Adnan Khan
Asim Khwaja
Benjamin Olken

Property tax experiment in Pakistan

Incentivising tax collection and improving performance

August 2016

Impact Evaluation Report 45

Public sector management

3ie International Initiative for Impact Evaluation

About 3ie

The International Initiative for Impact Evaluation (3ie) is an international grant-making NGO promoting evidence-informed development policies and programmes. We are the global leader in funding, producing and synthesising high-quality evidence of what works, how, why and at what cost. We believe that better and policy-relevant evidence will make development more effective and improve people's lives.

3ie impact evaluations

3ie-supported impact evaluations assess the difference a development intervention has made to social and economic outcomes. 3ie is committed to funding rigorous evaluations that include a theory-based design, use the most appropriate mix of methods to capture outcomes and are useful in complex development contexts.

About this report

3ie accepted the final version of this report, *Property tax experiment in Pakistan: incentivising tax collection and improving performance*, as partial fulfilment of requirements under OW2.178 issued under Open Window 2. The content has been copy-edited and formatted for publication by 3ie. Due to unavoidable constraints at the time of publication, some figures and tables are not optimal. All of the content is the sole responsibility of the authors and does not represent the opinions of 3ie, its donors or its board of commissioners. Any errors and omissions are the sole responsibility of the authors. Any comments or queries should be directed to the corresponding author, Adnan Q Khan at A.Q.Khan@lse.ac.uk

Funding for this impact evaluation was provided by 3ie's donors, which include UK aid, the Bill & Melinda Gates Foundation and the Hewlett Foundation. A complete listing of all of 3ie's donors can be found on the 3ie website.

Suggested citation: Khan, A, Khwaja, A and Olken, B, 2016. *Property tax experiment in Pakistan: incentivising tax collection and improving performance*, *3ie Impact Evaluation Report 45.* New Delhi: International Initiative for Impact Evaluation (3ie)

3ie Impact Evaluation Report Series executive editors: Jyotsna Puri and Beryl Leach
Managing editor: Deepthy Menon
Production manager: Pradeep Singh
Copy editor: Scriptoria
Proofreader: Mathew PJ
Cover design: John F McGill and Akarsh Gupta
Printer: Via Interactive
Cover photo: Abdul Latif Jameel Poverty Action Lab

Property tax experiment in Pakistan: incentivising tax collection and improving performance

Adnan Khan
London School of Economics and Political Science

Asim Khwaja
Harvard University

Benjamin Olken
Massachusetts Institute of Technology

3ie Impact Evaluation Report 45

August 2016

Acknowledgments

We thank the following people for their visionary leadership, insights and support: Muhammad Shahbaz Sharif, Chief Minister; Mujtaba Shuja-ur-Rehman, Minister Excise and Taxation; Javed Aslam (in his role as Chairman, Planning and Development Board and Chief Secretary, Punjab); Nasir Mahmood Khosa (Former Chief Secretary, Punjab); Tariq Bajwa (in his role as Finance Secretary, Punjab); Shamail Ahmad Khwaja, Fawad Hasan Fawad, Zahid Saeed, Shahid Ashraf Tarar and Khalid Masood Chaudhry (Secretaries of Excise and Taxation Department, Punjab during the course of this project); Dr Tauqeer Shah (Principal Secretary, Chief Minister Punjab); Imdad Ullah Bosal, Hamid Yaqoob Sheikh, Faisal Rashid, Mohammed Arshad, Saif Dogar, Socrat Aman Rana and Amjad Saleemi (in their different roles at Finance Department, Punjab); Mujtaba Piracha, Abdullah Khan Sumbal, Saima Saeed and Sadia Tariq (in their roles at Punjab Resource Management Program); Daud Khawaja, Anwar Rashid, Saeed Nawaz, Humayun Mazhar Sheikh, Naseem Sadiq and Ahmad Aziz Tarar (in their role as Director General, E&T Department); Akram Ashraf Gondal and Masood ul-Haq (in their role as Project Director, E&T Department); and Excise & Taxation Directors.

We also thank Ali Abbas, Zahir Ali, Osman Anwar-ul-Haq, Turab Hassan, Jonathan Hill, Kunal Mangal, Zahra Mansoor, Jessica McCann, Shahrukh Nawaz Raja, Obeid Rehman, Adeel Shafqat, Syed Sadaqat Shah, Mahvish Shaukat, Sameea Sheikh, Gabriel Tourek, He Yang, Erol Yayboke and Gabriel Zucker for excellent research assistance. Many thanks to the Technology for People Initiative for creating the Urban Immovable Property Tax (UIPT) Performance Dashboard.

Research discussed in this report has been funded by the International Initiative for Impact Evaluation (3ie) through the Global Development Network, the International Growth Centre and the National Science Foundation. All views and errors are solely ours.

Summary

Pakistan faces important policy challenges in improving service delivery and growth and development. Low levels of tax revenue act as a serious constraint to economic growth, provision of services and, more generally, to building an effective state. Pakistan does poorly on revenue collection, even when compared to other developing countries. To address this problem, in 2009 the Excise and Taxation (E&T) Department in Punjab, Pakistan began implementing a series of human resource (HR) reforms designed to appropriately incentivise tax collection and improve overall departmental performance. The property tax experiment in Punjab involves the design and evaluation of these performance pay packages to increase revenue generation while retaining or raising customer satisfaction and the accuracy of assessments.

Main results

The main results of the experiment were:

- **The incentive schemes produced substantial and unambiguous results on revenue collection.** Treatment circles outperformed control circles by a margin of over 12 percentage points in total collections over the two-year treatment period.
- **Of the three schemes, the Revenue scheme performed best in terms of impact on collections.** In both years, the Revenue scheme consistently had the largest effect and the largest return on investment (ROI). Furthermore, a third-party survey suggests that the E&T Department did not suffer any detectable quality of service costs (either in terms of customer satisfaction or assessment accuracy) as a result of incentivising inspectors.

Policy implications

We feel the points below are noteworthy and should form the basis of a more comprehensive performance-related pay system in the E&T system and potentially other related departments:

- **Performance pay works in raising revenues**. If revenue increase is an important outcome for the government, some form of monetary incentive has to be part of the performance management process for field-level staff.
- **Simpler and objective performance pay schemes perform better**. A key element of an effective performance pay scheme is simple and clear directions that explicitly link to performance on objective dimensions.
- **Performance pay schemes may need to be monitored to ensure customer satisfaction.** A general concern with performance pay schemes that only reward on collections is that they may lead to customer dissatisfaction and over-taxation. While our current findings do not show strong evidence for these concerns, it is recommended that the level of customer satisfaction be monitored regularly.

- **It may be more cost-effective to introduce performance pay 'periods' every few years.** Preliminary evidence suggests that the benefits of performance pay may continue even after the performance pay period is over; such a persistent effect means that it may be more cost-effective for the government to introduce performance pay schemes every few years. The precise length of time between successive performance pay periods should depend on how long it takes the tax base to expand.
- **Performance pay schemes may have to be designed differently for supervisory tiers**. The results of introducing the simplest Revenue scheme (which worked the best for field staff) were not conclusive for supervisory staff; hence further study is required to design an effective supervisory scheme.

Project design

This project was conducted as a randomized controlled trial (RCT). All 482 tax circles in Punjab were randomized into one of three treatment groups or a control group. The treatments were designed to measure the trade-offs that the government would experience in terms of increased revenue versus political costs in terms of dissatisfaction among the public. To our knowledge, this is the first RCT to assess these trade-offs.

The main treatment consisted of three performance pay schemes introduced for tax circle staff:

- **Revenue-based honorarium scheme.** This motivated tax collectors through the use of output-based incentives in combination with Revenue-based honoraria.
- **Revenue plus honorarium scheme.** This was designed to be similar in nature to the Revenue honorarium; however, checks against over-aggressive tax collection were incorporated by factoring in assessment accuracy and taxpayer satisfaction by utilising an objective third-party customer feedback assessment conducted on a randomly selected sample of properties in each tax circle.
- **Conditional fixed wage (flexible bonus) scheme.** In this scheme all circles are guaranteed a small honorarium but the majority of the honorarium is made at the end of the year and is conditional on performance.

In the second year of the study we added an additional control to assess monitoring and awareness effects. In general we treated this Information-only group as a version of a control and included it in the control group:

- **Information-only**. This group was provided information about their performance in the way the Revenue group was, except there was no monetary incentive provided. This information was essentially the information that the inspectors had provided us, but packaged in a way that may be more salient. Inspectors in this group may have felt more watched as well.

Based on feedback from the E&T Department, an additional scheme was added for supervisors:

- **Revenue-based honorarium for AETOs/ETOs scheme.** The revenue scheme was extended to Assistant Excise and Taxation Officers (AETOs) and Excise and Taxation Officers (ETOs) in the second year of the project (Financial Year[1] (FY) 2012–2013).

Data

The project utilized four primary types of data: (i) administrative data kept and maintained by circle-level staff, including ‘Section 9’ registers for newly assessed and re-assessed properties; (ii) HR records detailing inspector transfers and postings; (iii) a phone survey of inspectors measuring effort and monitoring from supervisors; and (iv) an independent survey of over 16,000 properties from all tax circles in Punjab that measured non-revenue outcomes such as customer satisfaction and accuracy of assessment. Administrative data were verified through quarterly checks.

[1]Financial year of the Pakistan government runs from 1 July of every to 30 June of next year.

Contents

List of figures and tables

Abbreviations and acronyms

AETO	Assistant Excise and Taxation Officer
E&T	Excise and Taxation
ETO	Excise and Taxation Officer
GARV	Gross annual rental value
GDP	gross domestic product
GIS	geographical information systems
HR	human resources
PEC	Performance Evaluation Committee
RCT	randomized controlled trial
ROI	return on investment
T&P	transfers and postings
UIPT	urban immovable property tax

1. Introduction

1.1 Background

The property tax experiment in Punjab involves the smart design and evaluation of performance pay packages to increase revenue generation while retaining or raising customer satisfaction and the accuracy of assessments. It began in 2008 when discussions were initiated between the Government of Punjab and the Principal Investigators. Through a rigorous collaborative initial effort, we and the policymakers identified low property tax collection as a pressing policy problem.

Existing relationships with members of its provincial government and a general receptiveness to new schemes made Punjab the ideal target for evaluating our performance pay schemes. Punjab has a long history of property taxation since its introduction in 1958. However, over the years, property tax collection in Punjab has stagnated (see Box 1). Within Pakistan, Punjab's property tax infrastructure is most similar to Sindh's, with Karachi in particular the leader in pushing new reforms such as digitization; Khyber Pakthunkhwa and Baluchistan lag behind the country's most populous provinces.

The property tax is a potentially attractive means of financing municipal government in Pakistan, especially considering its rapid urbanization and the devolution of service functions from the federal and provincial governments. It can provide local governments with access to a broad and expanding tax base, and also promote broader efficiency objectives, linking the provision of municipal services more closely to their financing and strengthening the contract between taxpayers and the state. This provided the context for discussions between Punjab's Excise and Taxation (E&T) Department and ourselves on reforms for increasing property tax collection by designing and evaluating pay-for-performance schemes for property tax collectors.

We used a step-by-step process to jointly diagnose the causal factors behind Punjab's inefficient property tax collection. Both parties recognized that efficient revenue collection is undeniably essential to service delivery, growth and development. Pakistan's tax revenues, however, including property tax, are low relative even to other developing countries. While many factors contribute to this inadequate tax collection, we acknowledged that without a motivated workforce, any structural or legislative reform would have limited success. As in other departments, staff in the E&T Department faced low motivation to perform: monetary and non-monetary incentives were limited and monitoring insufficient. Low motivation and compensation could also create opportunities for tax staff and citizens to collude to reduce taxpayer liability. Examining further, the penalties for underperforming (such as transfers and suspensions) lacked transparency and credibility and were not systematically applied.

After consulting members of all tiers of the E&T Department and other relevant departments, we decided to focus on designing objective and credible ways to build a motivated work force that were rewarded for strong performance and held

accountable for weak performance. We explored ways to design performance pay mechanisms that were conceptually sound and tailored to the department's needs. Given that recent wage increases in other departments had not been carefully evaluated, there was general scepticism that simply raising wages would help; therefore, we developed schemes whereby staff would only be rewarded for above-average performance. Moreover, there was a concern that increased revenue collection alone would not take into account methods of tax collection and their impact on the perceptions of taxpayers. Thus we designed three potential 'schemes'. These solutions ranged from those rewarding employees only in terms of revenue collection to more holistic programs that rewarded them for a range of factors, including taxpayer satisfaction and quality of service.

Box 1. Tax performance in Pakistan

- The E&T Department Punjab is a major revenue-collecting agency for the provincial government. Property tax is its second largest source of income, contributing 28 per- cent of its total revenue (Khan & Inam 2014).
- The tax-GDP (gross domestic product) ratio has been low in Pakistan, even compared to other developing countries, and declined from 13.8 percent in 1996 to 10.9 percent in 2010. This compared to 16.5 percent in Sri Lanka and 14 percent in India (Nabi & Sheikh 2011).
- Provincial tax revenue has remained stagnant, oscillating between 0.4–0.5 percent of GDP over the past 10 years (Nabi & Sheikh 2011).
- Property tax in particular has shown a particularly poor performance and is at roughly one-fifth the level of comparable countries, both in terms of local revenues and as a share of GDP. Many reasons have been cited for low collection, including legislative reform, lack of timely surveys and low motivation of government officials (Ellis *et al*. 2007).

1.1.1 Related literature

Revenue collection and public sector efficiency is a central question in all developing countries, including Pakistan. The literature suggests that developing countries do poorly on revenue collection compared to developed countries. The low level of tax revenues raised in developing countries results in the under-provision of public goods, heightens their vulnerability to economic crises and acts as a serious constraint to growth. Tax revenues as a share of GDP are 45 percent lower in poor countries than in developed countries (Gordon & Li 2009), with an estimated $285 billion per year loss due to tax evasion (Cobham 2005). Past work on this topic attributes the weak performance of tax collection to the corruption and tax evasion that result from low incentives for proper tax collection and administration (Schneider 2007; Friedman *et al*. 2000; Fisman & Wei 2004; Yang 2008).

Unfortunately, when it comes to tax revenues, Pakistan is an outlier even among developing countries. International comparison reveals that the present level of property tax collection in Pakistan is roughly a fifth of that of comparable countries

(World Bank 2006; GHK International 2009) even though property tax rates are similar. A central problem in collecting taxes is effectively incentivising public servants to work within official government systems while not resorting to rent-seeking. Tax administrations lose an estimated $285 billion per year due to tax evasion in developing countries (Cobham 2005). Yet designing and evaluating effective incentives schemes that are politically feasible and sustainable is an extremely challenging task.

1.1.2 Design

Utilising collections, personnel and primary survey data, the scheme designs were put to the test via a randomized controlled trial (RCT). More than 200 tax circles – geographical areas serviced by a set of tax collectors – were chosen for three schemes through public ballot and their performance was tracked over two years. These tax circles are defined internally within the E&T Department, and are generally unrelated to political or administrative units. Changes in the amount of tax collected and the performance of the inspectors in the three schemes were compared against a comparison group that operated under business as usual.

1.1.3 Results

Results of the evaluation indicate that the incentives worked well. Over the study period, treatment circles that received incentives outperformed circles in the comparison group by a margin of over 12 percentage points in terms of total collections, generating significant additional revenue for the government. Furthermore, a third-party survey suggested that the E&T Department did not suffer any detectable quality of service costs (either in terms of customer satisfaction or assessment accuracy) as a result of incentivising inspectors. On average, taxpayers in treatment circles report the same level of satisfaction with the department as taxpayers in non-incentivized circles. In addition, the project has generated significant by-products, including: digitization of the department's historical circle-level collections data; standardization of statement templates; property-level data and tax calculators; HR tracking data; quality checks on departmental data; and the creation of geo-coded circle-level maps that can allow the department to better track its performance. These have also led to the development of a data visualization and management tool, the UIPT Performance Dashboard, to serve as a proof-of-concept pilot that the department can then scale up.

1.1.4 Policy implications

Ultimately, the project has significant policy-level implications. Perhaps most significantly, the study has shown that performance pay *does* work in raising revenue. Additional findings show that simpler and objective performance pay schemes perform better, that it may be more cost-effective to introduce performance pay 'periods' every few years and that schemes should be designed differently for supervisory-level (versus field-level) staff.

1.1.5 Next steps

Though the results of the project have been significant, with clear policy implications, due to smart policy design, we are working with policymakers to refine it further to find the most cost-effective, efficient method of raising tax revenue. While the project provided insights on monetary incentives, it did not cover non-monetary incentives. As a result, we are now evaluating a complementary merit-based transfers and postings (T&P) system with the E&T Department. It has also begun discussing complementary reforms that would study how properly incentivising citizens by tying taxes to services directly can lead to a better tax payment culture and stronger trust in state systems. Together, these ongoing reforms will help inform a broader HR and citizen engagement policy that will have important policy implications for the E&T Department and other departments.

1.2 Uncovering causal factors

This collaborative process was at the heart of the property tax project collaboratively initiated in 2008.The collaboration started by narrowing down the causal factors behind the policy issue in a step-by-step procedure:

- Both parties recognized that efficient revenue collection is undeniably essential to service delivery and growth and development. Pakistan's tax revenues, however, including property tax, are low even relative to other developing countries.
- While many factors contribute to inadequate tax collection, we acknowledged that **without a motivated workforce, any structural or legislative reform would have limited success**.
- As in other departments, staff in the E&T Department faced low motivation to perform: monetary and non-monetary incentives were limited and monitoring insufficient. Low motivation and compensation could also create opportunities for tax staff and citizens to collude to reduce taxpayer liability.
- Examining further, the penalties for underperforming (such as transfers and suspensions) lacked transparency and credibility and were not systematically applied.
- After consulting all tiers of the E&T Department and other relevant departments, we decided to focus on designing objective and credible ways to build a motivated workforce that rewarded personnel for strong performance and held them accountable for weak performance.

1.3 Project timeline

The following provides a brief overview of the timeline of the project implementation starting from the pre-pilot phase.

1.3.1 Phase I

In July 2010, the pre-pilot phase (Phase I) was implemented in 11 circles chosen by a random ballot in Lahore. The pre-pilot was used to iron out kinks in project design,

communication and logistics. It was a success in that regard and established a strong precedent for the subsequent phases of the project. The pre-pilot provided valuable insight for the development of the full intervention. Results suggested that the benefits of this project could be very substantive, which led to the laying of groundwork for the Phase II scale-up.

1.3.2 Phase II pilot

The Phase II pilot began in July 2011 and extended incentive schemes to 161 tax circles chosen by ballot in the province of Punjab. Phase II provided a larger sample through which to scale and refine the pre-pilot results and offered an opportunity to formulate best practices for implementing broader civil service reform within the E&T Department. Consistent results showing gains linked to incentive packages throughout the fiscal year built enthusiasm for extending the schemes.

1.3.3 Phase II extension

In July 2012, the Phase II extension expanded incentive schemes to approximately 60 additional tax circles (for a total of 219 tax circles) and to two additional tiers of supervisory staff (AETOs and ETOs). Gains for participating circles remained consistent and established the reliability of incentives to stimulate sustained revenue generation. Phase II concluded in June 2013 and the final results are presented in this report.

1.3.4 Survey

In May–September 2012, January 2012–February 2013 and June–December 2013, surveys of households within the taxation circles participating in Phase II, as well as control circles, were conducted. The surveys collected data for the evaluation of tax official performance in terms of both customer satisfaction and the accuracy of their property tax liability assessments. These measures provided input for the implementation of some incentive packages as well as insight into the consumer side of property tax administration.

2. Intervention design

Successful policy design addresses the underlying causes that contribute to the policy issue by filtering frontier research knowledge through local contextual realities. This is best done when policymakers and researchers sit at the table together and show a willingness to listen and learn from each other.

The property tax experiment started precisely with such design conversations. In 2009, under the encouragement of the Punjab Resource Management Program, we began a series of discussions with a range of officials at the E&T Department, including the secretary, director general and directors and field staff (tax supervisors and inspectors). We explored possible ways to design performance pay mechanisms that were conceptually sound and tailored to the department’s needs.

There is a large body of literature, covering both development and management, which relates to issues of civil service inefficiency and reform. The case for performance-based honoraria is strongly made in the economic theory of incentives, as these directly link reward to performance and can thus increase effort. Whether it is driven by the presence of high-powered incentives or sustainable means to check corruption, the general concept is that performance pay, accountability and monitoring mechanisms can raise public sector efficiency. One view suggests that problems stem from low-powered incentives and that appropriate reward and penalty mechanisms based on performance are needed.

While wage and incentive schemes have previously been introduced in the public sector in Pakistan and elsewhere, they have rarely if ever been rigorously designed and scientifically evaluated. Our interaction with the Government of Punjab provided us with a rare opportunity to implement and evaluate incentive schemes in the public sector; this allowed us to not only explore key questions of economic importance but also to provide evidence-backed input to help with the design of innovative policy reforms.

The department did not want to judge performance simply by taxes collected, but by the level of service offered to taxpayers and the accuracy of tax assessment. We therefore designed three potential ‘schemes’. These solutions ranged from those rewarding employees only in terms of revenue collection, to more holistic plans that rewarded them for a range of factors, including taxpayer satisfaction and quality of service. Testing these schemes against each other allowed for the comparison of revenue generation with other metrics of performance and helped the department meet its fiscal and service quality goals.

Given that recent wage increases in other departments had not been carefully evaluated, there was also general scepticism that simply raising wages would help. In order to allay these concerns, each of the schemes was designed so that staff would only be rewarded for better performance compared to ‘business as usual’. The schemes established benchmarks based on historical collections, and individuals were rewarded if they performed above these standards. This also ensured that the schemes were cost-effective. Unlike the usual ‘target-based’ systems often used by departments, where staff are rewarded only when they achieve a target, these benchmark-based schemes not only ensured that every staff member felt that they could earn rewards if they worked hard, but also that their incentive to do so would continue even after they had met a particular goal. The project thus aimed to weigh the costs and benefits of each scheme by comparing them robustly.

This section presents the details of the performance pay- schemes introduced over the period starting FY2011-2012 to FY2012-2013. The design for all schemes was informed by international evidence and based on extensive consultation with all stakeholders involved, plus a pre-pilot consisting of 11 tax circles conducted in FY2010-2011.

2.1 Intervention

Below we present the details of each of the main three performance pay-schemes introduced for tax circle staff.

2.1.1 Revenue-based honorarium scheme

A common approach to motivating public servants is the use of output-based incentives in combination with audit mechanisms. Output-based incentives have been shown in many contexts to increase effort, though there is little rigorous evidence on incentives for civil servants. Muralidharan and Sundaraman (2011) and Basinga *et al*. (2010) demonstrated the effectiveness of output-based monetary incentives for government employees in education and health services. For many types of bureaucracy, however, output-based incentives are hard to design because there is no clear output on which to base incentives; such incentives are more feasible in tax collection because there is one clear output that the government cares about – tax revenue. Indeed, tax inspectors are one of the few cases where some governments actually do provide clear output-based incentives.

For the Revenue-based honorarium, the payout was determined by the formula below:

Total pay = baseline (current) salary + Revenue-based honorarium

Revenue-based honorarium = (bonus rate)(increase in revenue over and above benchmark)*

Table 1: Bonus rate cut-offs

Bonus Rate	Bin
40%	0–50th percentile
30%	50th–75th percentile
20%	75th percentile and above

Note: Percentiles were based on FY2009-2010 final net demand in case there was a benchmark change that happened in a circle due to e.g. circle rationalization. New circles selected under the expansion of the schemes were based on final FY2010-2011 net demand.

The **bonus rate** was determined using cut-offs in order to take into account the large differences in circle size, i.e. to promote equity, circles where revenue increases were relatively larger were compensated at a lower rate than smaller circles where it was more difficult to raise revenue. The final cut-offs were determined based on analysis of the Punjab data, and are shown in Table 1. Note that this bonus rate applies at the circle level, such that the total honorarium for each circle is distributed among the circle's tax-collecting staff.

The three members of the tax team in each circle are as follows: an inspector who leads the team, determines tax assessments and issues notices that demand

payment; a clerk responsible for record keeping; and a constable who assists the inspector in the field. The inspector receives 40 percent of the honorarium while the clerk and constable each receive 30 percent. The same split was applied to each of the three performance pay schemes.

Benchmark determination: The benchmarks were generated using historical revenue collection data for current taxes owed and arrears, i.e. each circle was rewarded separately for the collection of current payments due and arrears. In light of detailed discussions regarding benchmarks with the E&T Department, the historical benchmark for FY2011-FY2012 was the three-year average of FY07/08, FY08/09 and FY09/10 plus 10 percent to account for a natural increase. This benchmark was designed to include up-to-date data on collections for each circle. The 10 percent addition to the historical collection corrects in part for the natural rate of increase in tax collections during the period, which historically averaged about 8 per cent per year, while still ensuring that most inspectors will receive the bonus incentives on their marginal collections and not be so far below the benchmark that they will be discouraged from even trying to improve. In Year 2 of the project (FY12/13), benchmarks based on total collection (current + arrears) were used as opposed to distinct benchmarks for current and arrears payments as was the case in the first year of the pilot. This was based on learning from a series of detailed checks conducted by the Project Team which showed that the accuracy of the department data was better for total collections. In addition, the benchmark was updated by taking the three-year average of FY08/09, FY09/10 and FY10/11 plus 20 percent. If we were to incorporate FY11/12 collections into the benchmark, circle-level staff would have an incentive to collect less revenue during the remainder of FY11/12; thus this was avoided.

2.1.2 Revenue Plus honorarium scheme

The Revenue Plus honorarium scheme was designed to be similar in nature to the Revenue-based honorarium, but incorporated checks against over-aggressive tax collection by factoring in assessment accuracy and taxpayer satisfaction by utilizing an objective third-party customer feedback assessment conducted on a randomly selected sample of properties in each tax circle. The scheme recognizes that in addition to increased revenue collection, other components of the tax collection process are important. In addition to the bonus pay based on increase in tax revenue collected, tax personnel may also be held accountable for poor performance in situations such as under- or over-taxing properties (the Revenue scheme may have been particularly sensitive to this possibility) and rewarded in cases where they are doing a better job in terms of assessment accuracy and dealing with taxpayers.

Past research suggests that including a customer feedback component in the study design could potentially have significant beneficial effects for tax revenue and accuracy. For example, Olken (2007) finds that increasing such performance surveys on village road projects in Indonesia reduced ‘missing expenditures’ by a large and significant amount, and concluded that “traditional top-down monitoring can play an important role in reducing rent provision”.

In order to calculate the Revenue Plus honorarium, we objectively quantified performance both in relation to customer satisfaction and accuracy of assessment. In terms of customer satisfaction, we constructed the average scores for each circle from the taxpayer survey on questions regarding customer satisfaction and satisfaction with the outcome in dealing with the E&T Department. Circles were then ranked based on their average scores.

Similarly, for accuracy of assessment we constructed a measure of accuracy by randomly surveying a set of properties selected from the tax registers and independently surveying the property. The percentage difference between the actual assessment of the property and the assessment as judged by the independent survey team was calculated and circles were ranked in terms of the average percentage difference of the sampled properties in that circle.

The scores for customer satisfaction and accuracy were combined for each circle and circles were ranked based on their total score. Three equal-sized groups were constructed: above average, average and below average. Adjustments made to the honorarium are outlined in Table 2.

Table 2: Customer satisfaction and assessment accuracy adjustment – Year 1

Customer satisfaction and assessment accuracy ranking	**Honorarium adjustment**
Above average (top third)	Inspectors: add Rs15,000 per month
	Constables or clerks: add Rs11,500 per month
Average (middle third)	No adjustment
Below average (bottom third)	Inspectors: subtract Rs15,000 per month
	Constables or clerks: subtract Rs11,500 per month

Total pay (split between the circle staff in the 4:3:3 ratio) is then determined by:

Total pay = baseline (current) salary + Revenue-based honorarium +/– Customer Service and Accuracy adjustment

where the revenue-based honorarium part is calculated exactly as in the Revenue-based honorarium scheme.

Note that total pay was never less than the baseline (current salary), nor was any amount already paid out taken back. For example, supposing that the Revenue-based honorarium is less than Rs15,000 per month and the circle is ranked as ‘below average’, they continue to receive the baseline salary, but will just not receive any bonus.

Apart from a change in the way benchmarks were calculated for each circle (as was done in the Revenue scheme) from Year 1 to Year 2, the formula for calculating Revenue Plus honorarium payments remained the same in both years of the pilot.

2.1.3 Conditional fixed wage (flexible bonus) scheme

In the third performance pay scheme, the conditional fixed wage (also referred to as the 'Flexible Bonus') scheme, all circles are guaranteed a small honorarium but the majority of the honorarium is made at the end of the year and is conditional on performance. The scheme aims to mimic the way bonuses work in the private sector, where managers distribute a fixed bonus pool to talented employees based on all the factors they observe. For example, managers might be able to observe effort in addition to outcomes; they also might have information that certain areas were more difficult than others, and could adjust for these factors in ways that would be difficult in an objective, *ex-ante* specialized formulaic incentive system.

Several studies, both empirical and based on economic theory (Becker & Stigler 1974; Besley & McLaren 1993; Mookherjee & Png 1995), suggest that low wages may lead to unmotivated public servants and/or force them to resort to low performance and rent-seeking. The theoretical idea is that high wages act as an incentive, which can be taken away if performance is low. Van Rijckeghem and Weder (2001) showed that wage increases are inversely related to corruption indices through a cross-country study. Regardless of the theory, the idea that raising wages helps reduce rent-seeking is a cornerstone of many tax reform policies throughout the world.

In the first year of the project, each inspector was entitled to an honorarium of Rs30,000/month, and each constable and clerk to Rs23,000/month. Each quarter, 50 per- cent of this payment was made unconditionally, so that in Q1–Q3 each inspector received Rs15,000/month and each constable/clerk received Rs11,500/month. At the end of the fiscal year, the Performance Evaluation Committee (PEC) met to review all circles in the scheme. The PEC criteria used are detailed in Table 3. Staff working in circles deemed to be at the bottom were not paid anything further. Those who worked on circles judged to be in the middle were paid the 50 percent held back and thus ended up earning Rs30,000 (inspectors) or Rs23,000 (constables/clerks) per month. Finally, the top third circle employees were paid an additional bonus and to earn Rs45,000 (inspectors) or Rs34,500 (constables/clerks) per month.

In the second year of the project, while the ranking criteria were analogous, given that the benchmarks were raised for the other two schemes, the guaranteed income was lowered under the scheme. Thus the bottom third staff earned Rs7,000 (inspectors) or Rs5,000 (constables/clerks) a month; the middle third earned Rs22,000 or Rs16,500; and the top third Rs37,000 or Rs28,000. Moreover, unlike the first year, these bonus adjustments were made every six months, instead of just once in the year as was done in Year 1.

It is important to note that this scheme is different from the increase in pay that has been tried in various other departments previously. The main difference is that the

pay increase in previous departments was guaranteed. While these honoraria are also paid partially in advance, the idea is that they are not guaranteed but act like a performance bonus. This allows the department to adjust payments based on various factors that would be difficult to codify in a fixed formula. Thus it provides more flexibility, while at the same time introducing an element of discretion. This may be an advantage, but could also be less effective if there is uncertainty amongst employees as to how the reward will be calculated, which might lower the credibility of the scheme.

Table 3: PEC criteria for conditional fixed wage (flexible bonus) scheme

Category	Total score (99)	Variable	Sub-score
Recovery	39	% recovery against net demand	15
		% increase year-on-year total recovery	15
		Difficulty of circle	9
Demand	25	Net demand current % increase year on year	25
Customer satisfaction	10	Overall satisfaction	10
Directors' rating	10	Directors' rating	10
Accuracy of assessment	15	Quality of accuracy	15

2.2 Supervisory performance pay schemes

Based on feedback from the department, the Revenue scheme was extended to AETOs and ETOs in the second year of the project (FY12/13).

Half the ETOs and the AETOs were selected via a ballotto participate in the scheme. This was done in order to measure the effect of heightened, systematic oversight and support from the next level of department officials (above circle-level staff) on revenue collecting behaviour. This honorarium package was similar to the Revenue honorarium. The pay-out was determined by the formula below, which was computed for the average of the circles under the selected AETO or ETO's charge:

Total pay = baseline (current) salary + Revenue-based honorarium

*Revenue-based honorarium = 0.5 * (bonus rate)*(increase in **average** revenue minus **average** benchmarks)*

Average revenue was calculated as the sum of total revenue for all circles under the charge of a given AETO/ETO divided by the number of such circles. Similarly, average benchmarks were calculated as the sum of total (historical) benchmarks for all circles under the charge of a given AETO/ETO divided by the number of such circles. Each circles' benchmark was calculated exactly as it was for the circle-level staff (the three-year average of FY08/09, FY09/10 and FY10/11 plus 20 percent of this amount).

The **bonus rate,** similar to the revenue scheme, was determined using cut-offs in order to take into account the large differences in average circle size, i.e. to promote equity, circles where revenue increases are relatively larger were compensated at a lower rate than smaller average circles where it was more difficult to raise revenue. The thresholds used were exactly the same as those used for circle staff in the Revenue and Revenue Plus schemes, although instead of using one circle as a measure, average circle sizes under the ETO and AETO were used to determine the bonus rate.

In addition to the above schemes, in the second year we introduced an **Information-only** scheme for tax circles. This provided the same information as in the other performance pay schemes but did not offer any reward. It was simply meant as an alternative comparison group to the business as usual group. This helped ensure that the effects of the performance pay schemes were due to the performance pay component and not to simply having information provided to tax collectors in a specific format.

2.3 Other design considerations

There are several details in the project design that were important to ensure smooth and transparent implementation. These are mentioned briefly below:

- Circle changes and constructing benchmarks: an important part of the revenue and revenue plus schemes was ensuring that benchmarks were fair and appropriate. Benchmarks set too high would not allow staff to earn honoraria even when they worked hard, while those set too low would make the scheme less cost-effective for the government. Basing benchmarks on the historical collections of the circle was therefore deemed to be the fairest and most effective method that took into account each circle's collections and trends. In constructing these benchmarks, however, care had to be taken in both ensuring that changes in circle boundaries were carefully accounted for and that missing data were correctly input. Careful protocols and processes were followed to ensure that any benchmark adjustments were transparent, verified and approved by the PEC.
- Partially filled circles and staff movements: circles where not all three staff (inspector, constable and clerk) were posted received only their share of the honorarium, with the share of the missing staff member reverting back to the government. In addition, staff transfers were permitted with the understanding that any staff member leaving a performance pay scheme circle would no longer be eligible for any honorarium upon departure from that circle. Similarly, any new staff joining a performance pay scheme circle during the year were not eligible to receive an honorarium. These measures were put in place to ensure there was no bias in the selection of staff members in the performance pay schemes.
- Public ballot process: to ensure fairness, transparency and equity, staff were assigned to one of the three performance pay schemes or a control ('business as usual') group based on a public ballot carried out in each of the two years. The ballot was conducted with a large number of circle and supervisory staff

present. In the first year, 150 plus circles were selected into each of the three schemes, and in the second year an additional 60 circles were added (20 for each scheme). These issues are detailed further below.

2.3.1 Circle changes and constructing benchmarks

The project and research team, with feedback from various E&T staff and a senior-level committee from the ministry, ensured that benchmarks were constructed in a fair and transparent manner and were specific to each tax circle. This was done by basing benchmarks on the average revenue collection during FY07/08, FY08/09 and FY09/10. There were several reasons for choosing these years and excluding FY10/11:

- The key to a successful incentive program is how well it motivates employees to work harder. By including the FY10/11 recovery we would have run the risk of de-motivating the staff to work hard for the rest of the fiscal year. This is because lower collection in the current year means a lower benchmark next year, which means more potential money. This is known in the theoretical work on incentive theory as 'the ratchet effect'. The theory stresses that it is critical to avoid ratcheting up benchmarks, as this undermines incentives. Moreover, evidence from other countries shows that if incentive formulae are changed regularly, projects lose credibility and eventually no longer succeed.
- Nevertheless, in order to account for natural changes that occur over time, it had been recommended that the final benchmark be 10 percent higher than the average of the FY07/08, FY08/09 and FY09/10 collections. This was in line with average growth in collections observed in the data, addressed the government's financial constraints and was designed to reward only performance that went beyond 'business as usual'.
- Based on the above considerations, benchmarks for project implementation in FY12/13 were based on the three-year historical average of FY08/09, FY09/10 and FY10/11 inflated by 20 percent. A higher rate for inflating the benchmarks was preferred in Year 2 because results from Phase I suggested that inflating by 10 percent might be too conservative.

As a final consideration, two points are pertinent to note. First, when we encountered data insufficiency whilst estimating historical benchmarks, data were imputed by triangulation using the nearest corresponding years. Second, circles change over time. While compiling the historical collection data, we discovered that circles combine, split and merge in many different ways over time. These boundary changes are generally motivated by a multitude of idiosyncratic concerns; for example, circle boundaries may be redrawn to reallocate properties in rapidly growing areas, or to reapportion responsibility due to staff transfers. In order to generate accurate historical benchmarks for circles that experienced such changes, a more complex calculation was done. For example, if two circles merged to form a larger circle, their benchmarks were added together to form the benchmark for the combined circle. If a circle split into two new circles, then each new circle would get a benchmark that was rescaled in proportion to the relative size (in terms of net demand) of the two circles

(i.e. if the two circles were equal in size then they would both get a benchmark that was half the initial benchmark for the original circle). This general strategy allowed accurate benchmarks to be calculated for circles where merges or splits occurred and also dealt with more complicated cases where circles partially merged with other circles. See Box 2 for the detailed protocol followed in all such cases where benchmarks were changed.

Box 2. Benchmark change protocol

Given the importance of benchmarks, we took great care to ensure that any changes were carefully documented and approved. This was done by developing a protocol to deal with all such cases, including significant property destruction (greater than 5 percent of total demand), updated valuation tables or circle rationalization or extension. The protocol is outlined below:

- Any benchmark change required legitimate and verifiable external validation. The process therefore started with a request made to the project team for benchmark adjustment by the concerned parties. This request needed to be accompanied by relevant documents outlining the reason behind the change. If relevant, circle staff were required to submit relevant tax registers (PT-1, PT-8 etc.) and any external documents (e.g. the land acquisition proceedings by the revenue authorities), and a timeline of change (including dates of when demand changed and how long it took the E&T Department to process changes).
- Following submission, if deemed necessary, an independent field verification process was carried out.
- The PEC then reviewed the case and issued final approval for any benchmarks changes made. In order to avoid unnecessary requests, the PEC was also empowered to exact suitable penalties such as exclusion from the incentive schemes if there was reason to suspect that a request was based on fabricated or false data.

2.3.2 Staff movements and partially staffed circles

The project design accounted for the fact that not all tax circles had the complete team of three staff actually posted, and that over time staff can be transferred. For the former, we ensured that honorarium payments were made based on the number of staff present.

The project allowed for transfers and promotions of circle-level staff throughout the duration of the experiment. In cases where the staff member in question had served for more than 45 days in a specific quarter before being transferred and/or promoted, the honorarium calculation was prorated according to the number of days served in the balloted circle for that specific quarter.

2.3.3 Fair and transparent public balloting

Circles for each of the three schemes were selected through a public ballot ceremony to ensure transparency. Two ballots were held, one at the start of Year 1 and the second at the start of Year 2. Through the first ballot, 150 plus circles were selected for the project (around 50 in each scheme). In the second ballot, 60 additional circles were added to the existing three schemes in the project (20 to each of the three schemes) raising the total number of circles per scheme to approximately 70. Both ballots were attended by senior E&T staff, including the Secretary and the Director General of the E&TDepartment.

At each ceremony, we first conducted mock ballots to show the randomization process and ensure fairness. The final list of selected circles was presented to all participants in the ceremony and communicated to all directorates.

In order to ensure that there was no bias and the scientific evaluation design was not compromized, any staff members who were transferred into a performance pay circle *after* the ballot were not eligible for the honorarium for the given year (they were eligible in Year 2, provided they were still in the same circle). These restrictions were also necessary to avoid gaming of the experiment through manipulating postings. The same policy applied to higher-level staff members – transferring into an area where their predecessor received honorarium did not entitle the new appointee to the honorarium.

The department and the research team also adhered to strict protocols for ensuring that the tax staff included in the project were subjected to standard departmental disciplinary procedures, and that payments were only made to staff members that were eligible under departmental rules.

3. Evaluation design

The property tax experiment put in place detailed and systematic data acquisition, monitoring and evaluation methods. Regular monitoring processes ensured that data were accurately gathered, that any personnel movements and changes in the tax circle boundaries were accounted for, and that the schemes were properly understood and followed. The procedures put in place to gather the two main sources of department data (Recovery and HR) are described below.

To evaluate the impact of the project and identify how well each of the three schemes worked, the gold standard RCT evaluation approach was adopted.

3.1 Data

The project utilized three primary types of data: (i) recovery and collections numbers; (ii) personnel information; and (iii) non-revenue outcomes such as customer satisfaction and assessment accuracy. The first two were gathered using the department's own administrative data and subsequently verified by the project team. The third was gathered using a specially designed third-party survey of more than 16,000 properties from all tax circles in Punjab.

- **Collections data:** The primary source of collections data has been the quarterly circle-level recovery statements verified by ETOs and Directors. Regular channels for quarterly collection and digitization of revenue collection data were put in place to ensure the highest accuracy of data collected. The data were double entered and logical checks were used to eliminate any typographical or data entry errors. Through this process a total of six years of recovery data have been digitized, from FY07/08 to date.

- **Personnel data:** To know which staff members were present in treatment circles and eligible for incentive payments we designed an HR database that is updated every quarter. The database is maintained through an HR verification process whereby regional directors note any personnel movements at the end of each quarter and before payments are processed.

- **Survey data:** A second major source of data was an independent property survey that we conducted. This survey had three main purposes. First, it allowed us to obtain data on people's interactions with the E&T Department, both in terms of their overall perceptions of the quality of this interaction and on corruption. Second, we obtained an independent assessment of the property's characteristics (e.g. land area, covered area and location), which we could use to construct an independent assessment of the property's valuation and compare to the department's official assessments. Third, we could obtain information about the owners and property types, which allowed us to understand whether any observed impact of the schemes varied according to the types of properties and owners. To do so, we surveyed approximately 16,000 properties. Properties were sampled in one of two ways. First, to obtain a random sample of all properties (including those not necessarily on the tax rolls), we created GIS (geographic information system) maps of the circle boundaries for approximately 500 circles in Punjab, and used GIS software to randomly select five points within each circle (for both treatment and control circles). Survey enumerators were given the latitude and longitude coordinates of each point to locate that point on the ground. They then surveyed the property nearest that point, and selected seven more properties nearby (chosen by walking left from the point and choosing every other property) of which an additional four were surveyed based on a randomization table. Once this was completed, we matched these properties to the property-level administrative data to obtain the corresponding administrative records for these properties. We refer to this as our random sample, which comprises 25 properties per circle. Second, since we were particularly interested in properties whose tax valuation had changed as a result of our treatment, we sampled properties directly off the Section 9 registers that are maintained for newly assessed or re-assessed properties. Specifically, we randomly selected 10 properties in each circle among those that had been re-assessed during FY11/12 and FY12/13, and then located these properties in the field and surveyed them. We denote this as the 're-assessed' sample.

3.2 Monitoring procedures

A critical part of the project was setting up proper monitoring procedures to make sure the data gathered were accurate and consistent and that the project was being implemented as planned.

- **Treasury verification and post-payment checks:** To ensure that the department-level recovery data were accurate we compared the statement figures with district treasury totals. The treasury numbers provided an independent verification of the data at the district level. As the treasury data were a district-level verification, to further ensure accuracy we implemented a random audit procedure on collection totals to verify circle-level collection figures. Through this process we found no systematic discrepancies between the administrative data we received and what we found from independent audits. The average difference in collection numbers was less than 1 percent and could be attributed to the expected margin of error inherent in such comparisons.
- **Training and quizzes:** Detailed knowledge on any incentive scheme is essential for its success. Staff members should have a clear understanding of the criteria they will be judged and rewarded on and what actions they can take to improve their incentive payments. To ensure that the schemes were fully understood we also carried out staff trainings at the start of each financial year and before disbursement of each payment round. These trainings were carried out for the three performance pay schemes, the supervisory scheme and the Information-only scheme. Staff were also quizzed after each training session to judge their understanding. The quiz results revealed that shortly after the project started almost all inspectors were able to understand their respective schemes and could calculate their payments. Circle staff received their bonus separately from their base salary on quarterly government cheques, so they not only knew the precise amount of their bonus but also understood how it was calculated.

3.3 Evaluation design

The evaluation design was an RCT. Under this design more than 200 tax circles were chosen for three schemes through public ballot and their performance was tracked over two years. Changes in tax collected and the performance of the inspectors in the three schemes were compared against a comparison group (and in Year 2, an Information-only group as well) that operated under business as usual. Since balloting guaranteed that all groups were statistically equivalent, differences in collections could be attributed to the schemes. The RCT strategy design allows us to interpret any differences between the revenue-based and conditional fixed wage groups and a comparison group as causal.

4. Results

The results indicate that the incentives worked well. Treatment circles outperformed control circles by a margin of over 12 percentage points in total collections, generating an additional Rs240 million in revenue from the 218 circles in which the scheme was implemented over the two-year treatment period.

The Revenue scheme performed best in terms of impact on collections. In both years, the Revenue scheme consistently had the largest effect and ROI of the three schemes. Furthermore, the third-party survey suggested that the E&T Department did not suffer any detectable quality of service costs (either in terms of customer satisfaction or assessment accuracy) as a result of incentivising inspectors. On average, taxpayers in treatment circles report the same level of satisfaction with the department as taxpayers in non-incentivised circles.

The following discussion examines these results in detail. We present the results in chronological order, starting with the pre-pilot conducted in FY11/12 and continuing to the main project conducted over the FY12/13 period. Next, using data from the third-party audit survey, we discuss the effects of the performance pay schemes on taxpayer perceptions of the E&T Department and the accuracy of tax assessments. We conclude by presenting estimates of the ROI for the three schemes and discussing their implications for the department.

Note that the results shown here have already been shared with the department as part of our ongoing engagement. These results form the basis of our discussion for designing a viable policy uptake strategy.

4.1 Impact of the schemes on revenue performance

4.1.1 Pre-pilot results

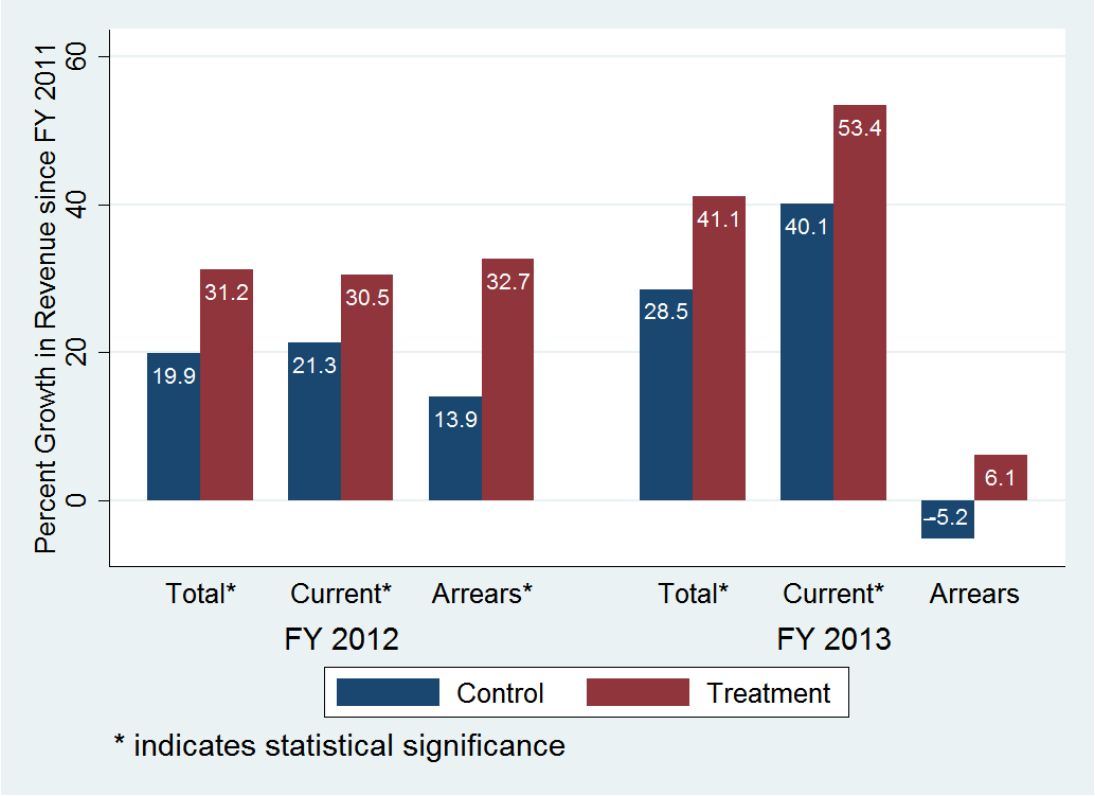
A pre-pilot scheme was conducted in FY11/12 with a small sample of 11 circles in the Lahore divisions. The intent of the pilot was to work out operational hitches, *not* to evaluate the effectiveness of the scheme. Design issues were resolved with the assistance of the High-Powered Committee and department feedback. Nonetheless, the results showed positive returns to incentive schemes. By the end of the fiscal year, treatment circles in the pre-pilot demonstrated on average 10 percent higher collections than control circles. These promising results warranted the subsequent scale-up to scientifically test and evaluate the efficacy of these schemes in detail.

4.1.2 Main project results

The data indicate substantial and unambiguous effects on revenue collection as a result of the incentive schemes. In both years of the project, circles under the performance pay schemes generated larger and statistically significant increases in revenue than control circles. A strong increase in current collections, which account for roughly three quarters of a circle's revenue, drove these results, especially in Year 2. These results are summarized in Figures 1 and 2. The figures show the impact on total recovery, as well as separating the effects on current and arrears

collections. The values shown are year-end outcomes (i.e. as of quarter four) for the three major performance pay schemes: Revenue, Revenue Plus and Flexible Bonus.

Figure 1: Impact of incentive schemes on revenue collection

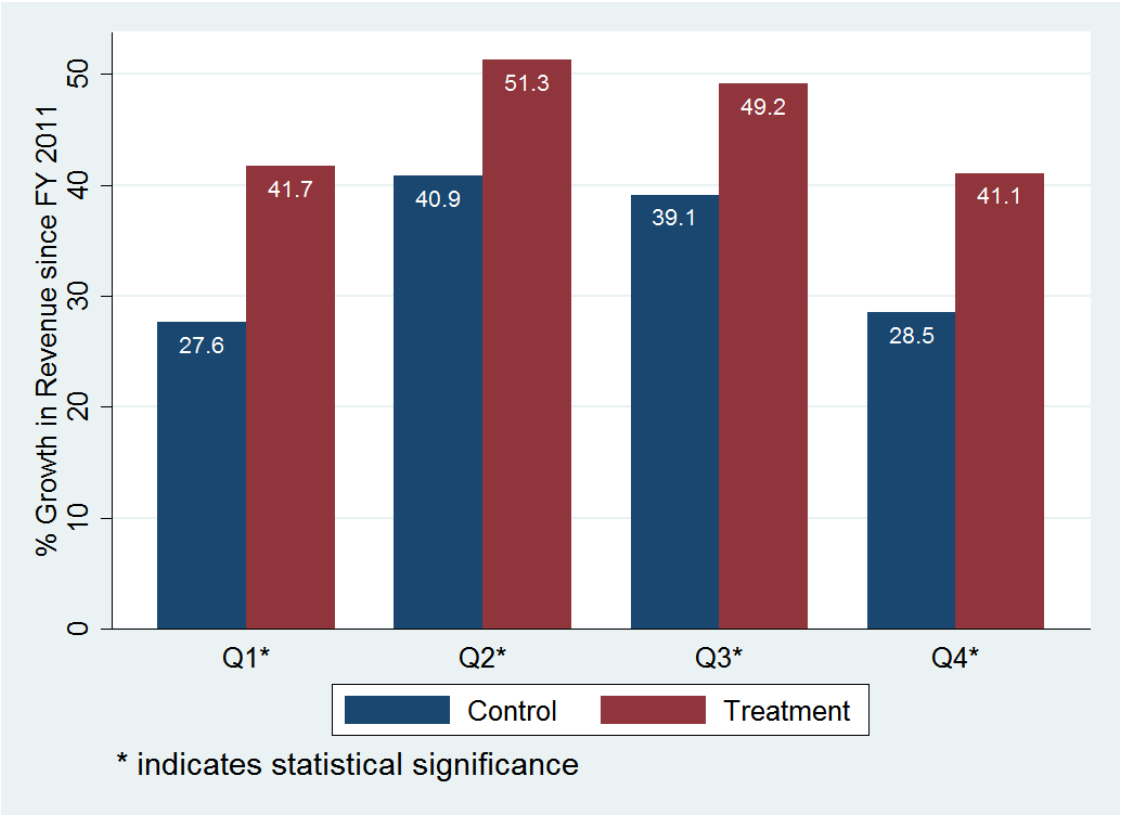


It is worth highlighting that the treatment effect widens in the second year (though the difference is not statistically significant). In terms of current collection, the difference increases from a 9.2 percentage point spread (30.5 percent increase in treatment circles compared to a 21.3 percent increase in control circles) to a 13.3 percentage point spread (53.4 percent increase in treatment circles compared to 40.1 percent in control circles).

The gap in terms of total collection increases as well, though by a smaller margin, due to a drop in arrears collection in the second year. This is to be expected: better performance in year one in both arrears and current collections implies that there will be fewer arrears generated in Year 2 and therefore it is less likely that tax staff will be able to show a significantly improved performance then.

For most of the report we will limit ourselves to showing quarter four results. This is because it does not appear that there are any time-specific patterns to the treatment effect. In Figure 2, we calculate the percentage increase in total recovery quarter by quarter in Year 2 compared with the corresponding quarter in the baseline year. Although there are shifts in the level of growth since baseline over time, the treatment effect (i.e. the difference between treatment and control) remains fairly constant over time, at around 10–14 percentage points.

Figure 2: Year 2 treatment effect on total recovery over time



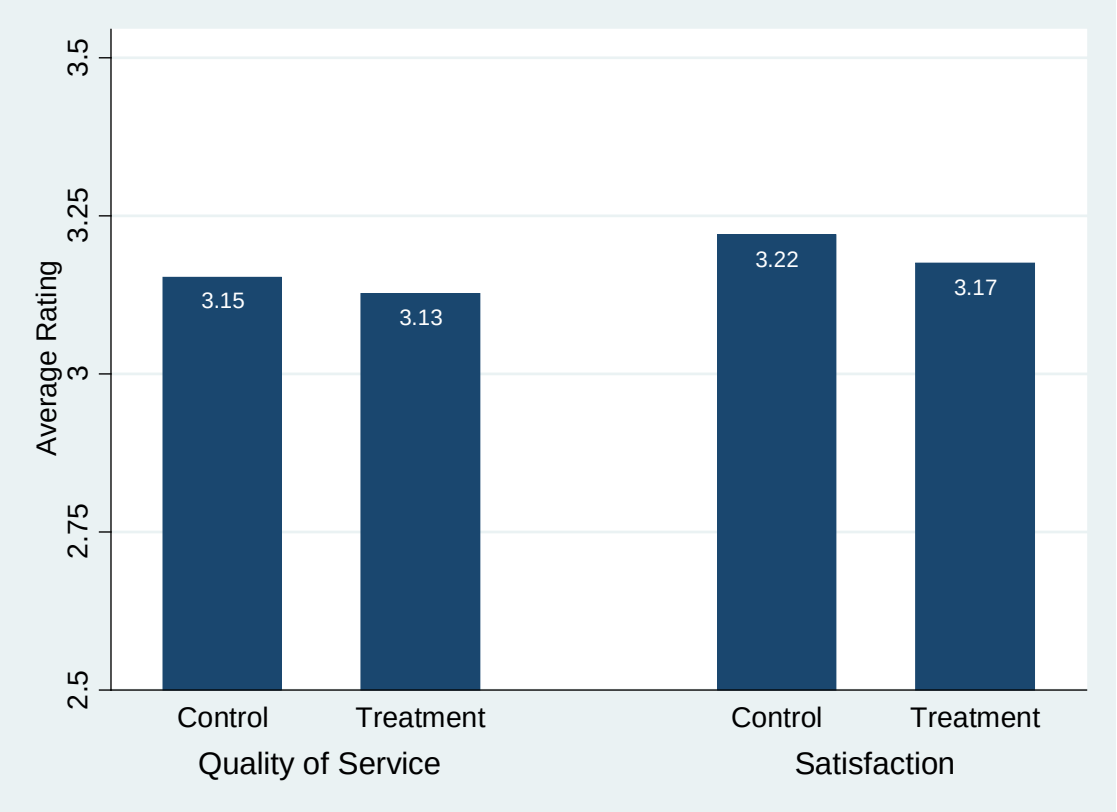
4.2 Impact of the schemes on non-revenue-based performance measures

A potential concern generated under performance pay schemes is that the pressure to collect more taxes may force staff to put excessive pressure on taxpayers. This may lead not only to greater taxpayer dissatisfaction, but also potentially to a worsening of assessment accuracy (if over-taxation occurs). While the Revenue Plus and Flexible Bonus schemes were designed to explicitly address this issue, we first examine here whether the three schemes overall had any adverse impact in such outcomes. We do so by looking at the satisfaction measures in the survey and creating an accuracy-of-assessment measure by comparing the tax assessment generated by the third-party survey with that reported by the tax staff in the official registers. As Figures 3 and 4 show, there seems to be little evidence of any substantial adverse effects on either of these measures.

4.2.1 Taxpayer satisfaction

Taxpayers in the survey were asked to rank the E&T Department separately in terms of satisfaction with the department and its quality of service on a 1–5 scale, with 1 (very poor) being the lowest rating and 5 (very good) being the highest rating; a score of 3 indicates average performance.

Figure 3: Quality and satisfaction



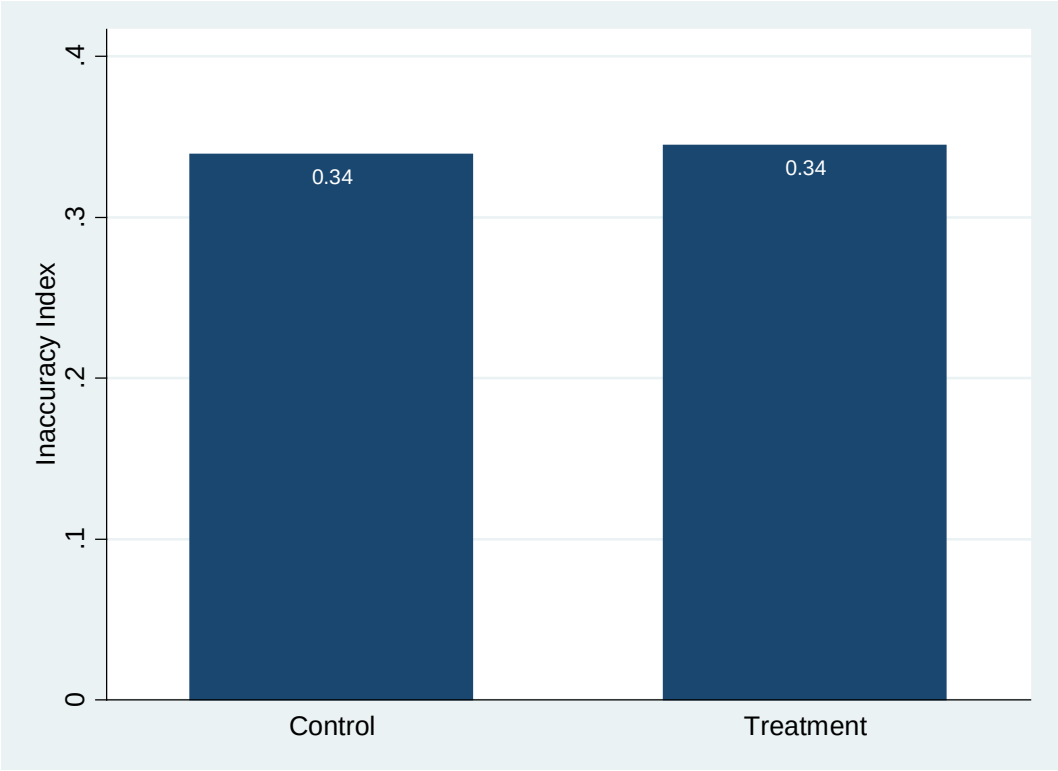
On average, taxpayers in the treatment and control circles reported very similar levels of satisfaction and quality of service. Although the average was slightly less in treatment circles, this difference was not statistically significant; therefore, from an empirical viewpoint, these measures are regarded as reflecting the same level between the two groups.

4.2.2 Tax assessment accuracy

These results were consistent with the lack of effects we found on the accuracy of assessment. Using the third-party survey as a benchmark, we found that accuracy of assessment in treatment circles was statistically equivalent to the accuracy of assessment in control circles. Inaccuracy is measured with an index constructed by comparing the gross annual rental value (GARV) from the third-party survey with the GARV from the official tax register. Higher values of the index correspond with more inaccuracy.[2]

[2]We should caution that the inaccuracy measure should not be interpreted literally since there is genuine measurement error in both the third-party survey and department data. The index is primarily therefore to be interpreted as a scale whose increasing values indicate an increase in inaccuracy but whose level value is less readily interpreted.

Figure 4: Inaccuracy

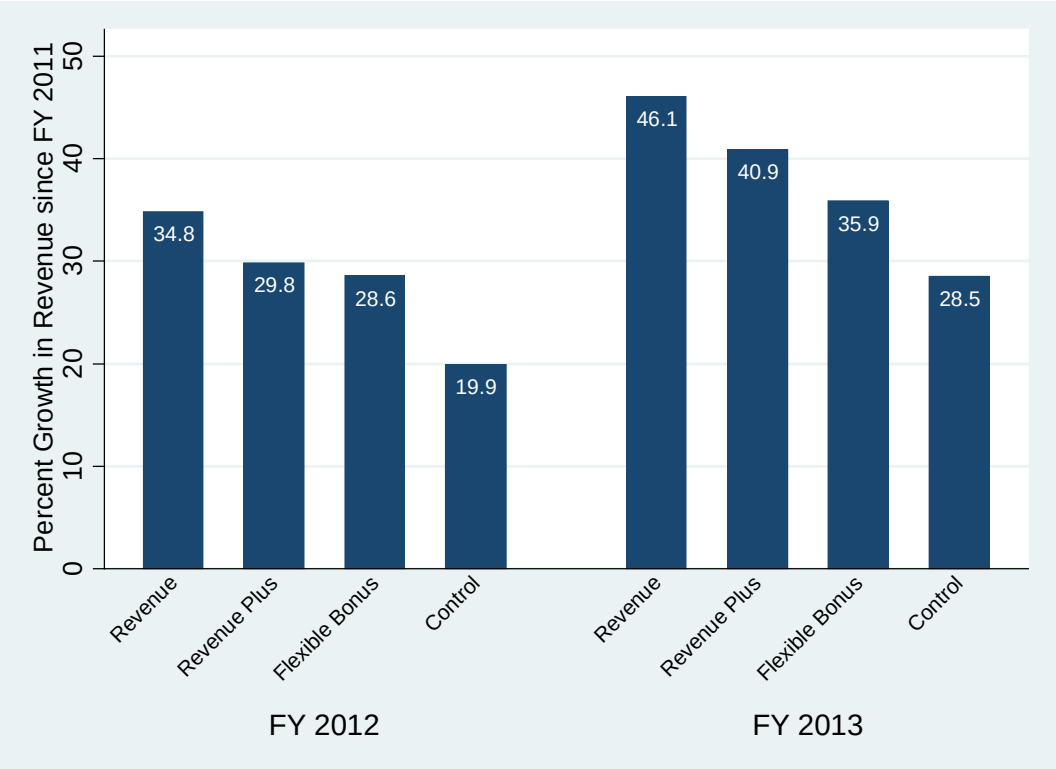


4.3 Separating impact on revenue and non-revenue outcomes for the different schemes

We now look at the relative performance of each of the three different types of performance pay schemes. The results show that the Revenue scheme was the most successful in generating additional revenue for the department, followed by Revenue Plus and lastly the Flexible Bonus (conditional fixed wage) scheme. In terms of non-revenue outcomes, there were no substantial differences between the schemes, and therefore it appears that the Revenue scheme is the best in terms of overall (revenue and non-revenue) outcomes as well.

In both years of the project we see a consistent pattern: the Revenue scheme has the best performance, followed by Revenue Plus and then the Flexible Bonus scheme. While in general the *differences* between schemes are not statistically significant it is so for current collection. In addition we do not find any statistically significant impact from the Information-only scheme and therefore for the analysis we include it as part of the control group.

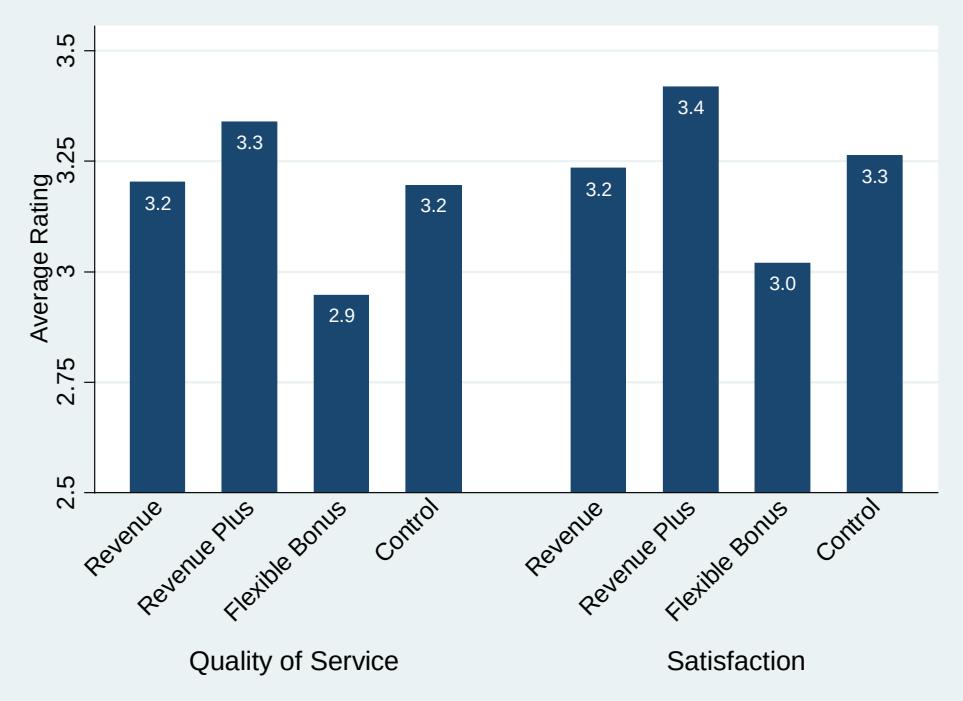
Figure 5: Impacts on total revenue collection by scheme



When we examined whether the increased revenue collected for each scheme was a result of an increase in the tax base (as it was for the schemes together), we found that while that holds for the Revenue Plus and Flexible Bonus schemes, for the Revenue scheme it appears that the increase was generated due to *both* an increase in the tax base and a higher recovery rate. (See Figure 5) On further examination, we found that this increase in recovery rate through the Revenue scheme seemed to be due entirely to increases in the recovery rate for arrears collection. It appears that as the stock of uncollected arrears decreased, tax staff became more assiduous in collecting the remaining arrears in Year 2, leading to a substantial increase in the recovery rate.

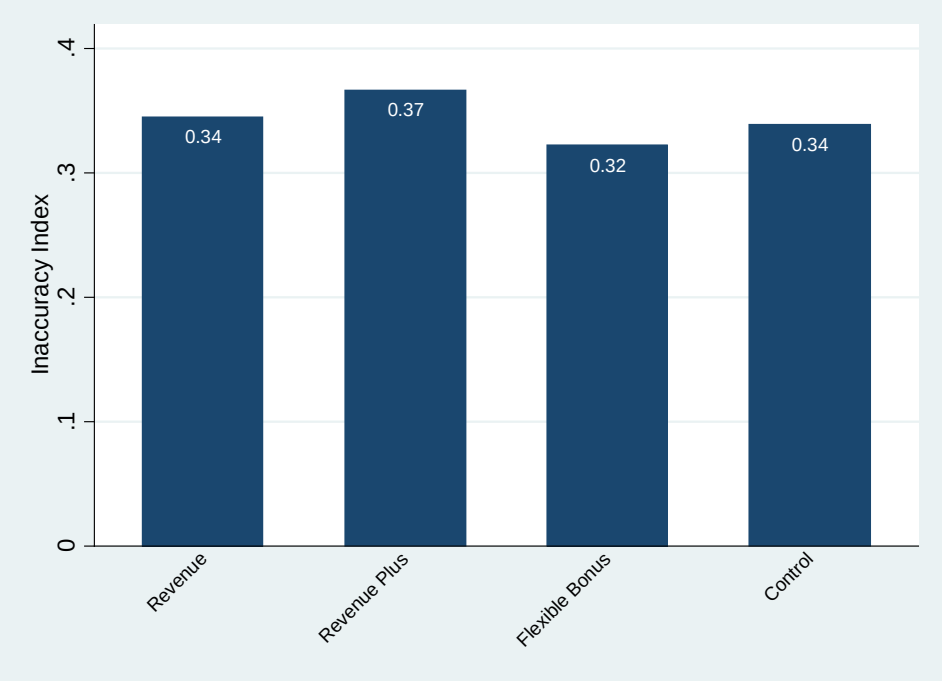
Finally, as before, we now examine the separate impact of each scheme on non-revenue outcomes. While we found that customer satisfaction and quality of service measures were not statistically any worse for the schemes together, when we split them up we found this was only the case for the Revenue and Revenue Plus schemes; the Flexible Bonus scheme appeared to have performed poorly in terms of these measures. This is shown in Figure 6. While the differences are not very large, it does suggest that the Flexible Bonus scheme did the worst both in terms of increase in revenue (which was positive, but by the smallest margin out of the three schemes) and these satisfaction and quality of service measures.

Figure 6: Quality and satisfaction by scheme



Furthermore, it does not appear that the schemes differ significantly in terms of their effects on accuracy of assessment. Using the same measure of inaccuracy as before, we found only minor deviations between the schemes and the control group. None of the differences were statistically significant (see Figure 7).

Figure 7: Inaccuracy by scheme



4.4 Analysing effects and understanding channels

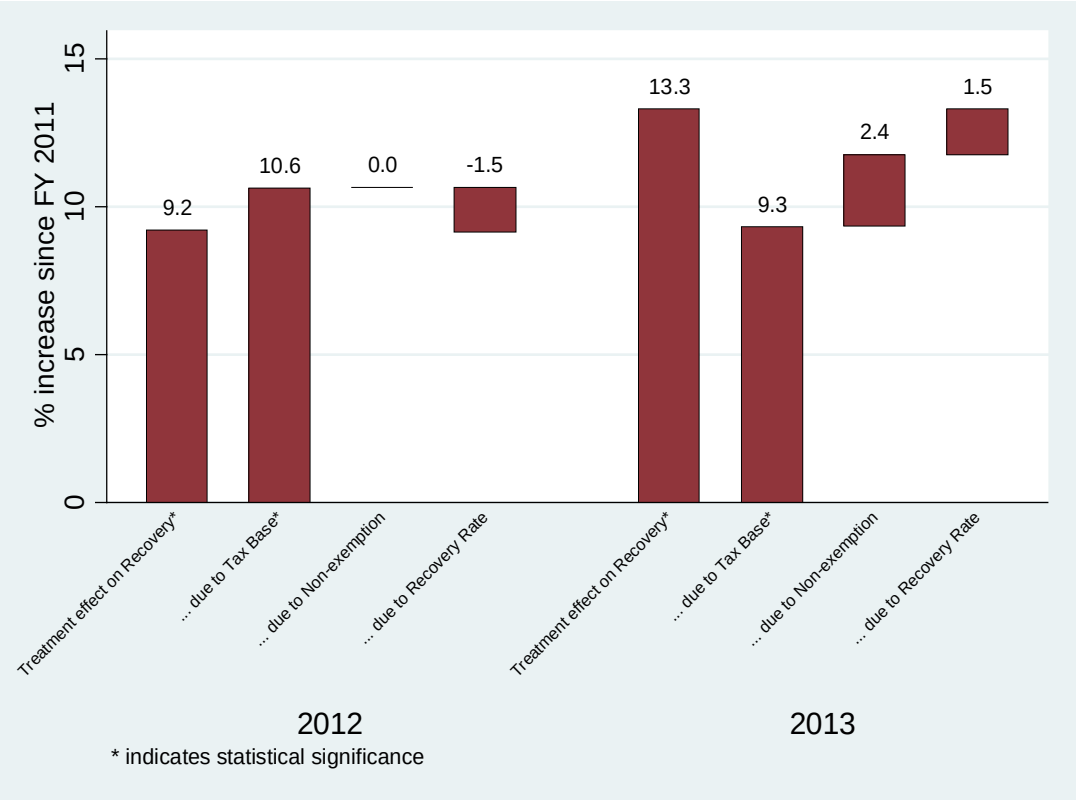
4.4.1 Analysing the impact on collections – tax base, exemptions and recovery rates

Tax staff have several ways of increasing tax revenues: they can increase the tax base, ensure that no excess exemptions are granted or increase the recovery rate. Examining the results in more depth, we found that the increase in collection generated through the performance pay schemes operated primarily through the tax staff increasing the tax base. This is quite promising, since increasing the tax base is likely to have longer-term benefits as well.

Figure 8 shows how the main treatment effect in each year (for current collections) can be separated into these three margins.[3] The tax base is measured by gross demand, the non-exemption rate is calculated as the ratio of net demand to gross demand (i.e. if no properties were exempted, the ratio of these values would be close to one) and the recovery rate is calculated as the ratio of recovery to net demand. We first present the overall effect again (in the first column). Thus 9.2 percentage points is the difference between the control and treatment groups in Figure I (FY2012–2013 current collections). Then in the next three columns we show how much of this effect comes from the three margins (the effect in these three columns can be added up to make the overall effect). As the figure shows, in both years most of the effect comes from an increase in the tax base (i.e. gross demand). Note that the numbers may not add up exactly due to rounding.

[3] Because recovery is the product of tax base, the non-exemption rate and the recovery rate, each of these margins is additively separable in logs. To ensure that the regression coefficients added up, we used a common set of controls for each of these specifications. But since the controls differ, the treatment is slightly different as well. To maintain consistency, we use the level obtained from the primary specification (in Figure I), and divide it proportional to the effects (in logs) we find in the analysis.

Figure 8: Margins of impact on current collections

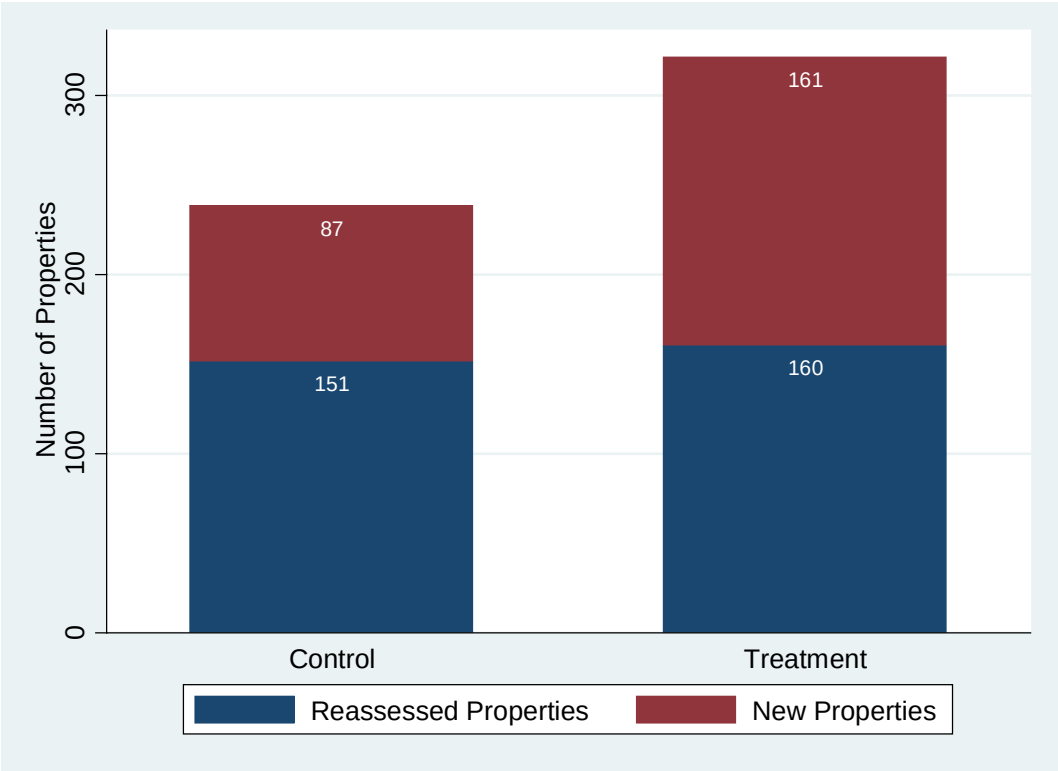


4.4.2 Understanding channels

This increase in the tax base was due to both new properties being added to the registry and revaluations of existing properties. The data indicate that increased collections are a result of effort on both fronts. In particular, we see a substantial increase in the number of properties added to Section 9 registers under treatment. The section register captures properties that either see a change in assessed value or are added to the tax rolls for the first time.

This is demonstrated in Figure 9, which shows the types of properties that were added to the Section 9 register for both the performance pay circles (the treatment group) and the business-as-usual circles (the control group). The analysis reveals interesting heterogeneity. We find that inspectors in treatment and control circles add re-assessed properties (i.e. re-valued properties that are already listed in the tax registers) at roughly the same rate (160 versus 151 properties added over the two-year treatment period, respectively), but inspectors in treatment circles add many more new properties. Over the two years of treatment, inspectors in treatment circles added an average of 161 new properties, compared to only 87 for inspectors in control circles. This represents an almost 100 percent increase in the rate of new property addition. This is, again, very encouraging, since the induction of new taxpayers into the system suggests that these effects are most likely to continue in future years.

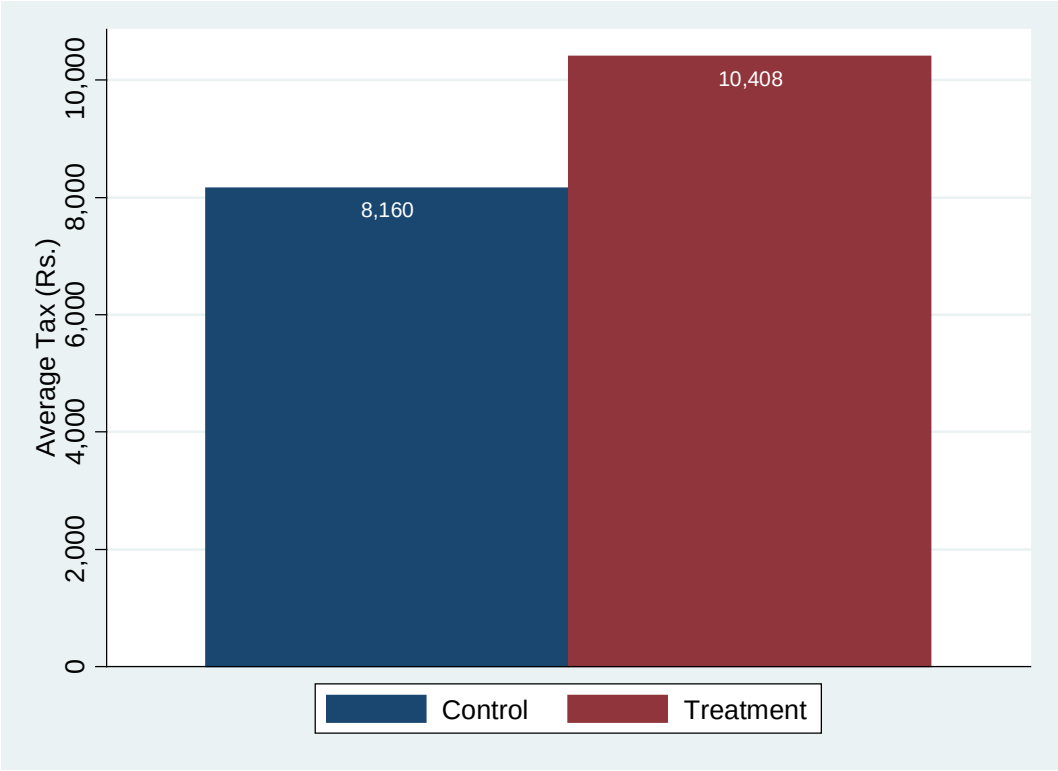
Figure 9: Treatment impact on number of properties added to Section 9 registers



While there is an increase in the number of properties added to the tax rolls or reassessed due to the incentive schemes, this is by itself not a large increase (around 8 per cent more properties over two years), so the question is whether this change can account for the total revenue increase observed.

Further examination of the data shows that in addition to adding more properties, inspectors in treatment circles also collected more tax from the properties they added to Section 9 registers. On average, a property added to Section 9 in treatment circles paid Rs2,248 more tax than a Section 9 property in a control circle. (see Figure 10)

Figure 10: Increased tax payments for Section 9 properties



The combination of increases in new properties and increased collections from Section 9 properties is sufficient to yield the overall treatment effect we observe in the administrative data.

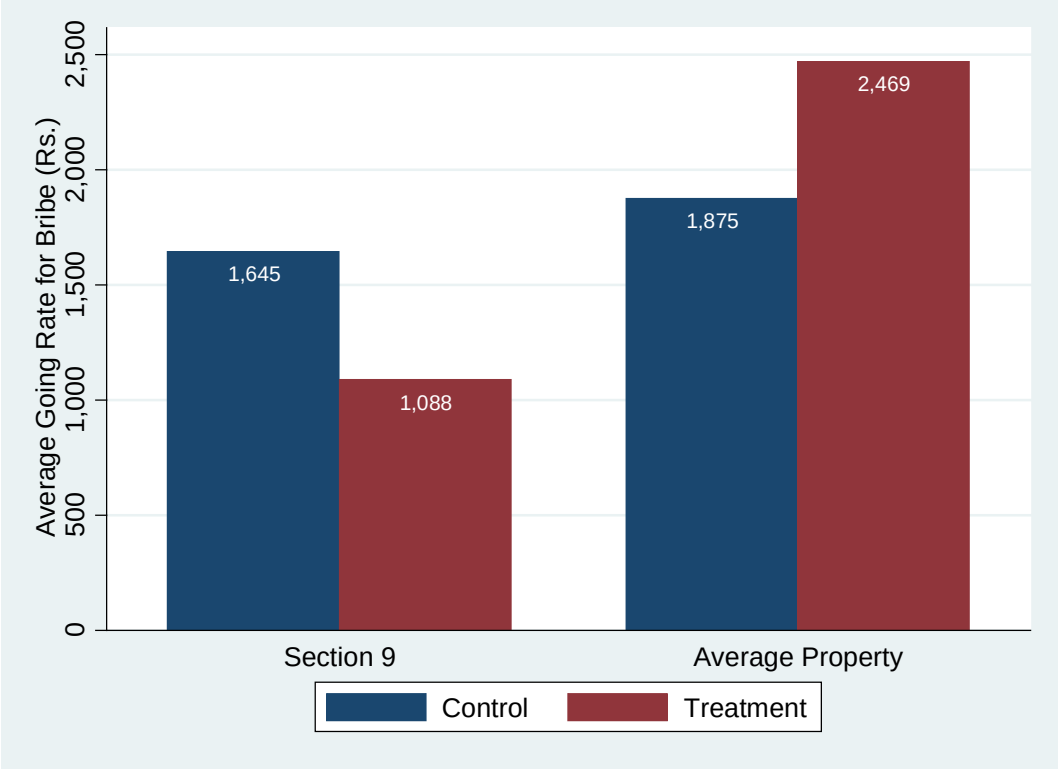
While these results are encouraging, they raise a broader question of whether there is heterogeneity in how the incentivised inspectors treat the properties under their charge. Our results so far clearly suggest that not all properties pay higher tax. While this could be due to the remaining properties already paying their full tax, it is also possible that these properties renegotiated non-payment of taxes. Anecdotal evidence suggests that there is widespread collusion between taxpayers and tax staff, where both parties agree on paying lower tax than is rightfully due. In this context, while tax payments will increase in cases where the greater incentives for tax collection for the tax staff lead to such collusion breaking down, this could also lead to a renegotiation. In case such re-negotiation happens, one will not see such properties paying more tax – instead they will report an increase in bribes paid to the tax collectors. There is also evidence to suggest that this may have happened. The results thus show that a substitution effect among tax collectors – and not increased diligence in collection – is responsible for the revenue increase, as collectors will skim off less due to the promise of greater formal returns.

While it is not feasible to directly ask whether a taxpayer paid any bribes to tax staff, in the taxpayer survey we did ask more generally about what the bribe rate would be for a property similar to the taxpayer's. This form of the question was more palatable to respondents, who otherwise shied away from talking about bribes. The

disadvantage, however, is that the magnitude of bribe payments cannot be compared directly to tax payments. This is because we do not know whether the respondent is reporting on behalf of herself or others. Nonetheless, the measure provides an indication of whether the *average* level of bribes in a given locality is increasing or not.

Figure 11 presents the results on bribe payments, separated by Section 9 status. The results reveal an interesting heterogeneity of response to treatment. For the average property, performance pay incentives increase the going rate of bribes, but decrease the going rate for Section 9 properties. The latter represent properties that have been reassessed and therefore it is likely that the collusion between the taxpayer and tax staff has broken down.

Figure 11: Patterns of bribe payments for the average property and Section 9 properties



Together these results suggest that in settings such as taxation where there may be prevalent collusion between taxpayers and tax collectors, strengthening the bargaining position of the tax collector (which is effectively what performance pay does) can have two distinct effects: a more desirable one where the collusion breaks down and the taxpayer pays the full amount of tax, or a less desirable one where the tax paid does not change but the taxpayer simply pays a higher bribe in order to remain in the collusive agreement. In our case we see both effects at play but the desirable one is large enough to generate overall gains. It does suggest, however, that one may want to have concurrent policies (such as audits, etc.) that minimise the less desirable cases.

4.5 Impact of the supervisory performance pay scheme

While in the second year of the project, supervisory (ETO and AETO) performance pay schemes were introduced, and we expressed a concern that the small number of such positions available (only 51 as opposed to the approximately 500 tax circles) may imply that the technical research design would not have a large enough sample size to be able to statistically detect any impact of this scheme. This was made even more challenging by the fact that some of the supervisors were transferred after the ballot. Therefore, as per the technical requirements of the evaluation design, they were no longer eligible for an honorarium.

Due to these technical limitations, the results of the project remain inconclusive with regard to the impact of introducing performance pay incentives for ETOs and AETOs. If anything, the evidence would suggest that introducing performance pay that rewards only on revenue collection is less effective at this level than at the circle-staff level. Based purely on the technical results, however, the more conservative conclusion would be that the results are inconclusive at this stage and may require further study.

4.6 Return on investment – overall economic impact of the schemes

From the Government's perspective, an important question is whether the performance pay schemes generated an overall economic return, i.e. did the increased revenues collected due to the schemes amount to more than the cost of the additional honoraria that were paid? Having a positive ROI was thus established as the threshold of success for the project.

The answer is indeed quite positive. The performance pay schemes resulted in a significant and positive return on investment for the Government of Punjab.[4] In total, circles under the scheme generated a profit of Rs11 million in the first year (from 150 circles) and a profit of Rs33 million (from close to 200 circles) in the second year, yielding a total of Rs44 million more revenue than the cost of the incentives.

[4] ROI is defined as the percent increase in additional revenue due to the scheme above the cost of the incentives, i.e.

$$ROI = \frac{Additional\ revenue - Cost\ of\ incentives}{Cost\ of\ incentives}$$

The cost of incentives is measured as the sum of honorarium payments to circle staff. The additional revenue collected due to the scheme is calculated using regression analysis with the following specification:

$$Log\ recovery_{cs} = \alpha_s + \beta Treatment_{cs} + \delta Baseline\ log\ recovery_{cs} + \varepsilon_{cs} + \varepsilon$$

where c indexes the circle in randomization stratum s. Because the ROI calculation is interested in level increases – not percent increases – the straightforward approach would have been to run the regression in levels. But since the level regression is also sensitive to outliers, we have used a log specification and then weighted observations by baseline recovery in order to reproduce the level effect, which gives the more economically meaningful result.

Overall ROI was calculated by comparing the Revenue, Revenue Plus and Flexible Bonus circles to control circles. The ROIs for each scheme were calculated by comparing each of these groups individually to the control circles.

4.6.1 Overall ROI

The following tables summarise the estimated ROI in both years of the pilot. Overall ROI was positive in both years and stronger in the second year.

As Table 4 shows, in the first year of the project, the 160 circles generated additional revenue (over control circles) of Rs98.2 million and the total honorarium pay was Rs87.3 million, giving an ROI of 12.4 per cent. In the second year (where we increased to 218 circles), the benchmarks were increased, and so even though we had a higher impact, the cost of the honoraria was no longer as high (per-circle cost in Year 1 was Rs545,945 compared to Rs497,189 in Year 2). Thus the ROI in the second year across all the schemes increased to 29.9 percent. Combined across the two years, the overall ROI of the three schemes was 22.1 percent.[5]

Table 4: Overall ROI

Period	Number of circles	Additional revenue due to scheme (Rs)	Cost of incentives (Rs)	ROI
Year 1	160	98,175,608	87,351,267	12.4%
Year 2	218	140,742,397	108,387,160	29.9%
Year 1 + year 2	—	238,918,005	195,738,427	22.1%

4.6.2 ROI by scheme

We already know that the Revenue scheme had the largest impact. Consistent with this, we also found that the Revenue scheme had the highest ROI.

Table 5 shows the ROIs separately for each of the three schemes for Year 1, Year 2 and then the two years combined. In Year 1 the Revenue scheme had an impressive ROI of 92 percent. This fell to 50.2 percent in Year 2, but still the overall ROI for this scheme was an encouraging 68.2 percent. By contrast, the overall ROI for the Revenue plus scheme was 11.9 percent, and it was negative for the Flexible Bonus scheme.[6]

[5]In the second year, we introduced an Information-only treatment, which provided feedback to circle staff on their performance, without providing any incentive pay. For the most part, this treatment had little effect – slightly positive, but insignificant. Because the ROI calculation is sensitive to estimating the treatment effects precisely, we have separated the Information-only treatment from the controls in the ROI calculation, though we have included it in the controls elsewhere in the report.
Including the Information-only treatment in the controls drops the ROI in Year 2: overall ROI drops from 29.9 percent to 13.6 percent. ROI for the sub-treatments drops to 33.5 percent for Revenue, 13.0 percent for Revenue Plus and –15.4 percent for Flexible Bonus.

[6] There are a number of factors at play that explain the difference in ROI between Year 1 and Year 2. For the Revenue scheme, we have a rather paradoxical result that ROI *decreases* in the second year even though the treatment effect is larger. This is mostly because the composition of circles that generated the additional revenue in Year 2 changed. In Year 1, a disproportionate share of the increase in revenue came from larger circles, which have lower reward rates and are weighted higher in the ROI calculation. Because the additional revenue was more evenly spread across circles in Year 2, overall ROI declined. For the Revenue Plus and Flexible Bonus circles, improved ROI in the second year was principally a result of adjustments to the scheme made at the end of the first year. Both of these examples highlight the important fact that the success of the scheme depends on the details of its design, as well as the need for constant monitoring and tweaking to ensure the scheme's success.

Table 5: Scheme-by-scheme ROI

Scheme	Year 1 ROI	Year 2 ROI	Year 1 + Year 2 ROI
Revenue	92.0%	50.2%	68.2%
Revenue Plus	–11.6%	28.0%	11.9%
Flexible Bonus	–44.4%	0.12%	–22.0%

4.6.3 ROI over time

The above calculations assume that performance pay schemes are in place every year, yet our results suggest that the increased revenue may be coming from increases in the tax base. If so, this increase may continue in subsequent years even if performance pay is not present.

To the extent that effects persist, the ROI calculations presented above are significantly understated, since any extra revenue after incentive payments stop comes free of cost. Table 6 projects how the ROI calculations would change under a number of scenarios, which consider a four-year overall period, assuming that the schemes are in place for the first two years but not in the last two. The scenarios vary by how we assume the treatment effect will decay over the last two years.

Table 6: Four-year projected ROI

#	Scenario	Treatment	Total ROI
1	Treatment effect remains at 66% in the first year after performance pay period and drops to 33% in the second year	Overall schemes	92.0%
		Revenue scheme	151.4%
2	Treatment effect remains at 75% in the first year after performance pay period and drops to 25% in the second year	Overall schemes	92.9%
		Revenue scheme	152.5%
3	Treatment effect remains at 50% in the first year after performance pay period and disappears in the second year	Overall schemes	57.3%
		Revenue scheme	110.0%

Even under relatively mild assumptions (such as Scenario #3), incorporating future revenue streams due to treatment makes a substantive difference on ROI: overall ROI more than doubles from 22.1 percent to 57.3 percent, and for the Revenue treatment, ROI increases by more than 40 percentage points, from 68.2 percent to 110.0 percent.

There is also some additional evidence to support the persistence of the treatment effects. In the FY2013- 2014 fiscal year, the performance pay schemes were no

longer in place as it was decided that a decision would be taken regarding their continuation after the final results of the project were presented and discussed. This provides us with a real check on whether the impact persists even after the performance pay period is over. Initial data from Quarter 2 of FY13/14 does suggest that effects for the Revenue scheme do persist over time, though it is too early to say what the exact trajectory of the effect over the post-treatment period will be. Under plausible scenarios, we can expect the overall ROI for the Revenue scheme over a four-year period to be in excess of 100 percent.

5. Research by-products

One benefit of the embedded researcher–policymaker engagement model is that in addition to the main project, it naturally makes possible by-products that are valuable for policy where researchers can add value at no cost to the government. The property tax experiment had a number of such valuable by-products that we have summarised in the following sections. These generated a range of other benefits over and above those arising from the main project.

Some of these by-products were integral to the project; for instance, discovering trends, identifying outliers and recalculating circle benchmarks would have been prohibitively time-consuming in the absence of data collation and digitization. Other by-products, like data visualization, were beneficial to the E&T Department but were not essential components of the intervention.

The extent to which these by-products required researcher involvement also varied. Data visualization, for instance, was developed solely by the project team and required extensive technical knowledge of web-based software development. On the other hand, the third-party property survey designed by the project team was only a complement to the property-level data collection initiative already in progress at the E&T Department.

Note that all circles – both treatment and control – benefited from these by-products. Since treatment effects were measured against controls, we could identify the effects of the performance pay incentives separately from those of the research by-products.

5.1 Revenue and personnel data collation and digitization

We collated and digitized the past six years of tax collection data at the circle level, starting with FY08/09. All circles provided us with quarterly circle statements, which were digitized and then verified through the consistency checks mentioned earlier.

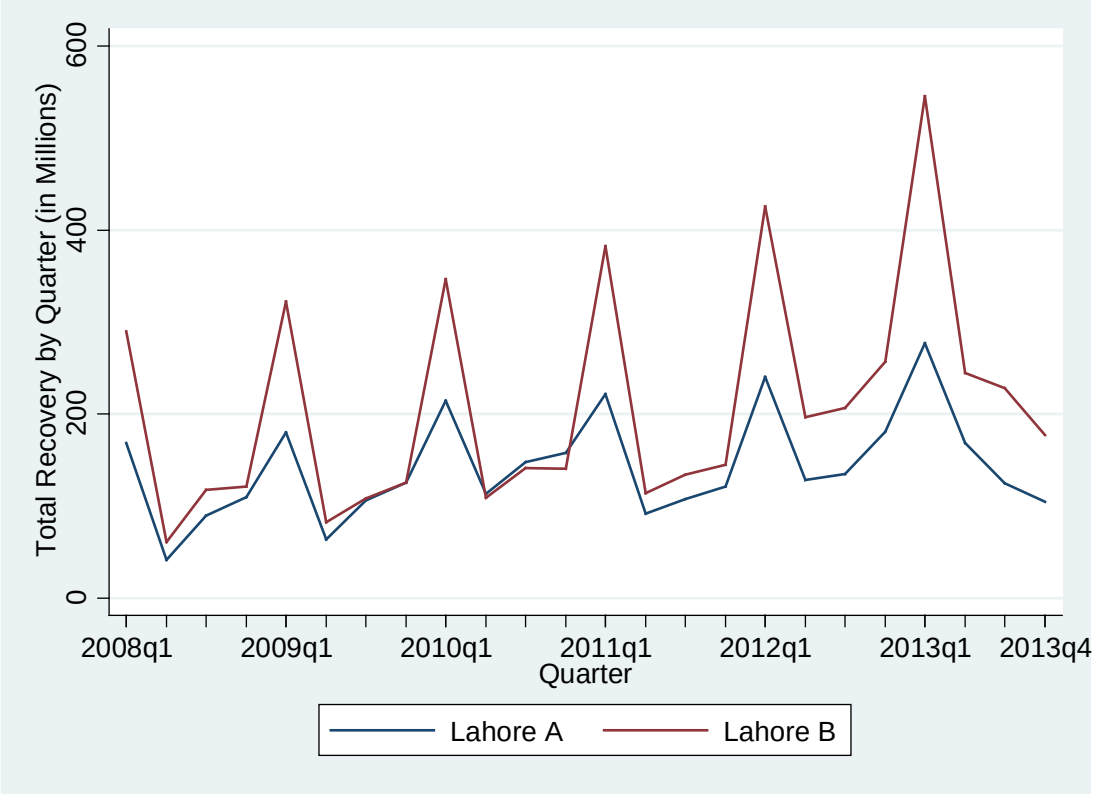
Combined with personnel data, which were also digitized, these data allow the department ready access to circle- and circle staff-level historical performances. Specifically, the data offer the opportunity to investigate pertinent questions about designing objective performance indicators for staff and having uniform data collection procedures across the province. It allows for easy tabulation and visualization of performance and staff movements. Moreover, it makes it easy to track changes over time, both in terms of how specific areas and regions are

performing with regard to changes in collections, but also in tracking how individual staff members perform over time.

Although the digitization was initially meant to gauge the performance of circles for treatment effects, it led to the formation of other highly beneficial methods of assessment. As a result of this exercise, the department is now piloting a new and more detailed reporting format (see below). This format will not only make it easier to understand the yearly workload of circle staff, but will also provide deeper insight into the variation in cyclical recovery activities, with the eventual goal of moving to a new system of computerized reporting.

Box 3. The value of digitising data

Digitising the tax data allows supervisors to see trends in department performance that would be difficult to spot otherwise. For example, the following chart plots recovery collection within each quarter over the past six fiscal years for the two divisions in Lahore.



The graph confirms the interesting seasonality in collections that the department is aware of. Most of the department's collection is done in the first quarter, and visually we can see that it is about double the collection amount in subsequent quarters.

We can also see that collection rates used to be similar in Lahore A and B after the first quarter, but this has stopped being the case since FY11/12.

Simple graphs like this are able to provide a wealth of information. Keeping track of these trends and patterns will help the department make more informed management decisions.

5.2 Standardizing statement templates

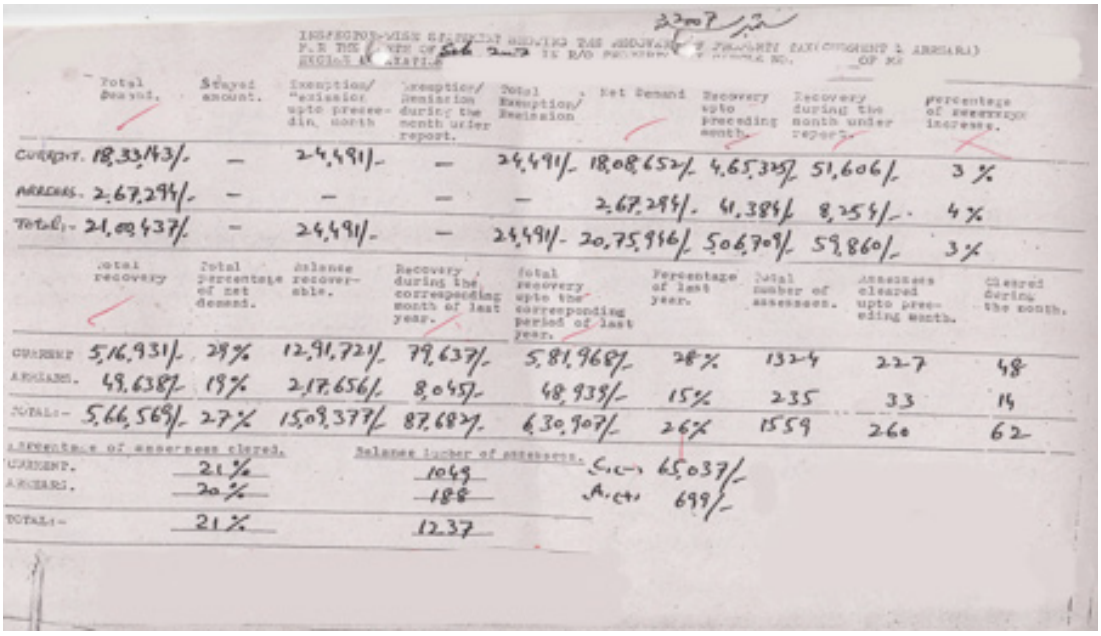
Circle statements are made at the circle level by the staff and are further verified by a higher-ranking tax official (in most cases the AETO or ETO). In many cases these statements are still made by hand and then compiled by the superintendent or assigned person in the divisional headquarters. Statements do not follow a standardized format, which leads to differential recording practices across the province (see Box 4).

In order to streamline data-reporting strategies, we also developed new templates for reporting data at the circle, AETO, ETO and director levels. These were complemented by a statement guide that walks officials through the statement in detail and explains the method for completion. The guide also gives the definitions of terms within the statement to avoid confusion and increase consistency within data reporting across the province. We hope that the department will incorporate this template in their regular work and that it will provide more information to allow for better assessments.

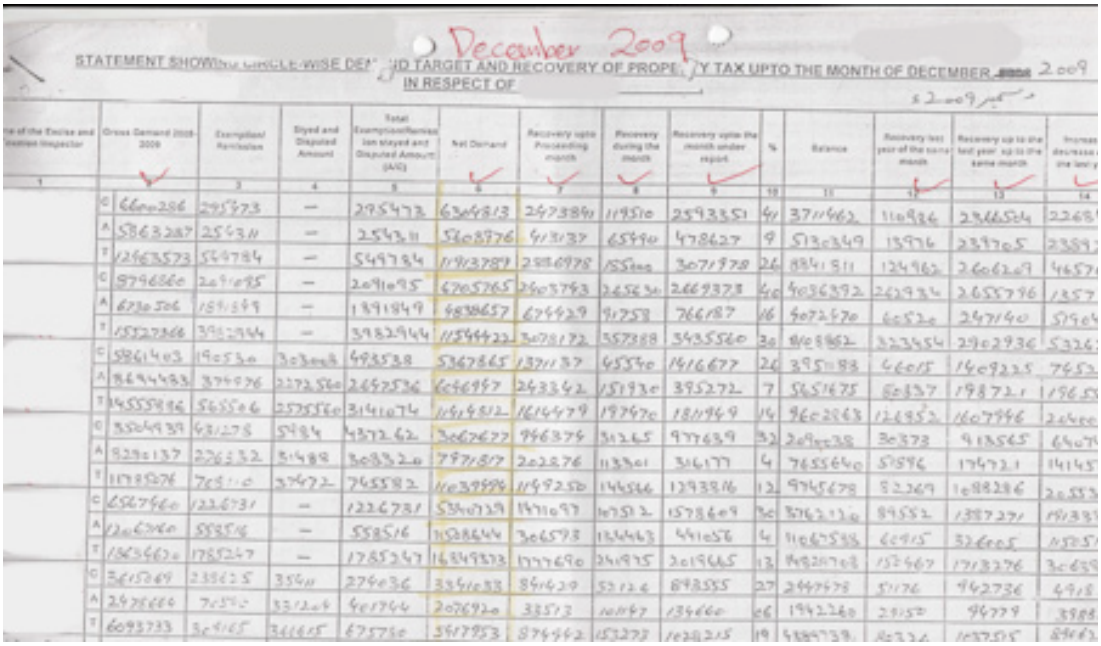
Box 4. Reconciling statement formats

We can see these differential practices in the examples provided below. The first statement is for Dera Ghazi Khan and the second for Lahore A. Both are hand-written; however, they follow very different formats. In fact, they even have different columns, which can lead to confusion within the department regarding definitions of terms. ‘Disputed’ amount, for example, should not be excluded from net demand, but due to confusion between ‘disputed’ amount and ‘stayed’ amount, both are excluded. This leads to lower net demand figures and higher recovery rates.

DG Khan circle statement



Lahore circle statement

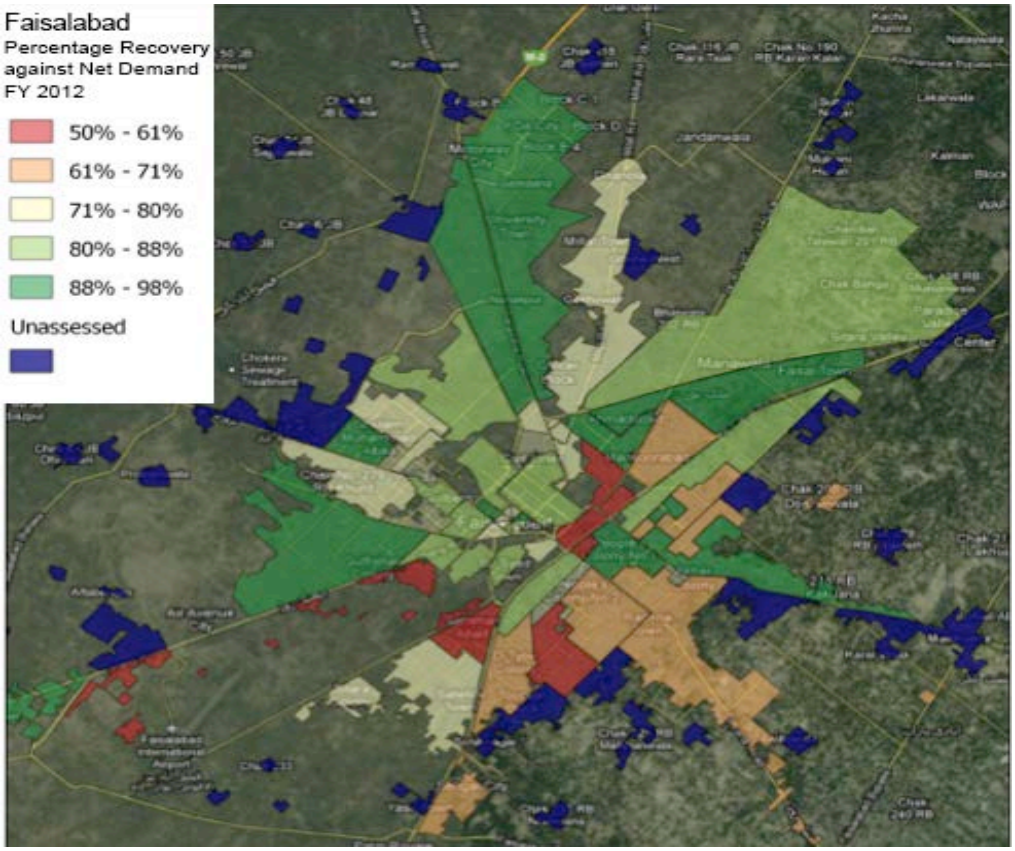


5.3 Circle mapping and tracking

We also mapped out the boundaries of tax circles using Google Earth. This extensive exercise was the first of its kind to be done in the E&T Department and served as the basis for developing effective data visualization tools (see below). We met with staff from each individual circle and relevant senior officers and mapped out each tax circle's boundary. This makes it easier to locate 'unassessed' areas that are legally taxable but not being taxed.

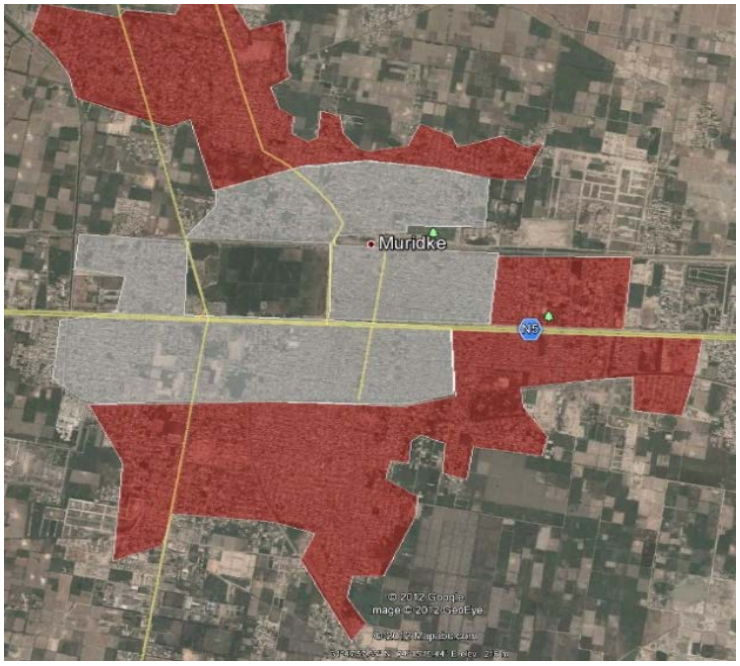
As an example, Figure 12showcases how circle mapping can allow a better representation of performance in a given region. The figure not only readily demonstrates which tax circles in Faisalabad had the best (and worst) performance in terms of recovery rates, but also highlights parts of the city that are not officially taxable units (i.e. are unassessed) but have had urban growth and therefore may be prime candidates for being declared as new rating areas. Specifically, this exercise revealed that as much as 36 per- cent of the city could conceivably be brought into new tax rating areas given the high housing density and urban expansion visible in the map (see Box 5). This also led to the development of the data visualization tool and allowed the department to visually assess performance, which it was not able to previously.

Figure 12: Tax circles in Faisalabad



Box 5. Helping identify new rating area

The image below shows Muridke, District Sheikhupura in December 2011. At that time the areas that are shaded were not officially taxed, as they had not been declared as a rating area (they now are). This can lead to situations where two areas have the same urban density but one is officially taxed while the other is not. An example of this is shown in the zoomed image of the boundary between rating and non-rating areas on the right panel – one side of the road is included in the tax circle, while the other is not yet one can see the housing density appears to be very similar.



5.4 Property-level data and tax calculators

The department is updating and digitising property-level records that can then be used to auto-generate assessed tax notifications and help in collections. A useful complement to this exercise was the third-party property survey, which covered over 25,000 properties across different rounds, mentioned in more detail earlier. Given that these third-party surveys were more detailed (a typical survey lasted 30 minutes) and provided comprehensive information on property and owner characteristics and views, they provided a consistency check and complement to the survey exercise being carried out by the department in 2013–2014, and the property-level data digitization being carried out. The property tax calculator developed for the project can also provide valuable feedback as the department moves towards automatically generated tax payable notices.

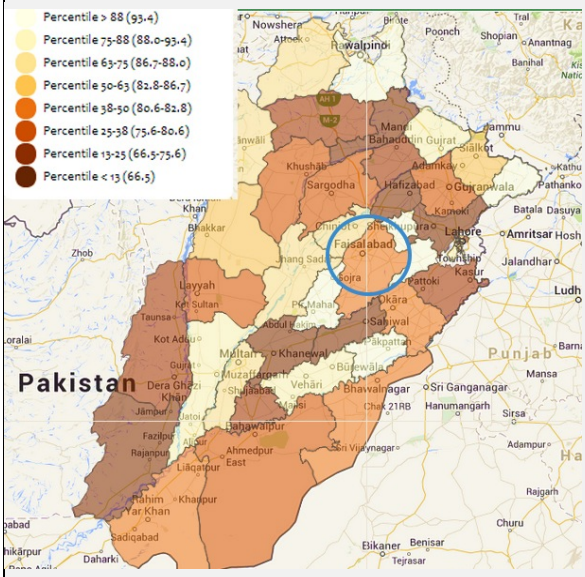
5.5 Data visualization

Finally, in order to fully utilize all of the above beneficial by-products, a ‘proof of concept’ web-based data visualization tool is being developed for the department. This tool is not intended to be a fully developed product –since that would require a full project to be setup–but it is meant to serve as an initial pilot to demonstrate the value of such tools. The pilot is sufficiently developed in that it can already facilitate supervisors’ monitoring of circle-wise progress in recovery, tracking increases in net demand, analysing trends in revenue increase and developing targeted strategies to improve division-wise property tax collection. It is also able to generate heat maps based on different variables and show how they vary across circles, districts and divisions. Heat maps will also provide patterns of high- and low-collection tax circles, which combined with maps of unassessed areas can help identify localities with a higher marginal return to effort for increasing property tax collection (see Box 6). Should the pilot be successful, it can form the basis for the department to request the full development and deployment of the tool.

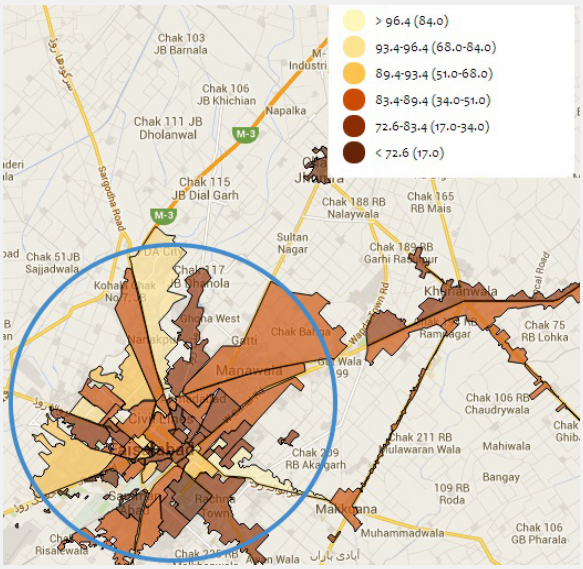
Box 6. The UIPT performance dashboard

The Urban Immovable Property Tax Performance dashboard, currently in its final stages of development with the department, will allow users such as department supervisors to compare a range of performance measures across different level of aggregations and compare performance over time. This not only provides a more contextual and relative sense of performance, but also identifies key areas and strategies for performance improvement. We provide some snapshots below to highlight features of this tool.

Comparing performance across districts and circles



Panel A



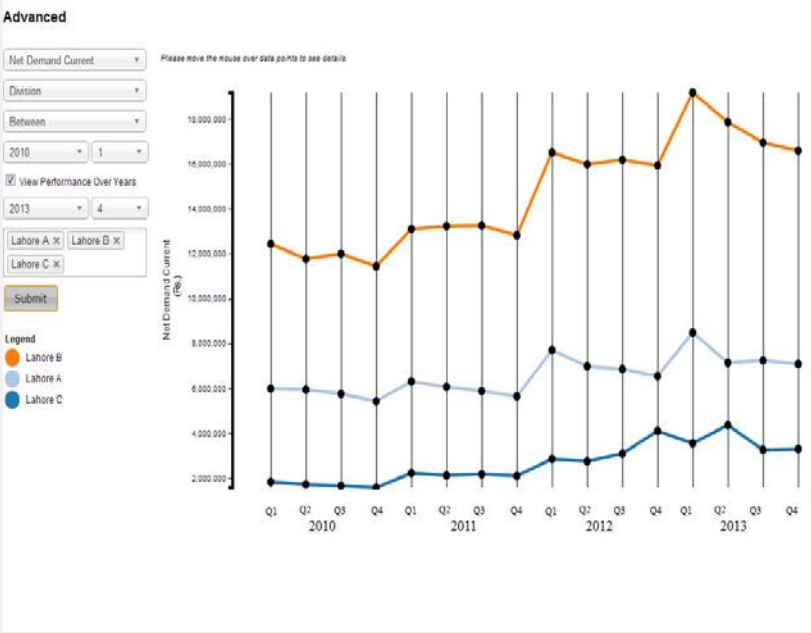
Panel B

Panel A provides a comparison of one possible metric of performance, recovery rate (i.e. percentage collections against total assessment) across districts. Lighter shades indicate better- performing districts.

Panel B then demonstrates how one can zoom into a particular city and examine how performance varies across tax circles within that city.

Comparing performance over time

This graph shows how the three regions of Lahore performed (in terms of total tax assessed) from 2010–2013.



6. Policy implications and next steps

This section summarizes some of the initial policy implications of the project. While ultimately we propose that each of these be discussed with the relevant stakeholders, the hope is that this section serves as the basis of these discussions by bringing valuable hard evidence to the table. The ultimate goal would be to inform policy and help improve performance.

6.1 Lessons learnt for performance pay schemes

Based on the results of the two-year study, we feel the following points are noteworthy and should form the basis of a more comprehensive performance pay system in the E&T system and potentially other related departments:

- **Performance pay works in raising revenues:** The findings provide conclusive and scientifically rigorous evidence (for the first time in Pakistan or anywhere, to our knowledge) that rewarding circle-level staff through monetary incentives leads to higher tax collection. The results broadly show that incentives work: incentives increase revenue by about 13 percentage points more than the usual performance, almost doubling the usual year-on-year rate of increase the department experiences. Moreover, the ROI is positive. This means that for every rupee invested by the government in the form of performance bonuses, it received more than that amount in the form of additional revenue generated. The main message of the project is, therefore, that if revenue increase is an important outcome for the government, some form of monetary incentive has to be an important part of the performance management process for field-level staff.

- **Simpler and objective performance pay schemes perform better:** The Revenue scheme–which was the simplest in that it rewarded only on one dimension, collections above a benchmark–performed best in terms of impact on collections. In both years, the Revenue scheme consistently had the largest treatment effect and the largest ROI of the three schemes. Revenue Plus performed second in terms of impact on collections; it has an ROI that is positive, but lower than that of the Revenue scheme. The Flexible Bonus scheme had been designed to provide the department with a holistic assessment of the performance of tax inspectors. It had also been designed to incentivise performance in multiple dimensions, including both revenue and non-revenue outcomes; however, this scheme performed worst on all the dimensions and also had a negative ROI. Since this scheme is very similar to the various across-the-board wage increases in other departments that have not been linked to performance, the evidence from this project suggests that these wage increases are unlikely to work as well. It therefore appears that performance pay, based on simple and clear directions that explicitly link to performance on objective dimensions, is a key element of an effective performance pay scheme.

- **Performance pay schemes may need to be monitored to ensure customer satisfaction:** The general concern with performance pay schemes that only reward on collections, such as the Revenue scheme, is that they may lead to customer dissatisfaction and over-taxation. While our current findings do not show strong evidence for these concerns, there is a possibility that these effects may show up if a broader range of satisfaction measures were to be considered. It is therefore recommended that the level of customer satisfaction be monitored regularly through the type of third-party surveys conducted as part of the project. These surveys could readily be combined with the *chalaan* (notice-sending process). Moreover, while the Revenue Plus scheme–which rewarded on collections as well as customer satisfaction and assessment accuracy–may have been less effective, the measures developed in the scheme can still serve as an important and separate check: staff would know that these data are being gathered, and that if there are sharp drops in satisfaction or assessment accuracy they may risk losing their performance pay.

- **It may be more cost-effective to introduce performance pay 'periods' every few years:** While these results will be more conclusive at the end of the current fiscal year, preliminary evidence suggests that the benefits of performance pay may continue even after the performance pay period is over. This is likely because the performance pay schemes work primarily by increasing the tax base, the benefits of which last over time. While further work is needed on this, a persistent effect means that it may be more cost-effective for the government to introduce performance pay schemes every few years. For example, once in every three years the government could have a year in which large performance pay rewards are issued– the precise length of time should depend on how long it takes the tax base to expand. This could also be timed with the year in which a property re-survey activity is being carried out. Doing so can both allow greater reward rates to be paid in the performance year and make the schemes more cost-effective for the government. For example, our preliminary analysis showed that under plausible assumptions the ROI for the Revenue scheme can increase from 68.8 percent to over 150 percent by having such performance pay periods. Still, we must remain wary of adverse incentives, whereby tax collectors may choose to not record increases in the tax base except during performance pay periods. While further discussion is warranted here, we feel this could be a promising way forward.

- **Performance pay schemes may have to be designed differently for supervisory tiers:** While the project established without any doubt that performance pay works for field-level staff, the results of introducing the simplest Revenue scheme (which worked the best for field staff) for supervisory staff were not conclusive. While this could partly be due to the fact that the number of supervisors needed to detect a statistically reliable effect was not sufficient (especially given that several of them were transferred out and were

thus no longer eligible for the scheme), there are reasons to suspect that simple performance pay schemes may not work well for supervisors. In particular, supervisors engage in a wider range of tasks and therefore rewarding them on the basis of one dimension – such as collections – may not be effective. The challenge here, especially given the lack of success of the multi-dimensional and more subjective Flexible Bonus scheme, is how to design such a supervisory scheme. This is an important area, however, and one where further study is required.

6.2 Related ongoing lessons and reforms

While this project has demonstrated that performance pay can help increase revenue collection, it is just one of a series of complementary reforms that are required to improve the property tax system in Punjab. Some of these reforms, like digitization of property records, have already been initiated. A major area of reform is rationalization of property tax rates, and removal of distortions. Since these reforms have been covered in other works (e.g. World Bank 2009; Nabi & Sheikh 2011), this report will focus on other relevant sets of reforms.

- **Human resource reform – merit-based transfers and postings:** While the current project focused primarily on financial rewards, an important aspect of HR policy is to also introduce non-financial reward systems– specifically, create good incentives for performance by implementing a merit-based and transparent transfers and posting (T&P) system. This is not only aligned with the Government of Punjab's policy of adopting a performance management system, but has also been the desired goal of HR policies of most public sector agencies, and the explicit objective of many civil service and tax reforms. The literature suggests that reallocating staff from low-productivity jurisdictions to high-productivity jurisdictions can have significant effects on performance (Iyer & Mani 2012; Hertzberg *et al*. 2010). Since past international reforms had not been scientifically evaluated, these did not generate rigorous evidence to sustain the reforms beyond a few years. By contrast, the current staff allocation project in Punjab has both been scientifically designed as a complement to the earlier incentive project and will be rigorously evaluated. Following approval from the government, this project has been underway since July 2013 and will be evaluated over a two-year period. The results from the evaluation of merit-based systems will provide valuable learning for both the allocative and incentive gains generated by an objective T&P system. These, along with the results from the current project, will help the department develop a comprehensive and optimal HR policy.

- **Selection of employees:** The property tax experiment demonstrated that performance can be improved by introducing reward systems even for existing employees, many of whom may never have been exposed to such systems and therefore may have faced a harder time in adjusting. The impact of performance pay systems may be enhanced even further through the induction of new employees: if wages are low and individuals are not rewarded for

performance, many good and able individuals may not even choose to apply for government jobs. Thus, in theory, performance pay can generate sustained performance increases over time as new, efficient and able people join the civil service. The current pay-for-performance projects had been designed and evaluated keeping selection fixed: that is, the performance incentives were offered to the existing tax collectors. Were it possible to also offer performance pay contracts to new inductees, a related project could scientifically examine whether this does indeed lead to the selection of higher performing employees. We would recommend that the department explore the feasibility of doing so.

- **Engaging citizens – rebuilding the social compact:** While the property tax experiment did incorporate citizen feedback, it was not designed to directly engage the taxpayer. Ultimately, however, raising taxes requires the participation and willingness of those who are taxed. Therefore, an important leg of tax reforms in any country is building a positive relationship with taxpayers. Although no one likes paying taxes, developed countries facilitate collection by linking individual's payments with their receiving better government services. In the case of property tax and municipal governments, this link can be made more concrete by explicitly linking local collections with local services identified by citizens. There are several underlying causes behind the currently low willingness to pay municipal taxes by the citizens. First, ***citizens have no idea how their tax payments are allocated.*** Taxes, especially local property taxes, are not classified or earmarked upon receipt, resulting in a perceived disconnect between payments made and services received. Second, ***citizens have low trust in the system.*** It is likely that due to inefficient service provision in the past or any number of other reasons, citizens are unwilling to pay taxes because they do not trust that their payment (investment) will produce positive services (return). To address these causes, the Departments of Local Government and Excise and Taxation and the research team are currently designing a reform involving a series of potential interventions that will then be subjected to a rigorous evaluation. The key features of the proposed design include establishing a link between property taxes paid and services received; specifically tying property tax collection to demanded services; and facilitating matching incentive grants. This project will actualize the Government of Punjab's vision of strengthening the link between citizens and state.

6.3 Next steps

As discussed above, the project has generated several interesting policy lessons, and given these lessons, an important next step would be to convene a set of discussions with relevant stakeholders where the results of the study are presented and its implications for policy reforms debated. We have already shared results with the Chief Minister of Punjab, and organized a policy dialogue on civil service reform attended by high-ranking civil servants from various departments. This project has presented a rare opportunity in terms of the detailed and rigorous scientific evidence

it has generated, and it is critical that policymakers have the opportunity to benefit from it and help implement better and more evidence-driven policy.

We are encouraging the E&T Department to take a lead in convening the set of discussions on how to incorporate the lessons from the project into policy decisions. These discussions are likely to involve stakeholders from within the department (circle staff and supervisory officers), relevant other departments (Finance, Planning and Development, Regulations, Law, etc) and the research team. These discussions would hope to cover the design, timing and feasibility of performance pay schemes for tax staff, but also discuss current ongoing reforms mentioned above and explore the feasibility of future reforms.

Appendix A. Sample design

The sample included all 482 urban tax circles in the state of Punjab, Pakistan. As per usual departmental procedure, however, circle boundaries changed over time to accommodate personnel requirements and shifting property growth rates across the province. For consistency, we maintained the original circle boundary definitions when tracking circle performance over time.

Appendix B. Descriptive statistics

Table 7 below covers the main variables of interest in the study, divided by data source.

Panel A presents summary statistics from the administrative data. Several observations are worth noting. First, current year revenues are substantially larger than arrears – the mean of log current revenues is 15.52 compared with just 13.9 for log arrears, implying that on average current revenue in the typical circle is about five times as large as arrears. This suggests that the main effects on total revenue will likely be felt through increases in current year revenue. Second, there is much more variation in arrears – the standard deviation in log arrears is about 1.5 times that of log current revenue – implying that detecting effects on arrears statistically will be more difficult. It is also interesting to note that the log recovery rate (the log of tax revenue divided by the tax base net of exemptions) is –0.14 for current year taxes – this implies that about 85 percent of all taxes that are demanded by the government are in fact paid. Thus while non-payment is a substantial issue, it is still the case that the majority of citizens do in fact pay the tax bills they receive. Thus any potential evasion may come from under-assessment of properties (as we will see in Panel B) rather than flagrant disregard of issued tax notices.

Panel B presents statistics from the property survey. First, observe that 84 percent of properties we randomly sampled in the field were successfully located on the tax registers. Again, while there are a substantial number of untaxed properties, it is not the case that only a few properties are on the tax rolls. Second, conditional on being on the tax rolls, on average properties appear to be under-taxed. This suggests a natural channel through which the treatment will operate, which in fact we found to be the case.

Table 7: Descriptive statistics

	Mean	SD	N
Panel A: Administrative Data			
Log Revenue (Total)	15.75	0.74	482
Log Revenue (Current)	15.52	0.73	482
Log Revenue (Arrears)	13.91	1.17	479
Log Tax Base (Total)	16.14	0.81	482
Log Tax Base (Current)	15.86	0.73	482
Log Tax Base (Arrears)	14.40	1.37	479
Log Non-Exemption Rate (Total)	-0.23	0.20	482
Log Non-Exemption Rate (Current)	-0.19	0.13	482
Log Non-Exemption Rate (Arrears)	-0.30	0.41	479
Log Recovery Rate (Total)	-0.16	0.18	482
Log Recovery Rate (Current)	-0.14	0.14	482
Log Recovery Rate (Arrears)	-0.19	0.29	479
Panel B: Survey Data			
Property successfully found in administrative records (dummy)	0.84	0.37	11,971
Quality of Tax Department [0-1]	0.53	0.22	6,050
Satisfaction with Tax Department [0-1]	0.55	0.23	6,050
Inaccuracy	0.34	0.27	9,879
Tax Gap	-0.099	0.42	9,879
GARV	31,915	248,026	11,186
Self-reported tax payment in FY 2013	4,246	20,255	10,047
Bribe Payment	2,073	3,932	5,993
Frequency of Bribe Payment	0.76	0.88	4,802

Notes: Panel A statistics from administrative data are shown at the end of Year 2 of the study (FY 2012-2013). Each observation is one of the 482 circles as defined at the time of randomization. Panel B statistics from the property survey are for randomly sample properties only. Subjective variables - i.e., Quality, Satisfaction, Bribe Payment, and Frequency of Bribe Payment - are reported for Phase 1 circles only (see text for more details).

Appendix C. Balance test

The randomization was carried out through public lotteries, with a large number of representatives from the E&T Department present. This helped minimize any perceived bias especially since the performance pay schemes were popular (most staff wanted to opt in). In order to reduce any concerns about differential selection across the schemes while maintaining informed consent, the lottery occurred in two stages. In the first stage, circles were selected to participate in the project and staff consent to participate was sought. Staff were told about the three possible incentive schemes while making clear that a second lottery would determine which scheme they would be assigned to.

Once consent was obtained, a second lottery was held to assign consented circles into particular incentive schemes. Over 95 percent of circle staff that were selected in the first lottery consented to participate. Given the extremely high consent rates observed in the first year, both stages were conducted in a single lottery in Year 2. The lotteries were held as close as possible to the start of the fiscal year on July 1. Each of these lotteries had 19 strata based on 11 administrative divisions and for all but the smallest few divisions, circle size. Table 8 shows sampling and assignment over the course of the intervention.

Table 8: Experimental design

	Randomization		Implementation	
	Year 1	Year 2	Year 1	Year 2
Revenue	53	72	47	68
Revenue Plus	54	74	48	68
Flexible Bonus	54	73	49	67
Information	0	70	0	66
Control	322	194	338	213

Notes: The first two columns (under Randomization) show the number of circles that were assigned to each of the three (or four) treatment types in each year. In cases where staff did not consent to treatment after the first ballot (in Year 1), circles were assigned treatment values of 1/3 for each main treatment type (i.e. Revenue, Revenue Plus, and Flexible Bonus). Values are rounded. The second two columns (under Implementation) show the number of circles that were actually implementing the treatment at the end of the fiscal year. Treatment wasn't implemented either because of lack of consent or because the initially selected circle staff were transferred to new posts. See text for more details.

Table 9 compares the selected circles to controls on their baseline characteristics in the administrative data based on the final randomization at the end of Year 2. Column 2 shows the difference between all incentive treatments and controls. All comparisons are done on the original randomization (i.e. intent-to-treat) results. Standard errors are shown in parentheses, and randomized-inference based p values are shown in brackets. None of the differences are statistically significant. The

remaining columns show the three incentive sub-treatments (Revenue, Revenue Plus and Flexible Bonus), Information-only scheme and supervisory treatments, all compared to their respective controls. Out of the 42 comparisons made (seven variables * six columns), only one is statistically significant at the 5 percent level (the non-exemption rate in the Flexible Bonus scheme), as one would expect by random chance.

Table 9: Balance test

	Main Treatment					Information Treatment		Supervisory Treatment	
	Control	Treatment	Revenue	Revenue Plus	Flexible Bonus	Control	Treatment	Control	Treatment
Log Revenue	15.47	-0.037	0.024	-0.053	-0.055	15.46	0.050	15.65	-0.167
		(0.042)	(0.059)	(0.057)	(0.058)		(0.065)		(0.089)
		[0.347]	[0.683]	[0.375]	[0.378]		[0.451]		[0.156]
Log Recovery Rate	-0.333	-0.015	0.007	-0.002	-0.039	-0.330	-0.026	-0.366	0.006
		(0.024)	(0.035)	(0.034)	(0.040)		(0.037)		(0.041)
		[0.532]	[0.827]	[0.965]	[0.271]		[0.488]		[0.909]
Log Non-exemption Rate	-0.251	-0.024	-0.001	0.009	-0.059	-0.245	-0.035	-0.267	0.009
		(0.019)	(0.021)	(0.023)	(0.038)		(0.035)		(0.019)
		[0.200]	[0.964]	[0.721]	[0.0280]		[0.292]		[0.791]
Number of staff posted	2.564	0.055	0.038	0.056	0.088	2.576	-0.077	2.549	0.054
		(0.053)	(0.070)	(0.076)	(0.077)		(0.080)		(0.062)
		[0.300]	[0.600]	[0.494]	[0.280]		[0.343]		[0.607]
All positions filled	0.519	0.059	0.043	0.094	0.054	0.531	-0.072	0.538	-0.011
		(0.044)	(0.065)	(0.064)	(0.066)		(0.065)		(0.056)
		[0.188]	[0.506]	[0.146]	[0.414]		[0.267]		[0.908]
Log benchmark	15.44	-0.017	0.036	0.014	-0.073	15.44	-0.010	15.59	-0.114
		(0.044)	(0.062)	(0.059)	(0.064)		(0.074)		(0.095)
		[0.665]	[0.565]	[0.845]	[0.252]		[0.896]		[0.390]
FY 10-11 log growth rate	0.0280	0.003	0.017	-0.005	-0.005	0.0233	0.026	0.0259	-0.008
		(0.013)	(0.014)	(0.016)	(0.023)		(0.022)		(0.019)
		[0.824]	[0.378]	[0.813]	[0.813]		[0.217]		[0.788]
P-val, joint sig.		0.412	0.793	0.006	0.455		0.261		0.003
P-val, from RI		0.359	0.796	0.010	0.571		0.430		0.136

Notes: Randomization balance test. Columns labeled Control reflect control group means. Values in the treatment columns are the coefficients of a regression of the baseline value of the variable indicated in the row on a treatment dummy (or the set of subtreatments dummies), controlling for the relevant randomization strata. In the Main Treatment tests, the Information treatment is included in the controls. In the Information Treatment tests, the Information treatment group is compared against pure controls. The Supervisory Treatment test compares without regard to Main Treatment or Information Treatment status. Robust standard errors in parentheses. Randomization inference (RI) based p-values in brackets. RI statistics are based on 1000 re-randomization iterations. * p<0.10, ** p<0.05, *** p<0.01. Stars reflect randomization inference based p-values.

Appendix D. Study methodology

The research design was an RCT, which lends itself to straightforward empirical analysis using instrumental variable regressions. We used instrumental variable regressions because tax inspectors occasionally transferred posts, which meant we did not have perfect compliance over the two years of the study.

Specifically, we estimate 2SLS (two-stage least squares) regressions, where the endogenous variable is the treatment status at any point in time and the instruments are the results of the lottery. Our primary specification for assessing circle-level outcomes using the administrative data is

$$\ln Y_{cst} = \alpha_s + \beta Treatment_{cst} + \gamma \ln Y_{cs0} + \epsilon_{cst}$$

where Y_{cst} is the outcome of interest for circle c in stratum s at time t, and $Treatment_{cst}$is a continuous variable that takes values from 0 to 1 that represent the fraction of treated circle staff present in circle *c* in the last quarter of the given fiscal year. Y_{cs0} is the value of the outcome variable at baseline (i.e. in the fiscal year prior to randomization).$Treatment$ is instrumented by a binary variable that represents the circle's randomization status into any one of the three incentive schemes. We include stratum fixed effects (α_s) given the lottery was stratified by these strata. All regressions based on administrative data are run using circle boundaries that existed at the time of randomization. We report robust standard errors clustered at the level of the robust partition of circles, i.e. the maximum set of circles that have been involved together in a set of splits and merges since randomization.

To estimate the impact of the separate sub-treatments, we estimate the analogous regression separately by treatment:

$$\ln Y_{cst} = \alpha_s + \beta_1 Revenue_{cst} + \beta_2 RevenuePlus_{cst} + \beta_3 FlexibleBonus_{cst} + \gamma \ln Y_{cs0} + \epsilon_{cst}$$

For survey-based outcomes, we run regressions at the individual property level. As discussed above, we have two separate samples, the general population sampled from random GPS points, and properties that were sampled because they had a change in their tax assessment (either previously assessed properties that were reassessed, or properties newly added to the tax rolls). When examining the general population sample, we run regressions of the form

$$Y_{ics} = \alpha_s + \beta Treatment_{cs} + \epsilon_{ics}$$

where i is an individual property. As above, we instrument for $Treatment$ with the randomization results. We include stratum fixed effects and cluster standard errors at the circle level. When available, we include controls for baseline level outcome variables.

Appendix E. Regression tables

Below (Table 10 and Table 11) are the primary regressions tables used in our analysis.

Table 10: Impacts on revenue collected

	Year 1			Year 2		
	(1) Total	(2) Current	(3) Arrears	(4) Total	(5) Current	(6) Arrears
Panel A: Main Treatment						
Any treatment	0.091***	0.073***	0.152**	0.094***	0.091***	0.113
	(0.028)	(0.027)	(0.069)	(0.031)	(0.032)	(0.083)
Panel B: Subtreatments						
Revenue	0.118***	0.109***	0.134	0.129***	0.152***	0.005
	(0.035)	(0.034)	(0.099)	(0.043)	(0.044)	(0.133)
Revenue Plus	0.080	0.086*	0.072	0.093**	0.081*	0.175
	(0.053)	(0.052)	(0.110)	(0.045)	(0.049)	(0.114)
Flexible Bonus	0.071*	0.024	0.243**	0.056	0.035	0.148
	(0.038)	(0.035)	(0.098)	(0.041)	(0.042)	(0.108)
N	481	481	481	482	482	479
Mean of control group	15.671	15.379	14.030	15.745	15.518	13.915
Rev. vs. Multitasking p.	0.323	0.193	0.830	0.233	0.049	0.262
Objective vs. Subjective p.	0.530	0.090	0.212	0.220	0.084	0.634
Equality of Schemes	0.562	0.143	0.433	0.359	0.086	0.527
Joint significance	0.004	0.010	0.073	0.012	0.005	0.305

Notes: This table presents results on the impact of the performance pay schemes on revenue-based outcomes. We use instrumental variables regressions, where treatment status is instrumented with randomization results. The unit of observation is a circle, as defined at the time of randomization. Outcome variable is log revenue collection as of the end of the fiscal year, for total revenue (Columns 1 and 4), current year revenue (Columns 2 and 5), and collections against arrears (columns 3 and 6). Specification follows Equation 5.3 of the main text, and includes stratum fixed effects. 'Any treatment' in Panel A includes the 3 subtreatments in Panel B. The Information treatment is included in the control group. We report p-values from tests of equality of coefficients as follows: Rev. vs. Multitasking tests for equality between Revenue and the average of Revenue Plus and Flexible Bonus; Objective vs. Subjective tests for equality of the average of Revenue and Revenue Plus against Flexible Bonus; Equality of Schemes tests whether all coefficients are equal; and Joint significance tests joint null that all coefficients are equal to 0. Robust standard errors in parentheses. Standard errors are clustered by a robust partition of circles, i.e. the group of circles such that all circles that merged or split with each other are included within the same partition. * p<0.10, ** p<0.05, *** p<0.01

Table 11: Impacts on non-revenue outcomes

	(1) Quality	(2) Satisfaction	(3) Inaccuracy	(4) Tax Gap
Panel A: Main Treatment				
Any treatment	-0.006	-0.011	0.004	0.007
	(0.022)	(0.022)	(0.012)	(0.022)
Panel B: Subtreatments				
Revenue	0.006	-0.006	0.002	-0.022
	(0.036)	(0.037)	(0.017)	(0.029)
Revenue Plus	0.040	0.029	0.028*	0.015
	(0.026)	(0.027)	(0.016)	(0.032)
Flexible Bonus	-0.060*	-0.053*	-0.016	0.029
	(0.031)	(0.032)	(0.018)	(0.031)
N	6050	6050	9870	9870
Sample	Phase 1	Phase 1	Full	Full
Mean of control group	0.538	0.555	0.339	-0.103
Rev. vs. Multitasking p.	0.683	0.876	0.813	0.159
Objective vs. Subjective p.	0.015	0.064	0.099	0.315
Equality of Schemes	0.014	0.059	0.090	0.344
Joint significance	0.035	0.129	0.160	0.533

Notes: This table presents results on the impact of the performance pay schemes on non-revenue outcomes. We use instrumental variables regressions, where treatment status is instrumented with randomization results. Unit of observation is a property. Specification follows Equation 5.5 of the main text, and includes stratum fixed effects. Quality and Satisfaction were measured on a 5 point Likert scale and re-scaled to a [0,1] interval. Tax Gap is the difference in the official gross annual rental value (GARV) minus our estimated GARV, divided by the sum of these. Tax Gap measures over/undertaxation, with positive coefficients indicating overtaxation. Inaccuracy is the absolute value of Tax Gap. Sample is restricted to Phase 1 of the survey for subjective outcomes (Quality and Satisfaction). The Information treatment is included in the control group. We report p-values from tests of equality of coefficients as follows: Rev. vs. Multitasking tests for equality between Revenue and the average of Revenue Plus and Flexible Bonus; Objective vs. Subjective tests for equality of the average of Revenue and Revenue Plus against Flexible Bonus; Equality of Schemes tests whether all coefficients are equal; and Joint significance tests joint null that all coefficients are equal to 0. Standard errors are clustered by robust partition of circles, i.e. the group of circles such that all circles that merged or split with each other are included within the same partition. * $p<0.10$, ** $p<0.05$, *** $p<0.01$

Appendix F. Power calculations

The project team ran two-sample t-test-based power calculations (adjusted for clustering) with a treatment sample size of 150 and a control group of the same size. The standard deviation of both groups was assumed to be .78 log points. The intra-cluster correlation used was 0.51. The test assumed a power of 0.8 and a Type I error probability of 0.05. The mean of the control group was 15.5 log points (of total recovery), while the treatment mean was 16.2 log points, giving us a minimum detectable ITT effect size of 0.63 and a TOT effect size of 0.79.

References

Basinga *et al*., 2010. Paying Health Care Providers for Performance: Evidence from Rwanda. World Bank. Available at: http://siteresources.worldbank.org/EXTDEVDIALOGUE/Images/537296-1238422761932/5968067-1269375819845/Rwanda_P4P.pdf

Becker, G and Stigler, G, 1974. Law Enforcement, Malfeasance, and Compensation of Enforcers. *Journal of Legal Studies*, 3(1), pp .1–18.

Besley, T and McLaren, J, 1993. Taxes and Bribery: The Role of Wage Incentives. *Economic Journal*, 103(416), pp .119–41.

Cobham, A, 2005. Working paper no. 129: Tax Evasion, Tax Avoidance and Development Finance. *QEH Working Paper Series*. Available at: http://www3.qeh.ox.ac.uk/pdf/qehwp/qehwps129.pdf

Ellis, P, Kopanyi, M and Lee, A, 2007. Property Taxes in the Large Cities of Punjab Province, Pakistan. *Journal of Property Tax Assessment and Administration,* 4(2), pp .31–52.

Fisman, R and Wei, S-J, 2004. Tax Rates and Tax Evasion: Evidence from "Missing Imports" in China. *The Journal of Political Economy,* 112(2), pp. 471–500.

Friedman, E, Johnson, S, Kaufman, D and Zoido-Lobaton, P, 2000. Dodging the grabbing hand: the determinants of unofficial activity in 69 countries. *Journal of Public Economics,* 76(3), pp .459–493.

GHK International 2009. Fiscal and Financial Management Reform Taxation Reforms. Government of Punjab. Available at: http://www.prmp.punjab.gov.pk/sites/prmp.pitb.gov.pk/files/Finance_Final_Report_021109_By_Mr_Patrick_Kevin_Doherty.pdf

Gordon, R and Li, W, 2009. Tax structures in developing countries: Many puzzles and a possible explanation. *Journal of Public Economics,* 93 (7–8), pp. 855–866.

Hertzberg, A, Liberti, JM and Paravisini, D, 2010. Information and Incentives Inside the Firm: Evidence from Loan Officer Rotation. *Journal of Finance*, 65(3), pp. 795–828.

Iyer, L and Mani, A, 2012. Traveling Agents: Political Change and Bureaucratic Turnover in India. *Review of Economics and Statistics*, 94(3), pp. 723–739.

Khan, S and Inam, A, 2014. Computing System Interventions for Property Tax

Mookherjee, D and Png, IPL, 1995. Corruptible Law Enforcers: How Should They Be Compensated? *Economic Journal*, 105, pp. 145–59.

Muralidharan, K and Sundararaman, V, 2011. Teacher Performance Pay: Experimental Evidence from India. *Journal of Political Economy*, 119(1), pp. 39–77.

Nabi, I and Sheikh, H, 2011. Reforming the Urban Property Tax in Pakistan's Punjab.*IGC Policy Note*. Available at: http://www.theigc.org/wp-content/uploads/2014/09/Nabi-2011-Policy-Brief.pdf

Olken, B, 2007. Monitoring Corruption: Evidence from a Field Experiment in Indonesia. *Journal of Political Economy,* 115(2), pp. 200–249.

Schneider, F, 2007. Shadow Economies and Corruption all over the World: New Estimates for 145 Countries. *Economics: The Open Access, Open Assessment, E-journal*, No. 2007–9.

Van Rijckeghem, C and Weder, B, 2001. Bureaucratic corruption and the rate of temptation: do wages in the civil service affect corruption, and by how much? *Journal of Development Economics*, 65(2), pp. 307–331.

Yang, D, 2008. Integrity for Hire: An Analysis of a Widespread Customs Reform. *Journal of Law and Economics*, 51(1), pp .25–57.

World Bank: 2006, Proptery taxes in the Punjab, Pakistan, Technical report. URL: https://openknowledge.worldbank.org/handle/10986/8277

World Bank: 2009, Government of the Punjab Property Tax Decentralisation program: Scopeevaluation report, Technical report. URL: https://openknowledge.worldbank.org/handle/10986/12378

Publications in the 3ie Impact Evaluation Report Series

The following reports are available from http://www.3ieimpact.org/evidence-hub/impact-evaluation-repository

Impact of mobile message reminders on tuberculosis treatment outcomes in Pakistan, *3ie Impact Evaluation Report 44.* Mohammed, S, Glennerster, R and Khan, A (2016)

Making networks work for policy: Evidence from agricultural technology adoption in Malawi, 3ie Impact Evaluation Report 43. Beaman, L, BenYishay, A, Fatch, P, Magruder, J, Mobarak, AM (2016)

Evaluating the effectiveness of computers as tutors in China, 3ie Impact Evaluation Report 41. Mo, D, Bai, Y, Boswell, M and Rozelle, S (2016)

Effectiveness of a rural sanitation programme on diarrhoea, soil-transmitted helminth infection and malnutrition in India, 3ie Impact Evaluation Report 38. Clasen, T, Boisson, S, Routray, P, Torondel, B, Bell, M, Cumming, O, Ensink, J, Freeman, M, Jenkins, M (2016)

Evaluating the impact of vocational education vouchers on out-of-school youth in Kenya, 3ie Impact Evaluation Report 37. Hicks, JH, Kremer, M, Mbiti, I, Miguel, E (2016)

Improving maternal and child health in India: evaluating demand and supply strategies, 3ie Impact Evaluation Report 30. Mohanan, M, Miller, G, Forgia, GL, Shekhar, S, Singh, K (2016)

A triple win? The impact of Tanzania's Joint Forest Management programme on livelihoods, governance and forests, *3ie Impact Evaluation Report 34.* Persha, L and Meshack, C (2016)

Removing barriers to higher education in Chile: evaluation of peer effects and scholarships for test preparation 3ie Impact Evaluation Report 36. Banerjee, A, Duflo E, Gallego, F (2016)

Sustainability of impact: dimensions of decline and persistence in adopting a biofortified crop in Uganda 3ie Impact Evaluation Report 35. McNiven, S, Gilligan, DO and Hotz, C (2016)

Micro entrepreneurship support programme in Chile, 3ie Impact Evaluation Report 40. Martínez, CA, Puentes, EE and Ruiz-Tagle, JV (2016)

Thirty-five years later: evaluating the impacts of a child health and family planning programme in Bangladesh, 3ie Impact Evaluation Report 39. Barham, T, Kuhn, R, Menken, J and Razzaque, A (2016)

Can egovernance reduce capture of public programmes? Experimental evidence from India's employment guarantee, 3ie Impact Evaluation Report 31. Banerjee, A, Duflo, E, Imbert, C, Mathew, S and Pande, R (2015)

Smallholder access to weather securities in India: demand and impact on production decisions, 3ie Impact Evaluation Report 28. Ceballos, F, Manuel, I, Robles, M and Butler, A (2015)

What happens once the intervention ends? The medium-term impacts of a cash transfer programme in Malawi, 3ie Impact Evaluation Report 27. Baird, S, Chirwa, E, McIntosh, C, and Özler, B (2015)

Validation of hearing screening procedures in Ecuadorian schools, 3ie Impact Evaluation Report 26. Muñoz, K, White, K, Callow-Heusser, C and Ortiz, E (2015)

Assessing the impact of farmer field schools on fertilizer use in China, 3ie Impact Evaluation Report 25. Burger, N, Fu, M, Gu, K, Jia, X, Kumar, KB and Mingliang, G (2015)

The SASA! study: a cluster randomised trial to assess the impact of a violence and HIV prevention programme in Kampala, Uganda, 3ie Impact Evaluation Report 24. Watts, C, Devries, K, Kiss, L, Abramsky, T, Kyegombe, N and Michau, L (2014)

Enhancing food production and food security through improved inputs: an evaluation of Tanzania's National Agricultural Input Voucher Scheme with a focus on gender impacts, 3ie Impact Evaluation Report 23. Gine, X, Patel, S, Cuellar-Martinez, C, McCoy, S and Lauren, R (2015)

A wide angle view of learning: evaluation of the CCE and LEP programmes in Haryana, 3ie Impact Evaluation Report 22. Duflo, E, Berry, J, Mukerji, S and Shotland, M (2015)

Shelter from the storm: upgrading housing infrastructure in Latin American slums, *3ie Impact Evaluation Report 21*. Galiani, S, Gertler, P, Cooper, R, Martinez, S, Ross, A and Undurraga, R (2015)

Environmental and socioeconomic impacts of Mexico's payments for ecosystem services programme, *3ie Impact Evaluation Report 20*. Alix-Garcia, J, Aronson, G, Radeloff, V, Ramirez-Reyes, C, Shapiro, E, Sims, K and Yañez-Pagans, P (2015)

A randomised evaluation of the effects of an agricultural insurance programme on rural households' behaviour: evidence from China, 3ie Impact Evaluation Report 19. Cai, J, de Janvry, A and Sadoulet, E (2014)

Impact of malaria control and enhanced literacy instruction on educational outcomes among school children in Kenya: a multi-sectoral, prospective, randomised evaluation, 3ie Impact Evaluation Report 18. Brooker, S and Halliday, K (2015)

Assessing long-term impacts of conditional cash transfers on children and young adults in rural Nicaragua, *3ie Impact Evaluation Report 17*. Barham, T, Macours, K, Maluccio, JA, Regalia, F, Aguilera, V and Moncada, ME (2014)

The impact of mother literacy and participation programmes on child learning: evidence from a randomised evaluation in India, 3ie Impact Evaluation Report 16. Banerji, R, Berry, J and Shortland, M (2014)

A youth wage subsidy experiment for South Africa, 3ie Impact Evaluation Report 15. Levinsohn, J, Rankin, N, Roberts, G and Schöer, V (2014)

Providing collateral and improving product market access for smallholder farmers: a randomised evaluation of inventory credit in Sierra Leone, 3ie Impact Evaluation Report 14. Casaburi, L, Glennerster, R, Suri, T and Kamara, S (2014)

Scaling up male circumcision service provision: results from a randomised evaluation in Malawi, 3ie Impact Evaluation Report 13. Thornton, R, Chinkhumba, J, Godlonton, S and Pierotti, R (2014)

Targeting the poor: evidence from a field experiment in Indonesia, 3ie Impact Evaluation Report 12. Atlas, V, Banerjee, A, Hanna, R, Olken, B, Wai-poi, M and Purnamasari, R (2014)

An impact evaluation of information disclosure on elected representatives' performance: evidence from rural and urban India, 3ie Impact Evaluation Report 11. Banerjee, A, Duflo, E, Imbert, C, Pande, R, Walton, M and Mahapatra, B (2014)

Truth-telling by third-party audits and the response of polluting firms: Experimental evidence from India, 3ie Impact Evaluation Report 10. Duflo, E, Greenstone, M, Pande, R and Ryan, N (2013)

No margin, no mission? Evaluating the role of incentives in the distribution of public goods in Zambia, *3ie Impact Evaluation Report 9*. Ashraf, N, Bandiera, O and Jack, K (2013)

Paying for performance in China's battle against anaemia, *3ie Impact Evaluation Report 8*. Zhang, L, Rozelle, S and Shi, Y (2013)

Social and economic impacts of Tuungane: final report on the effects of a community-driven reconstruction programme in the Democratic Republic of Congo, *3ie Impact Evaluation Report 7*. Humphreys, M, Sanchez de la Sierra, R and van der Windt, P (2013)

The impact of daycare on maternal labour supply and child development in Mexico, *3ie Impact Evaluation Report 6*. Angeles, G, Gadsden, P, Galiani, S, Gertler, P, Herrera, A, Kariger, P and Seira, E (2014)

Impact evaluation of the non-contributory social pension programme 70 y más in Mexico, *3ie Impact Evaluation Report 5.* Rodríguez, A, Espinoza, B, Tamayo, K, Pereda, P, Góngora, V, Tagliaferro, G and Solís, M (2014)

Does marginal cost pricing of electricity affect groundwater pumping behaviour of farmers? Evidence from India, 3ie Impact Evaluation Report 4. Meenakshi, JV, Banerji, A, Mukherji, A and Gupta, A (2013)

The GoBifo project evaluation report: Assessing the impacts of community-driven development in Sierra Leone, *3ie Impact Evaluation Report 3*. Casey, K, Glennerster, R and Miguel, E (2013)

A rapid assessment randomised-controlled trial of improved cookstoves in rural Ghana, *3ie Impact Evaluation Report 2*. Burwen, J and Levine, DI (2012)

The promise of preschool in Africa: A randomised impact evaluation of early childhood development in rural Mozambique, *3ie Impact Evaluation Report 1*. Martinez, S, Naudeau, S and Pereira, V (2012)

In 2009, the Excise and Taxation department in Punjab implemented a series of human resource reforms to address low levels of revenue collection. These reforms were designed to incentivise tax collection and improve overall departmental performance. This study involved the design and evaluation of these performance pay packages. Researchers found that the incentive schemes produced substantial and unambiguous positive effects on revenue collection. Treatment tax circles outperformed control tax circles by a margin of over 12 percentage points in total collections over the two-year treatment period.

Impact Evaluation Series

International Initiative for Impact Evaluation
202-203, Rectangle One
D-4, Saket District Centre
New Delhi – 110017
India

3ie@3ieimpact.org
Tel: +91 11 4989 4444

www.3ieimpact.org