentives	2.51 [0.03]	2.54 [0.03]	2.47 [0.04]	2.53 [0.04]	0.56	0.33	0.71	0.12	0.85	0.19
Management practice: Targeting	3.11 [0.04]	3.17 [0.04]	3.08 [0.04]	3.17 [0.04]	0.24	0.66	0.27	0.11	0.99	0.14
Management practice: Monitoring	3.67 [0.04]	3.64 [0.04]	3.58 [0.04]	3.65 [0.05]	0.65	0.15	0.78	0.32	0.87	0.26
Management practice: Roles	2.53 [0.03]	2.57 [0.03]	2.54 [0.03]	2.57 [0.03]	0.28	0.78	0.29	0.45	0.97	0.45
Management practice: Flexibility	2.53 [0.03]	2.61 [0.03]	2.57 [0.03]	2.55 [0.04]	0.07*	0.32	0.69	0.43	0.20	0.59
Management practice: Staff involvement	2.77 [0.03]	2.78 [0.03]	2.76 [0.03]	2.78 [0.03]	0.83	0.83	0.89	0.67	0.95	0.74
Management practice: Staffing	2.08 [0.03]	2.12 [0.03]	2.09 [0.03]	2.10 [0.03]	0.30	0.95	0.59	0.33	0.61	0.64
Number of principals	2.05 [0.03]	2.04 [0.04]	2.00 [0.03]	2.08 [0.04]	0.87	0.34	0.53	0.44	0.44	0.11
Observations	459	466	449	402						
F-test of joint significance (p-value)					0.62	0.66	0.14	0.37	0.23	0.37

Notes: * p<0.1; ** p<0.05; *** p<0.01

Table A4C: Randomization Balance Across Survey Characteristics
Means, Standard Deviations, and P-Values from T-Tests of Means

	(1) Short Week	(2) Long Week	(3) Short Month	(4) Graphical Week	P-Values from T-Tests of Means					
	Mean/SE	Mean/SE	Mean/SE	Mean/SE	(1)=(2)	(1)=(3)	(1)=(4)	(2)=(3)	(2)=(4)	(3)=(4)
Interview conducted in dedicated survey room	0.70 [0.02]	0.69 [0.02]	0.70 [0.02]	0.67 [0.02]	0.84	0.88	0.38	0.73	0.50	0.31
Interview conducted in complete privacy	0.97 [0.01]	0.96 [0.01]	0.94 [0.01]	0.96 [0.01]	0.75	0.09*	0.36	0.17	0.54	0.47
Respondent knowledgeable about own environment	0.48 [0.02]	0.49 [0.02]	0.52 [0.02]	0.50 [0.02]	0.81	0.18	0.45	0.27	0.60	0.59
Respondent knowledgeable about own environment and organization	0.52 [0.02]	0.51 [0.02]	0.48 [0.02]	0.50 [0.02]	0.81	0.18	0.45	0.27	0.60	0.59
Respondent was willing to reveal basic and confidential information	0.76 [0.02]	0.79 [0.02]	0.82 [0.02]	0.79 [0.02]	0.41	0.06*	0.45	0.30	0.96	0.29
Respondent seemed very patient	0.90 [0.01]	0.90 [0.01]	0.92 [0.01]	0.91 [0.01]	0.85	0.24	0.78	0.32	0.92	0.39
Interview perceived to go very well by enumerator	0.71 [0.02]	0.72 [0.02]	0.69 [0.02]	0.69 [0.02]	0.77	0.35	0.46	0.22	0.31	0.86
Observations	459	466	449	402						
F-test of joint significance (p-value)					0.98	0.12	0.76	0.35	0.90	0.73

Notes: * p<0.1; ** p<0.05; *** p<0.01

Table A5A: Randomization Balance Across Individual Characteristics in Diary Sample

Means, Standard Deviations, and P-Values from T-Tests of Means

	(1) Short Week	(2) Long Week	(3) Short Month	(4) Graphical Week	P-Values from T-Tests of Means					
	Mean/SE	Mean/SE	Mean/SE	Mean/SE	(1)=(2)	(1)=(3)	(1)=(4)	(2)=(3)	(2)=(4)	(3)=(4)
Age	38.96 [1.33]	38.23 [1.49]	36.75 [1.53]	36.80 [1.53]	0.72	0.28	0.29	0.49	0.51	0.98
Male	0.69 [0.07]	0.72 [0.07]	0.72 [0.08]	0.74 [0.07]	0.72	0.75	0.60	0.98	0.86	0.85
Diploma/TVET/Post-High-School	0.07 [0.04]	0.05 [0.03]	0.08 [0.05]	0.00 [0.00]	0.75	0.78	0.12	0.56	0.18	0.08*
Undergraduate	0.62 [0.07]	0.53 [0.08]	0.64 [0.08]	0.54 [0.09]	0.37	0.88	0.48	0.32	0.88	0.42
Masters	0.29 [0.07]	0.38 [0.08]	0.28 [0.08]	0.46 [0.09]	0.41	0.91	0.12	0.37	0.48	0.12
Agriculture	0.33 [0.07]	0.20 [0.06]	0.19 [0.07]	0.29 [0.08]	0.17	0.17	0.65	0.95	0.39	0.37
Education	0.18 [0.06]	0.28 [0.07]	0.22 [0.07]	0.17 [0.06]	0.29	0.62	0.94	0.60	0.29	0.60
Health	0.24 [0.06]	0.33 [0.07]	0.25 [0.07]	0.29 [0.08]	0.42	0.95	0.68	0.48	0.72	0.74
Revenue	0.11 [0.05]	0.15 [0.06]	0.19 [0.07]	0.11 [0.05]	0.60	0.30	0.97	0.61	0.66	0.36
Trade	0.13 [0.05]	0.05 [0.03]	0.14 [0.06]	0.14 [0.06]	0.19	0.94	0.90	0.19	0.17	0.96
Years in current position	3.05 [0.46]	3.11 [0.50]	3.51 [0.91]	2.84 [0.47]	0.93	0.64	0.75	0.69	0.70	0.52
Years in current organization	7.42 [0.93]	7.33 [1.21]	6.00 [1.03]	4.74 [0.78]	0.95	0.31	0.04**	0.41	0.09*	0.34
Years in civil service	16.89 [1.32]	14.78 [1.46]	14.44 [1.38]	13.49 [1.62]	0.28	0.21	0.10	0.87	0.55	0.65
Number of organizations worked in civil service	2.40 [0.18]	2.67 [0.38]	2.53 [0.20]	2.26 [0.20]	0.50	0.64	0.60	0.74	0.35	0.34
Grade	7.29 [0.29]	7.03 [0.30]	6.75 [0.42]	6.49 [0.44]	0.53	0.28	0.12	0.59	0.31	0.66
Director or acting director	0.11 [0.05]	0.17 [0.06]	0.14 [0.06]	0.09 [0.05]	0.40	0.71	0.71	0.67	0.26	0.49
Hours of work in a typical week	41.04 [1.10]	42.33 [1.66]	39.44 [1.62]	37.69 [1.56]	0.51	0.40	0.07*	0.22	0.05**	0.44
Observations	45	40	36	35						
F-test of joint significance (p-value)					0.56	0.90	0.51	0.68	0.28	0.78

Notes: * p<0.1; ** p<0.05; *** p<0.01

Table A5B: Randomization Balance Across Work Environment Characteristics in Diary Sample

Means, Standard Deviations, and P-Values from T-Tests of Means

	(1) Short Week	(2) Long Week	(3) Short Month	(4) Graphical Week	P-Values from T-Tests of Means					
	Mean/SE	Mean/SE	Mean/SE	Mean/SE	(1)=(2)	(1)=(3)	(1)=(4)	(2)=(3)	(2)=(4)	(3)=(4)
Access to electricity	5.10 [0.37]	6.12 [0.31]	5.49 [0.42]	5.98 [0.31]	0.04**	0.49	0.08*	0.23	0.76	0.35
Access to phone network	3.65 [0.19]	4.08 [0.16]	3.96 [0.21]	4.00 [0.18]	0.10	0.28	0.21	0.65	0.75	0.89
Access to internet	3.70 [0.13]	3.97 [0.14]	3.79 [0.15]	3.98 [0.13]	0.16	0.64	0.14	0.39	0.97	0.36
Access to computers	9.64 [0.07]	9.55 [0.16]	9.45 [0.18]	9.54 [0.09]	0.62	0.29	0.39	0.65	0.93	0.65
Access to vehicles	0.25 [0.04]	0.20 [0.04]	0.22 [0.05]	0.28 [0.05]	0.42	0.60	0.63	0.82	0.22	0.34
Staff capacity to produce memos	9.67 [0.11]	9.67 [0.13]	9.46 [0.17]	9.54 [0.15]	1.00	0.30	0.50	0.34	0.53	0.72
Staff capacity to produce presentations	7.46 [0.27]	8.12 [0.27]	7.50 [0.35]	7.61 [0.33]	0.09*	0.93	0.72	0.16	0.24	0.81
Staff capacity to produce presentations	8.20 [0.25]	8.27 [0.27]	8.24 [0.30]	7.84 [0.29]	0.85	0.91	0.36	0.95	0.29	0.35
Aggregate management practices (WMS)	0.68 [0.08]	0.81 [0.06]	0.63 [0.09]	0.74 [0.07]	0.21	0.71	0.57	0.12	0.46	0.36
Management practice: Incentives	0.49 [0.08]	0.61 [0.06]	0.40 [0.10]	0.61 [0.06]	0.22	0.49	0.26	0.06*	0.95	0.08*
Management practice: Targeting	0.94 [0.10]	1.10 [0.08]	0.88 [0.12]	1.02 [0.08]	0.21	0.68	0.57	0.11	0.45	0.34
Management practice: Monitoring	0.50 [0.08]	0.63 [0.07]	0.49 [0.10]	0.54 [0.09]	0.25	0.95	0.78	0.26	0.40	0.75
Management practice: Roles	0.74 [0.10]	0.87 [0.07]	0.71 [0.12]	0.79 [0.07]	0.28	0.86	0.71	0.24	0.42	0.59
Management practice: Flexibility	0.86 [0.08]	0.98 [0.06]	0.81 [0.10]	0.93 [0.06]	0.26	0.69	0.52	0.14	0.56	0.30
Management practice: Staff involvement	0.76 [0.08]	0.87 [0.07]	0.68 [0.10]	0.79 [0.08]	0.36	0.54	0.83	0.13	0.46	0.41
Management practice: Staffing	0.36 [0.06]	0.45 [0.06]	0.36 [0.07]	0.41 [0.06]	0.26	0.99	0.53	0.33	0.63	0.59
Number of principals	0 0	0 0	0 0	0 0	0.57	0.89	0.62	0.71	0.71	0.49
Observations	45	40	36	35						
F-test of joint significance (p-value)					0	0	0	0	0	0

Notes: * p<0.1; ** p<0.05; *** p<0.01

Table A5C: Randomization Balance Across Survey Characteristics in Diary Sample
Means, Standard Deviations, and P-Values from T-Tests of Means

	(1) Short Week	(2) Long Week	(3) Short Month	(4) Graphical Week	P-Values from T-Tests of Means					
	Mean/SE	Mean/SE	Mean/SE	Mean/SE	(1)=(2)	(1)=(3)	(1)=(4)	(2)=(3)	(2)=(4)	(3)=(4)
Interview conducted in dedicated survey room	0.84 [0.05]	0.80 [0.06]	0.86 [0.06]	0.74 [0.07]	0.60	0.84	0.27	0.49	0.56	0.22
Interview conducted in complete privacy	0.98 [0.02]	0.97 [0.03]	0.89 [0.05]	0.97 [0.03]	0.93	0.10	0.86	0.13	0.92	0.18
Respondent knowledgeable about own environment	0.44 [0.07]	0.45 [0.08]	0.33 [0.08]	0.34 [0.08]	0.96	0.32	0.36	0.31	0.35	0.93
Respondent knowledgeable about own environment and organization	0.56 [0.07]	0.55 [0.08]	0.67 [0.08]	0.66 [0.08]	0.96	0.32	0.36	0.31	0.35	0.93
Respondent was willing to reveal basic and confidential information	0.69 [0.07]	0.68 [0.07]	0.75 [0.07]	0.69 [0.08]	0.89	0.55	0.98	0.48	0.92	0.55
Respondent seemed very patient	0.84 [0.05]	0.88 [0.05]	0.94 [0.04]	0.97 [0.03]	0.69	0.16	0.06*	0.30	0.13	0.58
Interview perceived to go very well by enumerator	0.69 [0.07]	0.75 [0.07]	0.69 [0.08]	0.91 [0.05]	0.54	0.96	0.01**	0.59	0.06*	0.02**
Observations	45	40	36	35						
F-test of joint significance (p-value)					0.99	0.40	0.20	0.47	0.52	0.14

Notes: * p<0.1; ** p<0.05; *** p<0.01

Table A6: Heterogeneity: Absolute Difference by Individual Characteristics

OLS Estimates and Standard Errors
Robust Standard Errors Reported

	(1) Overall Absolute Difference	(2) Long Week and Education	(3) Short Month and Education	(4) Graphical Week and Education	(5) Long Week and Tenure	(6) Short Month and Tenure	(7) Graphical Week and Education	(8) Long Week and Manager	(9) Short Month and Manager	(10) Graphical Week and Manager	(11) Long Week and Grade	(12) Short Month and Grade	(13) Graphical Week and Grade
Long Week [2]	0.53***	0.28	0.54***	0.53***	0.45**	0.53***	0.54***	0.54***	0.53***	0.53***	0.63	0.53***	0.53***
Short Month [3]	0.00	0.18	0.00	0.00	0.02	0.00	0.00	0.00	0.00	0.00	0.12	0.00	0.00
Graphical Week [4]	-0.045	-0.046	-0.21	-0.045	-0.049	-0.13	-0.038	-0.044	-0.033	-0.051	-0.043	-0.26	-0.041
Age	0.62	0.61	0.47	0.62	0.59	0.49	0.66	0.62	0.73	0.58	0.63	0.27	0.64
Male	0.12	0.13	0.13	0	0.12	0.12	0.31	0.13	0.12	0.13	0.12	0.11	0.37
Diploma/TVET/Post-High-School	0.26	0.25	0.25	0	0.27	0.28	0.10	0.25	0.25	0.23	0.25	0.30	0.21
Undergraduate	0.0037	0.0031	0.0041	0.0037	0.0034	0.0034	0.0012	0.0037	0.0036	0.0037	0.0038	0.0040	0.0035
Masters	0.65	0.71	0.62	0.65	0.68	0.69	0.90	0.65	0.66	0.66	0.64	0.63	0.69
Years in current position	-0.085	-0.078	-0.083	-0.085	-0.086	-0.082	-0.085	-0.086	-0.084	-0.091	-0.085	-0.078	-0.079
Years in current organization	0.33	0.38	0.34	0.33	0.33	0.36	0.34	0.32	0.34	0.31	0.33	0.37	0.37
Years in civil service	0.12	0.055	0.18	0.12	0.11	0.12	0.15	0.12	0.11	0.12	0.12	0.14	0.13
Number of organizations worked in civil service	0.56	0.75	0.42	0.56	0.57	0.56	0.41	0.55	0.60	0.53	0.57	0.49	0.50
Grade	-0.063	-0.18	-0.065	-0.063	-0.078	-0.051	-0.066	-0.058	-0.071	-0.051	-0.047	-0.049	-0.069
Director or acting director	0.72	0.30	0.71	0.72	0.64	0.78	0.67	0.75	0.69	0.76	0.82	0.78	0.67
Hours of work in a typical week	-0.24	-0.36**	-0.24	-0.24	-0.25	-0.24	-0.24	-0.24	-0.25	-0.23	-0.23	-0.23	-0.25
Long Week [2] x Undergrad or Masters	0.17	0.04	0.17	0.17	0.13	0.18	0.13	0.20	0.17	0.17	0.25	0.19	0.13
Short Month [3] x Undergrad or Masters	0.011	0.012	0.012	0.011	0.013	0.0096	0.012	0.011	0.012	0.0096	0.011	0.0091	0.010
Graphical Week [4] x Undergrad or Masters	0.36	0.34	0.34	0.36	0.31	0.48	0.35	0.37	0.36	0.47	0.40	0.48	0.42
Long Week [2] x Years in civil service	-0.0094	-0.0090	-0.010	-0.0094	-0.0097	-0.0094	-0.011	-0.0093	-0.0095	-0.0098	-0.0092	-0.0090	-0.0097
Short Month [3] x Years in civil service	0.31	0.35	0.31	0.31	0.32	0.32	0.24	0.32	0.31	0.29	0.33	0.34	0.30
Graphical Week [4] x Years in civil service	-0.0060	-0.0054	-0.0061	-0.0060	-0.0072	-0.0063	0.00016	-0.0063	-0.0061	-0.0054	-0.0061	-0.0064	-0.0046
Long Week [2] x Manager	0.52	0.56	0.51	0.52	0.45	0.51	0.99	0.50	0.51	0.57	0.51	0.64	0.64
Short Month [3] x Manager	0.0011	0.00084	0.0018	0.0011	0.00038	0.00028	-0.0013	0.0030	0.0012	-0.00042	0.00060	0.0014	0.0036
Graphical Week [4] x Manager	0.96	0.97	0.96	0.96	0.99	0.99	0.96	0.96	0.99	0.96	0.95	0.95	0.88
Long Week [2] x Grade	0.021	0.020	0.018	0.021	0.020	0.019	0.027	0.020	0.022	0.022	0.023	0.012	0.032
Short Month [3] x Grade	0.39	0.41	0.39	0.39	0.40	0.44	0.26	0.42	0.38	0.37	0.35	0.68	0.22
Graphical Week [4] x Grade	-0.020	-0.019	-0.019	-0.020	-0.0098	-0.021	-0.020	0.0014	-0.00057	0.0074	-0.025	-0.034	-0.032
Individual controls	0.87	0.88	0.87	0.87	0.94	0.87	0.87	0.99	1.00	0.96	0.85	0.79	0.79
Work environment controls	-0.0047	-0.0049	-0.0043	-0.0047	-0.0043	-0.0047	-0.0048	-0.0046	-0.0047	-0.0054	-0.0046	-0.0048	-0.0052
Survey controls	0.39	0.38	0.39	0.45	0.43	0.39	0.38	0.40	0.39	0.35	0.40	0.38	0.36
Control Mean		0.26											
Adjusted R-squared		0.19	0.18										
Observations		0.56	0.56	0.12	0.26	0.0055	0.55	0.0057	0.56	-0.013	0.14		
				0.12									
				0.26									
					0.0055								
						0.0057							
							-0.013						
							0.14						
								-0.077					
								0.75					
									-0.083				
									0.70				
										-0.21			
										0.47			
											-0.014		
											0.78		
												0.031	
												0.32	
													-0.038
													0.32
Individual controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Work environment controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Survey controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Control Mean							0.59						
Adjusted R-squared	0.35	0.34	0.34	0.35	0.34	0.34	0.35	0.34	0.34	0.34	0.34	0.35	0.35
Observations	156	156	156	156	156	156	156	156	156	156	156	156	156

Notes: ** p<0.1, * p<0.05, *** p<0.01. Robust standard errors in brackets. The dependent variable in all columns is the absolute difference, across each time-use category, between the time-use diary and the survey response. Individual characteristics include: age, an indicator for whether the respondent is male, an indicator for the highest education qualification of the respondent. Work environment characteristics include the organizational access to electricity, phone networks, internet, computers, vehicles, and skilled staff, and aggregate management practices as measured by the World Management Survey. Survey environment characteristics include: an indicator for whether the interview was conducted in complete privacy; an indicator for whether the respondent was knowledgeable about own environment; an indicator for whether the respondent was knowledgeable about own environment and organization; an indicator for whether the respondent was willing to reveal basic and confidential information; an indicator for whether the respondent seemed very patient; an indicator for whether the interview was perceived to go very well by enumerator. Figures rounded to two significant figures.

Table A7: World Management Survey Indicators

Aggregate Index	Topic	Question	Score 1	Score 3	Score 5
Managemant: Incentives	Monitoring	In what kind of ways does your Directorate track how well it is delivering services? Can you give me an example?	Directorate does not track performance.	Directorate tracks a number of performance indicators. These are seen and reviewed by senior management only.	Full set of indicators are tracked formally and continuously. Reviews are conducted regularly and involve representative of all directorate staff groups. The results of the review are formally communicate to all directorate staff.
Managemant: Incentives	Monitoring	Are you involved in performance review for your Directorate? If so, how often does this occur?	Not involved in performance review;	Bi-annually	Monthly
Managemant: Incentives	Targeting	Does your Directorate have a clear set of targets derived from the organization's goals and objectives? Are they used to determine your work schedule?	The directorate does not have defined targets.	Targets are assigned to the directorate, as well as to the manager and employee levels, and these are generally well understood by mid-level staff. However the tasks assigned to staff are not always related to those targets.	Targets are clearly defined for the directorate, manager, and employee levels, and are well understood by all staff. All tasks are directly derived from the targets, which are regularly reviewed to ensure they remain on track.
Managemant: Incentives	Targeting	When you arrive at work each day, do you and your colleagues know what their individual roles and responsibilities are in achieving the organization's goals?	Staff do not know what their roles and responsibilities are.	Staff have a good idea of their roles and responsibilities but it is not always clear how they contribute to their organization's goals.	Staff have a very good understanding of their roles and responsibilities. Their own roles and goals are clearly interconnected to those of their organization.
Managemant: Incentives	Targeting	How are targets and performance measures communicated to staff in your directorate?	Neither targets nor performance measures are communicated to staff.	Targets and performance measures are formally communicated to managers and team leaders.	Targets and performance measures are formally communicated and understood by all staff.
Managemant: Incentives	Performance incentives	How would under-performance be tolerated in your Directorate? Can you give me an example of how such a case would be dealt with?	Poor performers stay in their positions (no consequences).	Poor performance is identified through evaluation and is addressed through concrete action. Although this applies to most staff, some individuals/staff groups get away with it.	Poor performers are identified through regular reviews and are put on a formal performance improvement plan immediately. This applies to all staff.
Managemant: Incentives	Performance incentives	Given past experience, have members of [respondent's organization] been disciplined for breaking the rules of the civil service?	There are no consequences for bad behaviour/ breaking the rules.	Bad behaviour is addressed through concrete action, but the underlying issues are not addressed.	Bad behaviour/ breaking the rules is addressed through concrete action. If any employee breaks the rules, the underlying issues will be identified and rectified. This applies to all employees.
Managemant: Incentives	Performance incentives	Does your Directorate use performance, targets, or indicators for tracking and rewarding (financially or non-financially) the performance of its employees?	Staff are rewarded equally (or not rewarded) irrespective of performance. Individual performance is not tracked formally	There is a formal staff evaluation system in place and performance is rewarded (financially or non-financially). However, there are no clear system or criteria for rewarding staff	There is a formal staff evaluation system and performance is rewarded (financially or non- financially). Rewards are given as a consequence of well-defined and monitored individual achievements. This applied to all staff.
Managemant: Incentives	Staffing	Do you think the management of your Directorate think about attracting talented people to your Directorate and then doing their best to keep them? For example, by ensuring they are happy and engaged with their work.	Directorate does not put emphasis on talent	Senior management believes that attracting and developing talent is important, but there is no clear system for identifying, attracting or retaining such talent.	Senior management believes that attracting and developing talent is important. There is a clear system for identifying and attracting talent, developing and retaining talent.
Managemant: Incentives	Staffing	If two senior level staff joined your Directorate five years ago and one was much better at their work than the other, would he/she be promoted through the service faster?	No promotion system (no one in the organization has been promoted for years) The promotion system is based on tenure	The promotion system is based on performance. Organization may have internal limitations (e.g. few position openings), but do everything to get around them (e.g. extra training).	Promotion system is based on performance. Organization actively identifies, develops and promotes top performers. Regular assessments, clear set of indicators and personalised career plans for individuals (regularly revised).
Managemant: Autonomy	Roles	When staff in your Directorate are given tasks in their daily work, how much discretion do they have to carry out their assignments? Can you give me an example?	How officers carry out their assignments is decided by senior managers. Officers have no say.	How officers carry out their assignments is jointly decided by the officer and senior managers. Senior managers tend to drive the decisions.	Officers have complete autonomy in deciding how to carry out their tasks.
Managemant: Autonomy	Roles	Can most staff in your Directorate make substantive contributions to the policy formulation and implementation process?	Staff do not contribute to policy formulation, nor to decisions about implementation.	Staff can contribute to policy formulation and decisions about implementation, but there is no formal forum through which to do this. Contributions typically only occur when problems arise.	Management expects all staff to contribute to policy formulation and decisions about implementation (formally or informally), and considers this part of their duties.
Managemant: Autonomy	Roles	Is the workload of achieving your Directorate's targets evenly distributed across its different employees, or do some groups consistently shoulder a greater burden than others?	A small minority of staff undertake the vast majority of work within the directorate.	The burden of the directorate's work is more or less distributed equally among staff. A small minority get away with working significantly less than others.	The burden of the directorate's work is distributed equally among staff. Tasks are assigned in such a way that the amount of time required and the level of difficulty are balanced out so no member of staff finds him/herself overburdened.
Managemant: Autonomy	Roles	Thinking about all the projects that your Directorate has been involved in since your appointment here, would you say that managers and supervisors try to use the right staff for the right job?	Staff are allocated to tasks randomly.	Managers try to use the right staff for the right job but do not go to great lengths to ensure this, or are met with institutional constraints which may prevent them from doing so.	The right staff are always used for a task. Allocation of tasks is based on staffs' documented skills and competencies.
Managemant: Autonomy	Flexibility	Does your Directorate make efforts to adjust to the specific needs and specific requirements of communities, clients, or other stakeholders?	The directorate uses the same procedures no matter what.	The directorate tailors procedures to the specific needs of its stakeholders, but struggles when those needs are complex.	The directorate tailors all procedures to the specific needs of its stakeholders. The evolution of those needs results in adaptation to plans, project and policies.
Managemant: Autonomy	Flexibility	How flexible would you say your Directorate is in terms of responding to new and improved work practices or reforms?	New practices are not adopted/ integrated in the directorate.	New ideas or practices are adopted, but in an informal and/ or isolated manner. The directorate encourages the adoption of new practices, however it is slow to integrate them into its operations (more than a year).	The adoption of new ideas and practices is an integral part of the directorate's work. New practices are regularly reviewed and considered, and once adopted and integrated across the directorate within 6 months.
Managemant: Autonomy	Staff involvement/ contribution	How do problems in your directorate get exposed and fixed?	Ad-hoc. no set process for improvement Deal with problems as they arise without following an established procedure Once fixed, no further action taken No suggestions from staff	Existing process to deal with problems Improvements made through meetings Focus on finding solutions, not prevention of future problems Suggestions from staff involved through meetings (formal or informal)	Exposing problems and suggesting solutions and improvements is part of all staffs' daily duty. Continuous improvement is part of the culture of the organization.
Managemant: Autonomy	Staff involvement/ contribution	What kind of feedback do you get in staff meetings?	No feedback from staff.	Staff provide feedback in meetings but in an unstructured manner. Focus on bad performance.	Staff provide the feedback on which action plans will be based. Focus on both good and bad performance. Details of the meetings are recorded and communicated to all staff.
Managemant: Autonomy	Staff involvement/ contribution	Let's say you've agreed to a follow up plan at one of your meetings, what would happen if the plan wasn't enacted?	No action taken. No changes made in the operations process.	Failure can be found in regular meetings (weekly, even monthly for long-term plans) or at standard points before the deadline. Plans can be altered in order to achieve expected results on time.	In addition to 4, tools can be checked up and reported to the manager in charge. Meetings (formal/ informal) are held to look into the root causes of problems and preventive actions are taken for future similar task.

Table A8: Robustness: Time Use and Service Delivery

OLS Estimates and Standard Errors

Robust Standard Errors Reported

Dependent Variable: Z-Score Across All Organization-Level Service Delivery Indicators (District-Level Only)

	(1) Time Use Only	(2) Management practices	(3) Sector FE	(4) Work environment characteristics and number of principals	(5) Working alone and incentives	(6) Coordination and incentives	(7) Working alone and autonomy	(8) Working alone and number of principals	(9) Coordination and number of principals	(10) Coordination and autonomy	(11) Alternative specification: complementarity	(12) Alternative specification: diminishing returns
Hours working alone (organization average)	0.0055* [0.0031]	0.0024 [0.0034]	0.0033 [0.0036]	0.0028 [0.0036]	0.0036 [0.0036]	0.0028 [0.0036]	0.0026 [0.0037]	0.0028 [0.0036]	-0.019 [0.012]	0.0030 [0.0037]	0.00065 [0.0079]	0.013 [0.013]
Hours coordinating (interfacing and in meetings, organization average)	0.0068** [0.0033]	0.0044 [0.0034]	0.0039 [0.0035]	0.0040 [0.0034]	0.0046 [0.0034]	0.0036 [0.0035]	0.0046 [0.0036]	0.0041 [0.0037]	0.0019 [0.0038]	0.0055 [0.012]	0.0021 [0.0077]	0.0032 [0.014]
Hours other (organization average: travelling and without work, organization average)						Omitted Category						
Management: Incentives	-0.00024 [0.0052]	0.11*** [0.036]	0.13*** [0.037]	0.12*** [0.037]	0.24*** [0.083]	0.19*** [0.074]	0.12*** [0.038]	0.12*** [0.038]	0.14*** [0.038]	0.12*** [0.038]	0.13*** [0.038]	0.13*** [0.038]
Management: Autonomy	-0.000054 [0.0060]	-0.040 [0.036]	-0.046 [0.037]	-0.054 [0.037]	-0.065* [0.038]	-0.046 [0.040]	0.00017 [0.075]	-0.066 [0.083]	-0.057 [0.036]	-0.053 [0.038]	-0.056 [0.038]	-0.058 [0.038]
Number of principals				0.036 [0.030]	0.031 [0.029]	0.034 [0.030]	0.036 [0.029]	0.036 [0.030]	-0.10 [0.083]	0.048 [0.10]	0.036 [0.030]	0.037 [0.029]
Hours working alone x Management: Incentives					-0.0068 [0.0043]							
Hours coordinating x Management: Incentives						-0.0038 [0.0039]						
Hours working alone x Management: Autonomy							-0.0036 [0.0042]					
Hours coordinating x Management: Autonomy								0.00070 [0.0038]				
Hours working alone x Number of principals									0.0097* [0.0049]			
Hours coordinating x Number of principals										-0.00073 [0.0056]		
Hours working alone x Hours coordinating											0.00013 [0.00044]	
Hours working alone*2												-0.00028 [0.00034]
Hours coordinating*2												0.000035 [0.00036]
Sector fixed effects	No	No	Yes	Yes	Yes	Yes	Yes			Yes	Yes	Yes
Work environment characteristics	No	No	No	Yes	Yes	Yes	Yes			Yes	Yes	Yes
Long week option included	No	No	No	No	No	No	No			No	No	No
Adjusted R-squared	0.012	0.037	0.028	0.030	0.034	0.029	0.028	0.027	0.041	0.027	0.027	0.025
Observations	279	279	279	279	279	279	279	279	279	279	279	279

Notes: * p<0.1; ** p<0.05; *** p<0.01. Robust standard errors in brackets. The unit of observation is the district organization. The dependent variable is the z-score of the service delivery index across all sectors in the sample. The sample is restricted to district-level organizations only. Time use measured by the long-week survey option are included. Management: Incentives is the average z-score across incentives, targeting, monitoring, and staffing components of management; Management: Autonomy is the average z-score across the roles, flexibility, and staff involvement components of management. Work environment characteristics include the organizational access to electricity, phone networks, internet, computers, vehicles, and skilled staff. Figures are rounded to two significant figures.

Table A9: Time Use, Motivation and Service Delivery

OLS Estimates and Standard Errors

Robust Standard Errors Reported

Dependent Variable: Z-Score Across All Organization-Level Service Delivery Indicators (District-Level Only)

	(1) Baseline	(2) Average individual characteristics	(3) Working alone and education	(4) Coordination and education	(5) Working alone and experience	(6) Coordination and experience	(7) Working alone and public service motivation	(8) Coordination and public service motivation	(9) Working alone and internality	(10) Coordination and internality
Hours working alone (organization average)	0.0014 [0.0032]	0.00076 [0.0032]	-0.0021 [0.0100]	0.0016 [0.0033]	0.0059 [0.0068]	0.00050 [0.0031]	0.0012 [0.0032]	0.0010 [0.0032]	0.0012 [0.0032]	0.00079 [0.0032]
Hours coordinating (interfacing and in meetings, organization average)	0.0065* [0.0037]	0.0044 [0.0036]	0.0044 [0.0036]	-0.011 [0.0078]	0.0045 [0.0036]	0.018** [0.0088]	0.0042 [0.0036]	0.0047 [0.0037]	0.0043 [0.0037]	0.0039 [0.0036]
Hours other (organization average: travelling and without work, organization average)			Omitted Category							
Management: Incentives	0.13*** [0.037]	0.11*** [0.037]	0.11*** [0.038]	0.12*** [0.037]	0.12*** [0.037]	0.11*** [0.037]	0.12*** [0.038]	0.11*** [0.037]	0.11*** [0.038]	0.11*** [0.038]
Management: Autonomy	-0.053 [0.037]	-0.041 [0.036]	-0.043 [0.037]	-0.042 [0.036]	-0.045 [0.036]	-0.036 [0.036]	-0.042 [0.036]	-0.041 [0.036]	-0.043 [0.036]	-0.041 [0.036]
Number of principals	0.034 [0.029]	0.028 [0.028]	0.027 [0.029]	0.017 [0.029]	0.028 [0.028]	0.024 [0.028]	0.025 [0.029]	0.029 [0.028]	0.029 [0.028]	0.027 [0.028]
Proportion With At Least Undergraduate		0.14* [0.076]	0.084 [0.20]	-0.15 [0.16]	0.14* [0.076]	0.16** [0.076]	0.14* [0.077]	0.14* [0.077]	0.14* [0.076]	0.14* [0.077]
Average Years in The Service		0.016*** [0.0041]	0.016*** [0.0041]	0.015*** [0.0040]	0.023** [0.0099]	0.031*** [0.0041]	0.016*** [0.0041]	0.016*** [0.0041]	0.016*** [0.0041]	0.016*** [0.0041]
Average Public Service Motivation		-0.17*** [0.048]	-0.17*** [0.048]	-0.17*** [0.048]	-0.17*** [0.048]	-0.17*** [0.048]	-0.074 [0.098]	-0.23** [0.10]	-0.17*** [0.048]	-0.17*** [0.048]
Average Internal Locus of Control		0.041 [0.039]	0.040 [0.038]	0.040 [0.038]	0.041 [0.038]	0.045 [0.039]	0.040 [0.039]	0.040 [0.039]	0.090 [0.095]	0.053 [0.079]
Hours working alone x Proportion With At Least Undergraduate			0.0034 [0.011]							
Hours coordinating x Proportion With At Least Undergraduate				0.021** [0.010]						
Hours working alone x Average Years in The Service					-0.00042 [0.00048]					
Hours coordinating x Average Years in The Service						-0.0012* [0.00068]				
Hours working alone x Average Public Service Motivation							-0.0058 [0.0057]			
Hours coordinating x Average Public Service Motivation								0.0044 [0.0066]		
Hours working alone x Average Internal Locus of Control									-0.0028 [0.0046]	
Hours coordinating x Average Internal Locus of Control										0.0028 [0.0055]
Sector fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Work environment characteristics and number of principals	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Long week option included	No	No	No	No	No	No	No	No	No	No
Adjusted R-squared	0.035	0.13	0.13	0.14	0.13	0.14	0.13	0.13	0.13	0.13
Observations	279	277	277	277	277	277	277	277	277	277

Notes: * p<0.1; ** p<0.05; *** p<0.01. Robust standard errors in brackets. The unit of observation is the district organization. The dependent variable is the z-score of the service delivery index across all sectors in the sample. The sample is restricted to district-level organizations only. Time use measured by the long-week survey option are excluded. Management: Incentives is the average z-score across incentives, targeting, monitoring, and staffing components of management; Management: Autonomy is the average z-score across the roles, flexibility, and staff involvement components of management. Work environment characteristics include the organizational access to electricity, phone networks, internet, computers, vehicles, and skilled staff. Average Public Service Motivation is the organization average z-score of the public-service motivation scale as per Perry (1996). Average Internal Locus of Control is the organization average z-score of the internal locus of control scale as per Levenson (1981). The Figures are rounded to two significant figures.

Table A10: Categorizing Management Practices
Correlation Coefficients and Factor Loadings

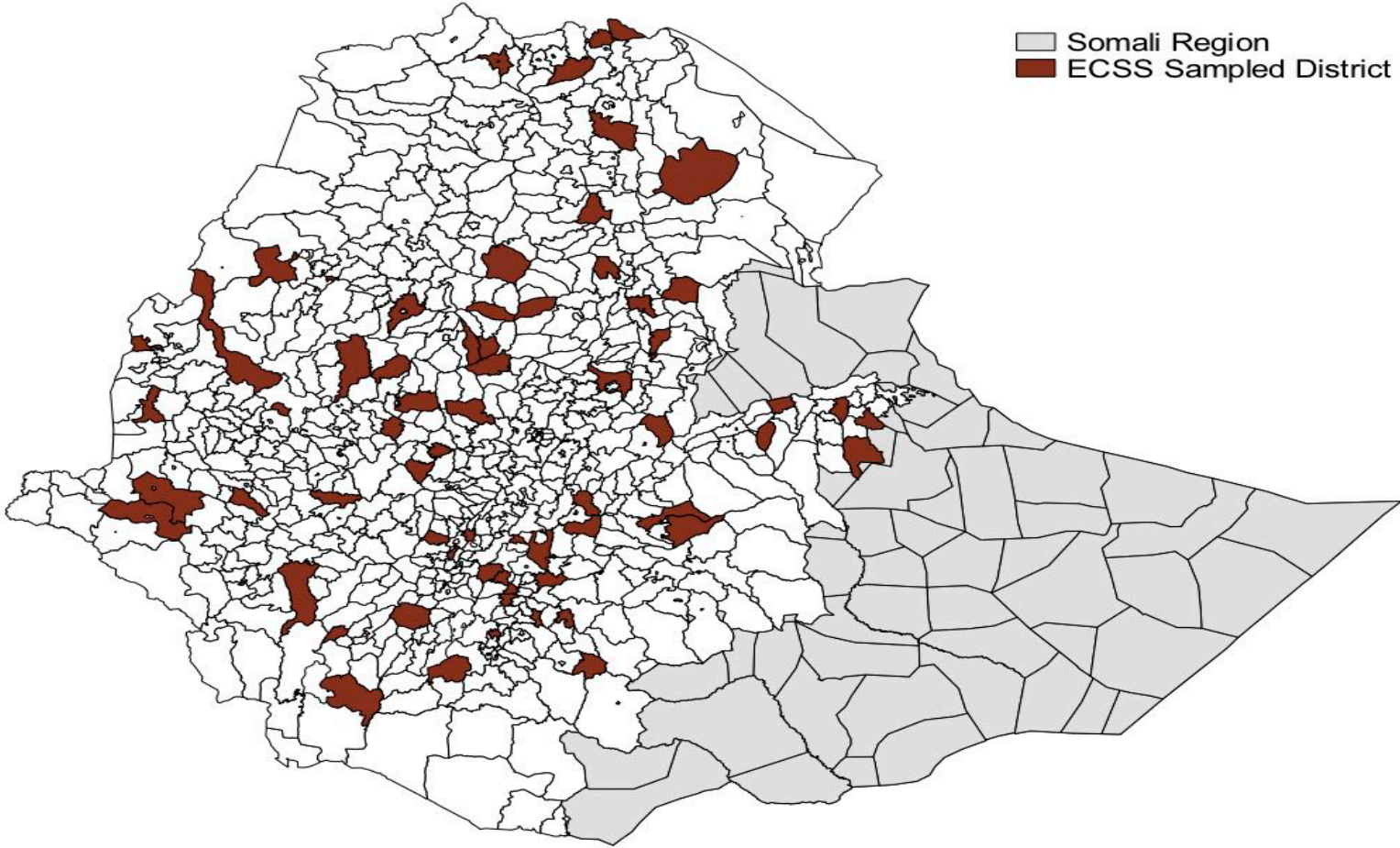
Panel A: Correlation Coefficients							
	(1) Monitoring	(2) Targeting	(3) Performance incentives	(4) Staffing	(5) Roles	(6) Flexibility	(7) Staff involvement and contribution
Monitoring	1						
Targeting	0.60	1					
Performance incentives	0.60	0.60	1				
Staffing	0.40	0.45	0.57	1			
Roles	0.51	0.66	0.67	0.58	1		
Flexibility	0.30	0.43	0.43	0.50	0.55	1	
Staff involvement/contribution	0.58	0.72	0.70	0.53	0.70	0.57	1

Panel B: Factor Analysis			
	(1) Factor 1	(2) Factor 2	(3) Difference [(1)-(2)]
Monitoring	0.68	-0.34	1.0
Targeting	0.79	-0.16	0.95
Performance incentives	0.80	-0.087	0.89
Staffing	0.65	0.19	0.46
Roles	0.82	0.092	0.73
Flexibility	0.63	0.42	0.21
Staff involvement/contribution	0.87	0.0053	0.86

Observations	380
---------------------	-----

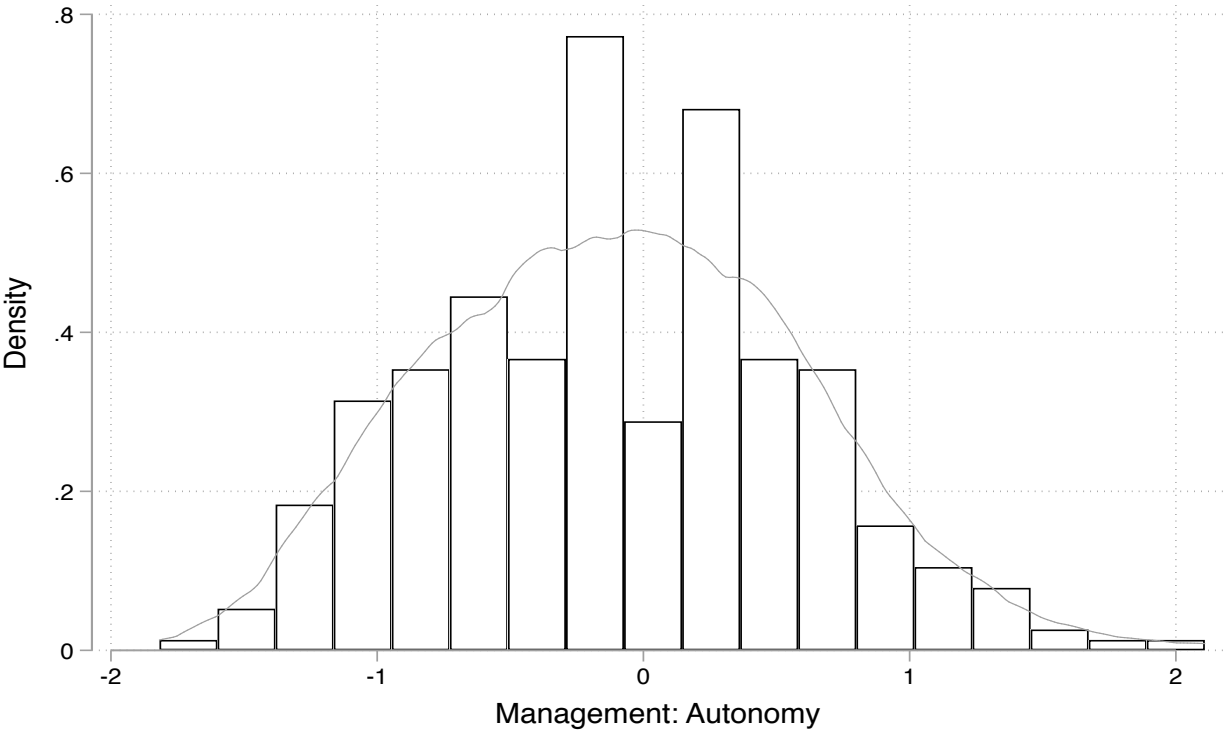
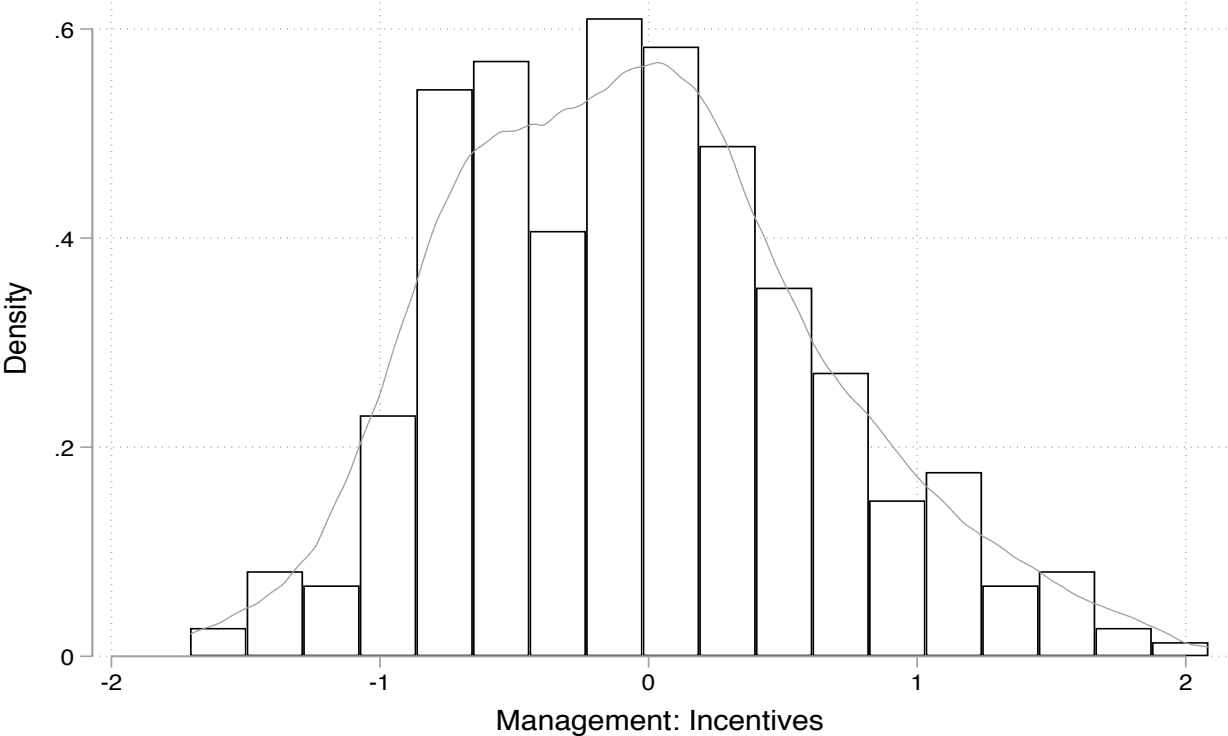
Notes: The top panel table presents correlation coefficients between each of the 7 topics of management practices. Each topic is the mean across the z-score of each item within the topic. The mean of each topic is then constructed at the organization level. The unit of observation is the organization. The bottom panel presents the factor loadings for two factors using a maximum likelihood method and the difference between the loadings for each of the 7 topics. Figures are rounded to two significant figures.

Figure A1: Map of Sampled Districts for Ethiopian Civil Servants Survey



Notes: The figure shows a heat map of districts across Ethiopia based on the Census 2007. The districts shaded in dark red are those sampled in the Ethiopian Civil Servants Survey. The gray areas are districts within the Somali region of Ethiopia, which was mostly excluded from the sampling frame of the survey due to security considerations.

Figure A2: Distribution of Management Practice Indices



Notes: The figure shows the histogram and kernel density function of Management: Incentives (top) and Management: Autonomy (bottom) in z-scores.