

Improving Tax Compliance without Increasing Revenue: Evidence from Population-Wide Randomized Controlled Trials in Papua New Guinea

CHRISTOPHER HOY

World Bank

LUKE MCKENZIE

Australian Treasury

MATHIAS SINNING

Crawford School of Public Policy, Australian National University

I. Introduction

Tax as a share of gross domestic product (GDP) and levels of tax compliance are substantially lower in the poorest countries of the world than in the richest countries (Besley and Persson 2014; Brockmeyer et al. 2019; Slemrod 2019). In these settings, authorities typically have less capacity to enforce tax legislation, and attempts to deter taxpayers from being noncompliant by threat of punishment are often not seen as credible (Kleven, Kreiner, and Saez 2018; Bergolo et al. 2019; IMF 2020). As a result, governments have begun exploring alternative, nondeterrence approaches to encourage taxpayers to comply, such as by providing them with information that tax revenue funds public goods and services (Mascagni 2018; Antinyan and Asatryan 2019). The effects of these “nudges” are likely to be inversely related to the cost of compliance faced by taxpayers, especially when enforcement is weak, because they may believe they can easily avoid being punished for noncompliance (Gordon

This study was preregistered with the American Economic Association RCT Registry (ID no. AEARCTR-0004056). We greatly appreciate the hard work of the staff at the Papua New Guinea Internal Revenue Commission (especially Jack Akia, John Baru, Kaia Fabila, and Patrick Urum) and at Abt Associates (particularly Iain Smith and Gordon Maitava) that made this study possible. We thank Robert Breunig, Anne Brockmeyer, Guillermo Cruces, Ryan Edwards, Christian Gillitzer, Rema Hanna, Stephen Howes, Ricardo Perez-Truglia, Fabrizio Santoro, Carlos Scartascini, Joel Slemrod, Teodora Tsankova, and participants at the 2020 Centre for the Study of African Economies Conference and the 2020 Midwest International Economics and Development Conference for helpful comments. We gratefully acknowledge financial support from the Australian Department of Foreign Affairs and Trade and the Australian Research Council (LP160100810). The views expressed in this paper are those of the authors and do not necessarily reflect those of the Australian government. Contact the corresponding author, Christopher Hoy, at choy@worldbank.org.

Electronically published November 2, 2023

Economic Development and Cultural Change, volume 72, number 2, January 2024.

© 2023 The University of Chicago. All rights reserved. Published by The University of Chicago Press.

<https://doi.org/10.1086/721650>

and Li 2009). However, to be able to meaningfully increase revenue, nudges would need to have a positive effect on taxpayers who face a high cost of compliance (i.e., those that have a large tax liability). This raises the question as to which types of taxpayers respond to credible, nondeterrence nudges in low-income and lower-middle-income countries and consequently whether these messages are an effective way to raise revenue.

We examine this issue by conducting two population-wide randomized controlled trials involving all firms registered to pay salaries and wages tax (SWT) and value-added tax (VAT) in Papua New Guinea (PNG).¹ PNG faces development challenges that are similarly faced by other resource-dependent, fragile states. It is a lower-middle-income country that has one of the highest rates of extreme poverty in Asia and ranks in the bottom 20% of countries globally in terms of the United Nations Human Development Index (UNDP 2019; World Bank 2019).² Tax revenue has remained at about 13% of GDP over the past 5 years, which is in line with the average for lower-middle-income countries (World Bank 2019). Tax compliance is extremely low. In 2018, about 95% of registered taxpayers did not file the required monthly SWT and VAT declarations with the tax authority, the Internal Revenue Commission (IRC), and no one has been imprisoned for noncompliance. To develop an understanding of what types of messages are likely to be effective in this environment, we conducted a survey of taxpayers, a series of focus groups, and extensive consultation with the IRC. This process revealed that deterrence messages were not likely to be seen as credible by most taxpayers, and there were genuine concerns that these types of nudges may even backfire because the tax authority is unable to follow through on threats of punishment for noncompliance.

Our two distinct but complementary trials test the effect of credible, non-deterrence nudges across all taxpayers and for multiple tax types through use of a range of communication channels at different points in time. Our first trial includes all taxpayers registered for SWT and VAT that could be contacted by mobile phone (23,489 enterprises), and they were randomly assigned to receive

¹ We use the term “population-wide” to refer to the universe of firms registered to pay SWT and VAT in PNG. This excludes firms that were not registered with the tax authority. In addition, we are able to contact only firms with valid phone numbers (in the first trial), mail addresses (in the second trial), and email addresses (in the second trial). Firms are required to provide their contact details at the time of registration, and it is possible that for some firms this information was inaccurate or out of date at the time our study took place. We were unable to determine the fraction of taxpayers who were successfully contacted. Thus, we present intent-to-treat effects throughout the paper, which underreport the corresponding average treatment effects on the treated (i.e., those who actually received the nudges).

² There were only four countries outside sub-Saharan Africa that had a lower Human Development Index ranking in 2018 than that of PNG (the Syrian Arab Republic, Afghanistan, the Republic of Yemen, and Haiti), three of which were undergoing severe conflict.

either a simple reminder (treatment group 1), information about the public benefits of paying taxes (treatment group 2), or no information (control group). The two treatment groups were contacted via SMS message 10 and 3 days before SWT declarations were due in April 2019. The second trial involved all taxpayers who were noncompliant for either SWT or VAT and were contacted by the IRC via post and email over a 9-month period (May 2019 to February 2020). Taxpayers with an odd tax identification number (TIN) received both a flyer highlighting the public benefits of paying taxes and a standard letter from the IRC. In both trials, we study the effects of the nudges on declarations filed and on the amount of tax paid. We are particularly interested in understanding the differences among three broad types of registered taxpayers: firms that had not filed previously (“nonfilers”), firms that had filed previously but claimed they had no tax liability³ (“zero filers”), and firms that had filed previously and paid tax (“nonzero filers”). These are the same groupings that were used by Santoro et al. (2020).⁴

Standard theoretical models of tax compliance (Allingham and Sandmo 1972) have been extended recently to illustrate that different types of taxpayers decide whether to comply on the basis of a broad range of costs and benefits, not just the risk of being punished for noncompliance (Slemrod 2019). We build on this existing theory by introducing the idea that in addition to financial costs (the amount of tax due), taxpayers face transaction costs associated with compliance (the time and effort spent on filing a declaration). Two predictions emerge from our model. First, taxpayers with a history of compliant behavior (i.e., zero filers and nonzero filers) are expected to be more responsive to nudges than those that have not filed before (i.e., nonfilers) because their familiarity with filing a declaration means they face a lower transaction cost of compliance. Second, nudges are expected to be more effective at increasing the number of declarations filed by taxpayers who have no tax liability (e.g., zero filers) because they face only transaction costs of compliance compared with those who face both transaction and financial costs (e.g., nonzero filers). Most previous studies in low-income and lower-middle-income countries have been unable to test these predictions and how nudges affect different types of taxpayers in general, because these studies have focused on comparing the effects of various messages on a narrow subset of taxpayers (Kettle et al. 2016; Mascagni, Nell, and Monkam 2017; Shimeles, Gurara, and Woldeyes 2017; Bergeron, Tourek, and Weigel 2023). In

³ They could have no tax liability for one of three reasons: they claim to have no employees (SWT), had insufficient sales to meet the threshold for paying VAT, or their sales were of an exempt class (VAT).

⁴ Santoro et al. (2020) alternatively label the groupings as “nonfilers,” “nil filers,” and “active” and have slightly different criteria for categorization given it is on the basis of an annually lodged tax.

addition, the results from existing research on tax compliance in upper-middle-income and high-income countries may not translate into lower-income settings characterized by low compliance rates, weak enforcement, and where third-party reporting is limited (Besley and Persson 2014; Pomeranz 2015; Kleven, Kreiner, and Saez 2018; Brockmeyer et al. 2019).

The findings from both trials illustrate that it is primarily taxpayers who face the lowest cost of complying with tax legislation that respond to credible, nondeterrence nudges. Our first trial reveals that both nudges almost doubled the number of on-time SWT declarations made by zero filers (i.e., those that had filed a declaration in the previous 15 months and claimed they had no SWT to pay because they have no employees). However, there were no significant effects on nonzero filers and nonfilers. Similarly, the findings of our second trial indicate that the inclusion of the flyer along with a standard letter from the IRC primarily increased the number of SWT and VAT declarations made by zero filers (i.e., those who previously reported they had no employees in the case of SWT and those who reported that their revenue stream was below the exemption threshold or were of an exempt class in the case of VAT). Thus, in both trials there was no change in the total amount of tax revenue collected.

We also conduct further analysis to reveal two key findings suggesting that being contacted (or being contacted in an unconventional way) by the tax authority may have led taxpayers to conclude they are under greater scrutiny. First, there were no statistically significant differences between the effects of the different treatments on filing on-time declarations in the first trial, even though one message provided details about due dates for specific tax types and the other provided only general information. Second, differences in declaration behavior between taxpayers in the treatment groups and the control group in the first trial were only short lived, suggesting that the effect we observe is an immediate response to having received a nudge as opposed to a sustained change in behavior. We also provide possible alternative explanations that may be driving the results, such as some of the messages could be lowering transaction costs or increasing the utility of taxpayers from compliance.

Our research contributes to the existing literature in several ways, which are in relation to the scale and setting of the study, the underlying mechanisms we uncover that drive the behavior of different types of taxpayers, and the innovative design of our trials. First, we analyze, to our knowledge, the first population-wide randomized controlled trial examining the effect of nudges on tax compliance in a low-income or lower-middle-income country.⁵ The

⁵ We are only aware of one other population-wide study in a low-income or lower-middle-income country, that by Santoro et al. (2020), who examine the effectiveness of letters from the tax office

population-wide nature of our trials allows us to provide insight into the effect of nudges when they are scaled up across all taxpayers, which makes the findings more relevant to policy makers and more generalizable (Muralidharan and Niehaus 2017; DellaVigna and Linos 2020). This also enables us to examine heterogeneity in the responsiveness of different types of taxpayers on the basis of their history of compliance and their location, which helps us to isolate the channels driving the findings we observe.

Second, we provide evidence across tax types, communication channels, and time periods supporting theoretical predictions that nudges primarily increase the declarations of taxpayers who face the lowest cost of compliance. We extend existing theory (Allingham and Sandmo 1972; Slemrod 2019) by introducing transaction costs into a simple model of tax compliance and theorize that nudges will have different effects on taxpayers depending on their compliance history. Our results are consistent with these predictions: we show that taxpayers who claim they have no tax liability (zero filers) are the most likely to respond to a nudge and that nonfilers are the least likely to respond. Thus, our study extends to a lower-income setting the trends observed by prior work in upper-middle-income and high-income countries; namely, that taxpayers who have the least to lose from compliance are often the most likely to respond to messages from tax authorities (Carrillo, Pomeranz, and Singhal 2017; Perez-Truglia and Troiano 2018).

Finally, our study contains several innovative design features. Our analysis is, to our knowledge, one of the first to test the effectiveness of nudges across different types of taxes that constitute a major source of revenue (SWT and VAT make up more than half of the total tax revenue collected in PNG; Mascagni 2018; OECD 2019). Our nudges were designed to be easily scalable and did not require tax authorities to make substantial changes to existing processes (Muralidharan and Niehaus 2017; Mascagni 2018). We also implemented our interventions in line with best practice by communicating with taxpayers via SMS messages and email, in addition to use of the traditional method of postal services (Brockmeyer et al. 2019; Ortega and Scartascini 2020; Ortega, Scartascini, and Mogollon 2020). When discussing our empirical findings, we pay particular attention to the first trial because it makes a greater contribution

in Eswatini. It was conducted in the second half of 2019, after ours. Research in these settings has been limited by only involving a sample of taxpayers in a particular city or a specific type of noncompliant taxpayers (Kettle et al. 2016; Mascagni, Nell, and Monkam 2017; Shimeles, Gurara, and Woldeyes 2017; Collin et al. 2021; Okunogbe 2021; Bergeron, Tourek, and Weigel 2023). These studies focused on comparing the effect of different kinds of nudges as opposed to the effect of nudges on different types of taxpayers (Mascagni 2018; Antinyan and Asatryan 2019).

to the literature about tax compliance of formal firms. There are three reasons we believe this trial makes a greater contribution to the literature. First, it consisted of multiple treatments, allowing us to compare the relative effectiveness of different types of messages. This enabled us to examine different motivations for why people pay tax and relate this to parameters in our model. Second, it allowed us to consider whether the treatments affect on-time filing as well as whether the effects were sustained over time. The issue of on-time filing in low-income and lower-middle-income countries has received less attention in the literature. Finally, it involved a new form of interaction between the IRC and PNG taxpayers, whereas the second trial consisted of adding a flyer to an existing interaction. Consequently, the first trial intervention was expected to be more influential.

The remainder of this paper is structured as follows. We provide details about the conceptual framework and summarize existing research on this topic in section II. We discuss the setting of the study in section III. We describe the study design and explain our empirical model in section IV. We present the findings of our analysis in section V. In section VI, we discuss the implications of our study and conclude.

II. Conceptual Framework and Related Literature

A. Conceptual Framework

Seminal theories of tax compliance argue that taxpayer behavior depends on the costs and benefits of tax compliance (Allingham and Sandmo 1972), similar to economic theories about crime (Becker 1968). Put simply, Allingham and Sandmo (1972) theorize that people choose between the certain cost of paying tax (t , which is assumed to be a fixed amount that represents a taxpayer's tax obligation) and the potential punishment they could face if they do not file (γp , where γ is the probability of being found noncompliant, and p is a fixed amount that represents the punishment a taxpayer will receive).

These theories have been extended in recent years to take into account broader reasons for why individuals may choose to pay tax (Slemrod 2019). For example, taxpayers may value the public benefits financed by tax revenue or the value of complying with social norms (Hallsworth 2014; Kettle et al. 2016). The utility gain from paying tax is represented formally as a in the model below. Recent studies have also recognized that taxpayer behavior is more likely to be determined by the taxpayer's belief about the probability of being punished, γ_b , as opposed to the actual probability, γ . We follow a simple model of tax compliance in the literature, discussed by Kettle et al. (2016), to illustrate that in a given reporting period, the individual utility associated with compliance, U_c , and noncompliance, U_n , can be written as

$$U_c = y - t + a$$

and

$$U_n = y - \gamma_b p,$$

where y is income before tax. According to this model, taxpayers comply if $U_c > U_n$, which requires that

$$t < \gamma_b p + a.$$

We extend this simple model by introducing the idea that there are transaction costs associated with filing a declaration, r , along with the financial costs of paying tax, t (Meiselman takes a similar approach; Meiselman 2018). We update the utility functions above accordingly and therefore predict that taxpayers will comply if

$$t + r < \gamma_b p + a.$$

Although this model is far from comprehensive, it provides a basic framework to illustrate the different channels through which nudges can affect taxpayers' behavior. Nudges can make the risk of punishment more salient for taxpayers (in our model, this would be consistent with increasing the term γ_b).⁶ Nudges may also provide taxpayers with information that increases their utility from compliance, such as letting them know the public benefits of paying tax or suggesting they will be abiding by social norms (in our model, this would mean increasing the term a). Another channel through which nudges may affect compliance is by reducing the transaction costs for taxpayers by reminding them of due dates or by simplifying the process (in our model, this would mean reducing the term r) (Hallsworth 2014).

Our simple model also helps us to illustrate how the effects of nudges are likely to vary according to the costs of compliance (both financial and transaction costs) faced by different types of taxpayers. Specifically, on the basis of the taxpayer's history of compliant behavior, we make predictions about the effect of nudges on the likelihood a taxpayer files a declaration. We consider three broad types of taxpayers: those that have not filed a declaration previously (nonfilers), those that have filed in a previous reporting period but claimed to have no tax liability (zero filers), and those that have filed in a previous reporting period and paid tax (nonzero filers).⁷ In the following, we consider the possibility

⁶ A more traditional approach of increasing the costs of noncompliance would be to directly increase the term p .

⁷ To be precise, we define not filing previously as not having filed a declaration in the past 15 months. We observe the filing behavior of taxpayers only in the 15 months before the trial.

that transaction costs associated with filing a declaration, t , beliefs about the probability of being punished, γ_b , and utility gains from paying taxes, a , may differ among nonfilers, zero filers, and nonzero filers. For taxpayers who have not filed a declaration in a previous reporting period, their compliance costs have always outweighed the benefits from compliance. Formally,

$$t + r^n > \gamma_b^n p + a^n,$$

where the superscript “n” refers to nonfilers. In contrast, zero filers and nonzero filers have filed a declaration in a previous reporting period (i.e., during the reporting period in which they filed $t + r^f < \gamma_b^f p + a^f$, where the superscript “f” refers to having filed a declaration in a previous reporting period). For simplicity, we assume that all taxpayers who have filed in a previous reporting period face the same transaction costs (r^f) resulting from the time taken to complete a declaration and file it with the tax authority.⁸ We also assume that nonfiling taxpayers face larger transaction costs than those faced by taxpayers that have filed in a previous reporting period (i.e., $r^f < r^n$) because they may not be familiar with how to complete a declaration and how to file it with the tax authority. Given these assumptions, the following two predictions emerge:

PREDICTION 1. A nudge will be more likely to lead taxpayers who have filed previously to file a declaration compared with taxpayers who have not filed previously.

This prediction is a consequence of nonfilers facing larger transaction costs than those faced by zero filers and nonzero filers (i.e., $r^f < r^n$). In addition, it is reasonable to believe that a higher share of nonfiling firms may no longer be in operation, and consequently no longer need to file a declaration, compared with firms that have filed recently (McKenzie and Paffhausen 2019).

PREDICTION 2. Among taxpayers that have filed previously, a nudge will be more likely to lead taxpayers that had no tax liability to file a declaration compared with taxpayers that had a positive tax liability.

This prediction is a consequence of the fact that the total cost of complying for taxpayers that claim to have no tax liability (zero filers) is equal to their transaction costs, r^f , while the total cost of compliance for taxpayers that have a positive tax liability (nonzero filers) is equal to their transaction costs plus financial costs from paying tax, $r^f + t$.

⁸ Arguably, transaction costs are lower for taxpayers who have no employees or sales because they are not required to make complex calculations to determine their tax liability. If this is the case, our second prediction is even more likely to hold.

The population-wide experiments that follow allow us to test these predictions.

B. Related Literature

Since the 1970s, there have been numerous studies examining the effectiveness of strategies used to increase tax compliance; however, the use of messages drawing on insights from behavioral economics has only recently been tested (Holzinger and Biddle 2016). Studies using nudges have illustrated that tax revenue can be increased through different types of messages that draw on a range of motivations to comply, including the risk of punishment, social norms of compliance, legal obligations, or the public benefits of paying tax (Halls-worth 2014; Mascagni 2018). An extensive number of trials have been conducted to test the effect of these types of nudges in high-income countries, and in general they show that deterrence nudges are more likely to raise revenue than are nondeterrence nudges (Antinyan and Asatryan 2019; Slemrod 2019). However, even in these settings, far less attention has been given to how the effect of nudges varies among different types of taxpayers (De Neve et al. 2019).

There is a growing body of literature about tax compliance in upper-middle-income countries, largely in Latin America, that examines various strategies that can be implemented even when enforcement is not as strong as it is in many high-income countries. In particular, there has been a strong focus on the potential of third-party reporting to improve tax compliance (Pomeranz 2015; Brockmeyer et al. 2019; Naritomi 2019). These studies in Chile, Costa Rica, and Brazil have consistently illustrated that when available, tax authorities can use third-party reporting to improve compliance. There is also growing evidence that different types of taxpayers respond differently to information from the tax authority (Castro and Scartascini 2015; Carrillo, Pomeranz, and Singhal 2017; Bergolo et al. 2019; Holz et al. 2020; Brockmeyer et al. 2021). While this body of work has greatly improved the knowledge base about tax compliance in a setting outside high-income countries, very little is known about whether these findings would extend to countries with lower incomes, high rates of noncompliance, weaker enforcement, and limited third-party information.

Only a small number of field experiments have tested how nudges affect tax compliance in low-income and lower-middle-income countries.⁹ Santoro et al.

⁹ In addition to the studies discussed here, a number of related studies have focused on the effects of “naming and shaming” taxpayers for noncompliant behavior or publicly celebrating those who are compliant in low-income and lower-middle-income countries (Chetty, Mobarak, and Singhal 2014; Slemrod, Rehman, and Waseem 2022).

(2020) conduct a population-wide trial of about 20,000 taxpayers in Eswatini in sub-Saharan Africa. They posted letters with different messages to both individuals and companies. They do not directly compare the effectiveness of the different messages. In some but not all settings, the different messages have similar results. They also consider heterogeneous effects of nudges on taxpayers depending on the taxpaying histories of the taxpayers. They find nonfilers to be most responsive to nudges. Although nonfilers begin to file, they do not begin to pay tax. Throughout, they find significant heterogeneity, emphasizing the importance of studying different types of taxpayers, as our work does.¹⁰

Collin et al. (2021) compare the effectiveness of various nondeterrence SMS messages at improving tax compliance in Dar es Salaam, Tanzania. Their three messages (simple reminder, shared benefit, and social pressure) each increased tax payment. In Guatemala, Kettle et al. (2016) examine the effectiveness of sending different kinds of letters to a small subset of noncompliant individuals and enterprises.¹¹ They find that simply receiving a letter increased declarations. Letters that highlighted the risk of punishment and letters that referred to social norms were the most effective in increasing tax revenue. Shimeles, Gurara, and Woldeyes (2017) show that hand-delivered letters, which refer to the risk of punishment or to the public benefits of paying taxes, increased the amount of tax paid by enterprises in Addis Ababa, the capital city of Ethiopia, in the short run. They restrict their sample to taxpayers who were not exempt from paying tax and had never been contacted by the tax authority before. Similarly, Mascagni, Nell, and Monkam (2017) find positive effects on tax paid by individuals and enterprises in Kigali, the capital city of Rwanda, from reminder SMS messages and emails that included either a punishment or a public-benefit motivation. Their analysis is limited to firms that have previously filed and recently registered their contact details with the tax authority. In Kananga, Democratic Republic of the Congo, Mascagni, Nell, and Monkam (2017) find a hand-delivered enforcement letter to be effective at increasing both compliance and revenue for property taxes compared with a control letter. The authors introduced randomized letters toward the end of their campaign, so their sample is restricted. Finally, Okunogbe (2021) analyzes the effectiveness of enforcement messages on property owners in Congo Town, a section of Liberia's capital city, Monrovia. By leveraging data from a new property

¹⁰ The definition of "nonfilers" by Santoro et al. (2020) is substantially different from what we use in our study. Specifically, they define nonfilers as taxpayers who did not file last year's income tax return; in our study, nonfilers did not file in each of the past 15 months, which is a much stronger indication of persistent noncompliance, and consequently they are less likely to be influenced by a nudge.

¹¹ Kettle et al. (2016) focus on all noncompliant taxpayers who opted in to pay profits tax, which captures only 3.5% of declarations.

database and threatening punishment, the revenue authorities were able to send enforcement messages that quadrupled the tax-payment rate.

III. Setting of the Study

A. Taxation in Papua New Guinea

Tax revenue as a share of GDP in PNG has been declining since its peak in 2011 and was at 13.7% in 2017, which is about the average for lower-middle-income countries (World Bank 2019). The three types of taxes that generated the highest shares of revenue in 2017 were SWT (34%), corporate income tax (26%), and VAT (21%) (OECD 2019). This tax structure is somewhat similar to that in comparable countries, except that PNG is particularly reliant on SWT.

All firms in PNG are expected to register with the IRC, and as part of this process they receive a TIN. It is not possible for firms to engage in international trade or bid for government projects without a TIN. The IRC regularly updates the tax register, and in 2018 the commission performed an extensive process of removing from the register those firms that were no longer operational.¹² At the start of 2019, 61.4% of firms in the IRC tax register were less than 5 years old.

All registered firms in PNG are required to file SWT declarations with the IRC by the seventh day of each month and VAT declarations by the twenty-first day of each month. SWT consists of taxes withheld from employees who are required to pay personal income tax; sole traders do not need to pay SWT (but they are still required to file a declaration). VAT is applied at the rate of 10% for the sales and import of goods and services in PNG in most sectors of the economy; however, there is a minimum revenue threshold below which firms do not have to pay VAT (but they are still required to file a declaration). Sales of certain types of goods and services are also exempt from VAT. The requirement for all registered firms to regularly file declarations regardless of whether they have a positive tax liability is common practice in most low-income and lower-middle-income countries (Santoro and Mdluli 2021).

To file a tax declaration, firms must complete a form that provides information about their business activities and tax liabilities over the previous month. An example of the declaration form that is required to be submitted to the IRC for SWT is provided in figure C1 (figs. A1, A2, B1, and C1–C3 are in the online appendixes). In the case of SWT, taxpayers are expected to calculate the withholding tax for each of their employees, while in the case of VAT, taxpayers are expected to identify how much of their revenue and expenditure is non-exempt. Accurately completing these calculations can be quite challenging for

¹² This involved more than 50 staff working full-time for 1 year contacting taxpayers to check whether they are still operational.

many taxpayers.¹³ The majority of SWT and VAT declarations made by small businesses and sole traders are filed in person at an IRC office. In the days immediately prior to due dates, there are long queues of taxpayers waiting to file their declarations.

The formal sector in PNG is largely concentrated in the capital and largest city, Port Moresby, which is the case in many low-income and lower-middle-income countries. The population of Port Moresby is about four times larger than that of PNG's second-most populous city, Lae, and is home to 56% of firms that are registered with the IRC. Furthermore, 94% of the declarations are processed through the IRC head office in Port Moresby because the commission has only three small offices located elsewhere throughout the country. Thus, firms located outside Port Moresby are substantially less likely to come into contact with the IRC, and, unsurprisingly, these firms have lower rates of compliance.

B. Tax Compliance in Papua New Guinea

Tax compliance has multiple dimensions. Three key dimensions are whether individuals and firms have registered with the tax authority for the relevant taxes, whether registered individuals and firms file declarations, and whether registered individuals and firms ultimately pay the correct amount of tax. The tax authority is able to accurately measure only the second of these because the extent of the other dimensions of compliance depends on information that is not reported directly. Consequently, we focus on measuring tax compliance by assessing whether registered firms file their SWT and/or VAT declarations (either on time or late). We also analyze the amount of tax paid.¹⁴

The rate of compliance among taxpayers in PNG is very low. More than 75% of firms registered for SWT and VAT did not file a declaration in 2018 (see figs. 1 and 2). We refer to the taxpayers that have not filed a declaration as “nonfilers” and to the remaining 25% of taxpayers, who have filed at least once in 2018 or in the first 3 months of 2019, as “filers.” To test the theoretical predictions discussed in section II.A, we estimate heterogeneous treatment effects on the basis of these categories to account for the possibility that these types of firms will respond differently to the nudges we provide.

¹³ The IRC provides guidance on its website; however, this may still be too advanced for many taxpayers. For example, in early 2019, a six-page guide for employers was released that explains how to calculate SWT for employees. The guide assumes among other things that all taxpayers have access to Microsoft Excel.

¹⁴ We do not consider firms that are not registered because our analysis is based on administrative records from the IRC.

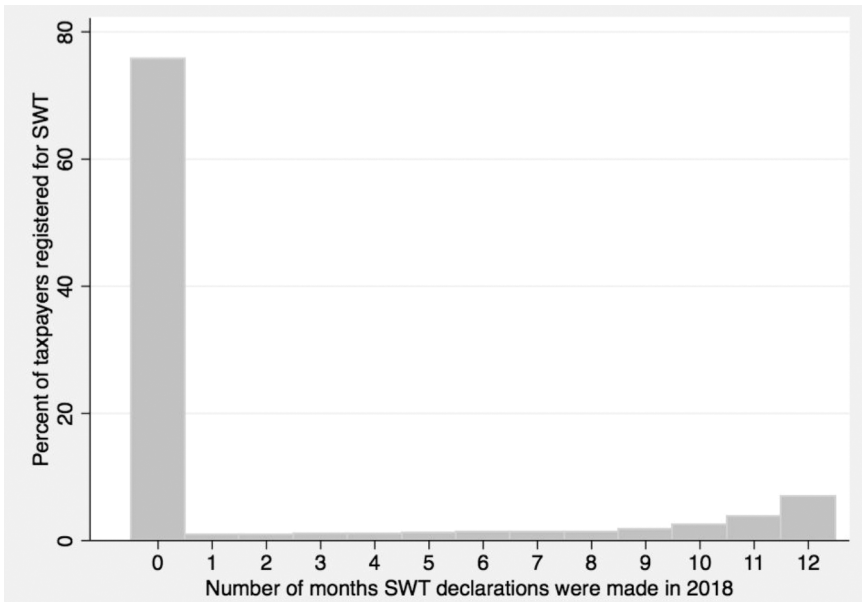


Figure 1. Breakdown of registered taxpayers by degree of SWT compliance. Seventy-six percent of taxpayers who were registered to pay SWT did not file a declaration in 2018, and only 7% filed a declaration each month.

We observe a considerable amount of heterogeneity in tax-compliance behavior among firms that have filed in the past 15 months. About 25% of these firms file a declaration every month (we refer to them as “regular filers”). The remaining 75% file an average of once every 2 months (we refer to them as “semi-regular filers”). Importantly, only about half of SWT and VAT declarations are for a non-zero amount. The reason for this is many taxpayers state in their declarations that they have no tax liability. This is the case for SWT when firms have no employees and for VAT when firms do not earn enough revenue from activities subject to VAT (or do not earn any revenue at all). To test the theoretical predictions discussed in section II.A, we examine heterogeneous treatment effects on the basis of whether taxpayers have previously reported that they had no tax liability (zero filers) or had a positive tax liability (nonzero filers) because we expect the responses of taxpayers to vary with the compliance costs they face.

Existing enforcement activities by the IRC are similar to those used by tax authorities in other low-income and middle-income countries (Mascagni 2018). First, the focus of compliance activities is primarily on ensuring that firms register and then on whether they file declarations. A lower priority is assigned to whether declarations are accurate. Firms that do not file a declaration have a small chance of receiving a letter in the post or via email asking them to file their overdue declaration. If they do not respond, three more letters are sent. The

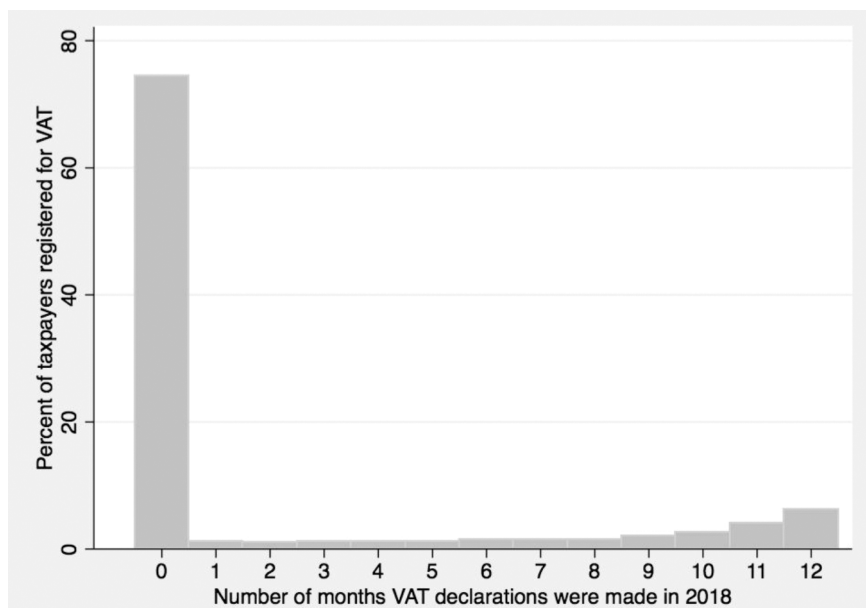


Figure 2. Breakdown of registered taxpayers by degree of VAT compliance. Seventy-four percent of taxpayers who were registered to pay VAT did not file a declaration in 2018, and only 6% filed a declaration each month.

mechanism for determining which firms are contacted is based on the number of outstanding declarations or the amount of debt overdue.

C. Local Explanations for Why Tax Compliance Is So Low

We conducted focus groups and a short survey in order to understand local explanations for why about 95% of registered firms in PNG do not file SWT and VAT declarations every month and to help inform the development of messages that would make firms more likely to be compliant.¹⁵ The survey was conducted in February 2019 and had 137 respondents, of which 58% was IRC staff that regularly interact with taxpayers and also pay tax. The remaining respondents were primarily owners of small and medium-sized enterprises registered with the IRC. The survey was not intended to be representative of the universe of taxpayers and IRC staff. The answers did not differ significantly between the types of respondents. We summarize below the main reasons why respondents believe firms do not pay tax and the main types of messages that they believe could change tax behavior. In appendix A, we provide more details about the implementation of this scoping survey.

¹⁵ Registering firms is a key focus of the IRC, and the anecdotal evidence is that a high share of all firms are registered.

1. Respondents' Views on Why People Do Not Pay Tax

Respondents had mixed views about why some Papua New Guineans do not pay the right amount of tax (see fig. A1). Uncertainty about how the tax revenue would be used was ranked as the most important reason. Respondents mentioned that many taxpayers are not aware that tax revenue helps to fund services that benefit the public, such as health and education. The other main reasons provided were that firms find paying tax too complicated, they forget to pay, or they think they can avoid paying tax and not face punishment.

2. Respondents' Views on What Messages Would Lead Firms to Pay More Tax

Two types of messages were expected to be the most effective in encouraging taxpayers to pay tax (see fig. A2). These were messages that reminded firms of their tax obligations ("reminder message") or that explained how the tax revenue was spent ("public-benefit message"). Messages that refer to social norms or highlight the risk of punishment were viewed as not being credible. More than five times as many respondents thought a reminder or public-benefit message would be effective compared with the number of respondents who thought a message that referred to social norms or that highlighted the risk of punishment would be effective.¹⁶

IV. Methodology

A. SMS Trial Design

This trial focused on the entire population of firms registered to pay SWT and VAT in PNG that had a valid mobile phone number (23,489 firms).¹⁷ The 23,489 firms were randomly assigned to treatment group 1 (7,828 firms), treatment group 2 (7,830 firms), and to the control group (7,831 firms). Consistent with our pre-analysis plan, stratified randomization was implemented by using the following set of baseline characteristics: location, industry, age of firm, size of firm, and an indicator for not having filed an SWT or a VAT return in 2018 (i.e., a measure of past compliance status). The randomization was done using the Stata command *randtreat* (ver. 1.4, April 5, 2017). The stratified randomization resulted in no statistically significant differences in baseline characteristics among the groups (see table B1; tables B1–B4 are available in the online appendixes). Our main outcomes of interest are the number of declarations filed on time and the amount of tax paid on time. We also use information about firms' taxpaying

¹⁶ Trials in high-income countries find deterrence messages can be effective in improving tax compliance (Antinyan and Asatryan 2019). It is possible that respondents could be underestimating the effectiveness of messages with unpleasant tones and overestimating more pleasant messaging.

¹⁷ Because we focus on firms with a mobile number listed with the IRC, some large enterprises are excluded because they provided the IRC with a landline phone number.

behavior over a period of 4 months from when the SMS messages were sent, which allows us to measure whether the treatment effects were only short lived or sustained over a longer period.¹⁸

The treatments we provide in both trials were informed by other studies (e.g., Mascagni, Nell, and Monkam 2017) and by extensive consultation with senior management of the IRC and focus groups (see sec. III.C). We did not include a deterrence nudge because research in low-income and lower-middle-income countries indicates that the effects of deterrence and nondeterrence messages are often similar (Mascagni 2018) and because our consultations revealed that the threat of punishment from the tax authority for noncompliance was not likely to be seen as credible by most taxpayers in this setting. We also did not include a nudge that referred to social norms of tax compliance, because informing firms that about 95% of taxpayers are noncompliant was unlikely to trigger a positive reaction. Moreover, because we were particularly interested in the effect of our nudges on different types of taxpayers—in contrast with previous research that focuses on the effects of different types of nudges—we did not consider deterrence or social-norms nudges to be essential.

The two types of nondeterrence messages provided to firms in the treatment groups aimed to address different reasons for why people do not pay tax (see sec. III.C). Treatment 1 provided firms with a reminder about what they were required to do to comply with tax legislation (fig. 3). This reminder SMS message was sent 10 days before SWT declarations were due (the relevant due date for our trial was April 7, 2019). Firms also received a second reminder SMS message 3 days before the due date.¹⁹ Treatment 2 provided firms with information about the public benefit of paying taxes (fig. 4). This SMS message is based on what has been used in other countries, but it was modified to suit the local context.²⁰ This message was also sent 10 and 3 days before SWT declarations were due.

B. Flyer Trial Design

This trial studied all firms that were contacted by the IRC via post and email for being noncompliant for either SWT or VAT over a 9-month period (May

¹⁸ The SMS bulk-service provider did not share information about unsuccessful SMS messages. As a consequence, the effects of the SMS trial presented in this paper may be interpreted as intent-to-treat effects, which may underestimate the treatment effects on the treated.

¹⁹ The second reminder was identical to the first reminder, but instead of “REMINDER” it stated “FINAL REMINDER.”

²⁰ Our SMS message is very similar to a message used by Mascagni, Nell, and Monkam (2017) in Rwanda. They used the following message: “By paying your taxes you make it possible to educate our children, fund our health care, and keep us safe. Pay taxes. Build Rwanda. Be proud.”

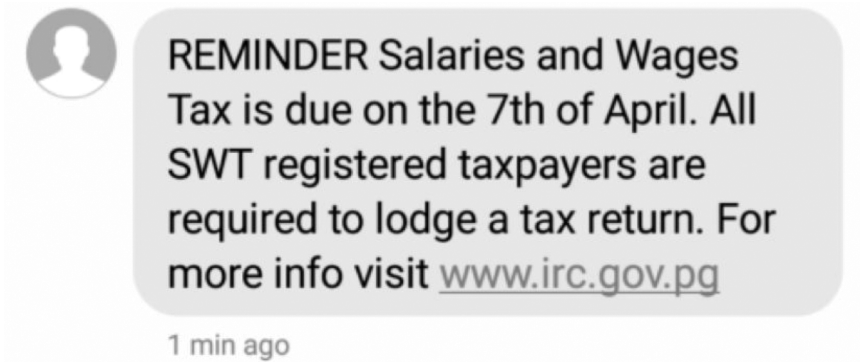


Figure 3. The SMS message that was sent from the IRC to taxpayers in treatment group 1 (reminder message).

2019 to February 2020).²¹ In total, the IRC sent 9,709 letters and emails to noncompliant taxpayers. We present the template of the letter (called a “demand notice”) that was sent via post and email in figure C2. It is likely that most of these taxpayers received demand notices previously (unfortunately the tax authority does not track how many letters have been sent to a taxpayer from the date of the taxpayer’s registration). Thus, they would already be quite familiar with this form of correspondence from the tax authority. To measure the effectiveness of adding a flyer to the demand notice, taxpayers were assigned to two groups on the basis of whether they had an odd or an even TIN.²² TINs are a sequence of randomly allocated numbers that were determined at the time of registration. As a result, we can exploit TIN allocation as a form of random assignment. Firms with an odd TIN received a flyer (the treatment group), and those with an even TIN did not (the control group). The demand notice includes language that attempts to deter taxpayers from being noncompliant. Our trial compares a treatment group receiving both a demand notice and a nondeterrence flyer with a control group receiving a demand notice only.

To implement the trial, we leveraged an existing random split in the compliance enforcement division of the IRC across two floors. The compliance enforcement team on one floor focused exclusively on firms with an odd TIN, and the team on the other floor focused exclusively on firms with an even

²¹ This trial was implemented after the SMS trial. We also examined the interaction between SMS and flyer treatments. Our findings indicate that the interacted treatments have no significant effect on our outcome variables. Note that at the end of February 2020, there were no known cases of COVID-19 in PNG. We have no reason to expect that businesses in PNG would have foreseen the enormous spread of COVID-19 in March 2020.

²² A potential strength of our study design is that, in contrast to the control group of the first trial, the control group of the second trial may be viewed as an “active control” or “placebo” because the demand notice is sent to taxpayers in the treatment group and taxpayers in the control group.

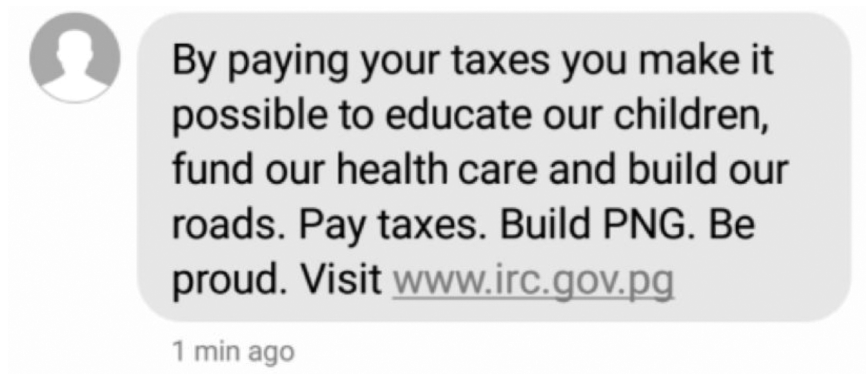


Figure 4. The SMS message that was sent from the IRC to taxpayers in treatment group 2 (public-benefit message).

TIN. This minimized the risk of an incomplete intervention due to taxpayers in the control group being treated or taxpayers in the treatment group not receiving the treatment.²³ As we were unable to stratify the randomization of the second trial, there was a slight imbalance between treatment and control groups (see tables B2 and B3). To address this issue, we control for observed characteristics in our regression analysis (see sec. IV.C below).²⁴ The outcome measures considered for this trial are the likelihood of filing an SWT or a VAT declaration with the IRC and the size of tax payments.

The flyer that was added to the demand notice from the IRC to firms in the treatment group drew upon a public-benefit motivation for why taxpayers should comply. This included information about how tax revenue helps to provide access to free education to more than 2 million children and free health care to all citizens. The flyer also used the same wording as the SMS message sent to treatment group 2 in the first trial (i.e., “educate our children, fund our health care and build our roads”). The flyer was sent as an attachment via email and was included in the envelopes containing the letters from the IRC. Each flyer had an English version printed on one side and the Tok Pisin (a near-universal dialect in PNG) version on the other side. A copy of the English version of the flyer is shown in figure C3.

²³ The enforcement division of IRC had previously been split across two floors with a nonrandom allocation of taxpayers between the teams. In early discussions with the IRC, one of the authors mentioned that allocating all the odd TINs to one team and all the even TINs to the other team would allow the IRC to monitor the activities and effectiveness of the two teams across two broadly similar groups of taxpayers. The IRC decided to implement this structure. Staff were randomly allocated to either the odd TIN or even TIN teams. This change occurred several months before the second trial began.

²⁴ We also estimate regression models without control variables (not presented in the paper). The results indicate that the inclusion of control variables in our regression models has almost no impact on the estimated treatment effects.

C. Empirical Model

To comply with international research standards, we preregistered our trial with the American Economic Association RCT Registry (ID no. AEARCTR-0004056). The most straightforward type of empirical analysis we conduct is an ordinary least squares (OLS) regression, which measures the effect of receiving a treatment compared with not receiving a treatment (i.e., the control group) on whether taxpayers filed a declaration and on the total amount of taxes paid.²⁵ In the first trial, our analysis involved combining both treatment groups and comparing members of this combined group with the control group. The results from our analysis provide information about the effect of receiving an SMS message on taxpayer behavior. In the second trial, we compare the behavior of taxpayers who received a public-benefit flyer (those with an odd TIN) and a standard deterrence letter from the IRC with the behavior of those who received only a standard deterrence letter from the IRC and not a flyer (those with an even TIN). In both cases, we control for a set of baseline characteristics (such as location, industry, and age of firm).²⁶

Our empirical strategy can be formalized as follows. To study the effect of treatment t on an outcome measure of interest, we estimate the model

$$Y_i^t = \beta_0^t + \beta_1^t T_i^t + \mathbf{X}_i^t \beta_2^t + \varepsilon_i^t, \quad (1)$$

where Y_i^t is a given outcome of taxpayer i , T_i^t is the treatment indicator for the comparison of treatment group t with the control group, ε_i^t is the model error term, and $\mathbf{X}_i^t$ is a vector of baseline characteristics.

We also use the baseline characteristics to estimate heterogeneous treatment effects. We are particularly interested in the history of tax compliance because treatment effects are predicted to vary on the basis of this dimension (see sec. II.A). We use baseline information about whether taxpayers filed a declaration during the period January 1, 2018, to March 27, 2019, and compare the treatment effect on nonfiling and filing firms. We also examine heterogeneous treatment effects between zero filers and nonzero filers on the basis of whether firms had previously reported to have a tax liability. Our ability to detect a

²⁵ As we consider only two outcomes of interest and heterogeneous treatment effects along a small number of prespecified dimensions, it appears very unlikely that multiple hypothesis testing changes the conclusions drawn from our results.

²⁶ We control for the baseline characteristics of taxpayers as opposed to the characteristics of firms reported in their tax declarations after the treatment. This is because after the nudge is provided, the treatment is not exogenous to the information taxpayers provide in their declarations. For example, the treatment may lead some firms to change their reporting of the number of employees they pay in their SWT declarations. Thus, we do not include the number of employees in the current period as a control variable, and we do not consider heterogeneous treatment effects along this dimension.

statistically significant effect is constrained by the number of taxpayers that fall into these categories.

We conduct three types of additional analyses to allow us to examine the mechanisms behind the effects we observe. First, in our first trial, we examine whether there are differences in responses of taxpayers on the basis of which treatment they received. To investigate this, we conduct separate regression models to compare the members of one of the two treatment groups with the control group. Second, to examine whether the effects of the treatments of the first trial are sustained over time, we conduct multiple OLS regressions on the basis of the model specified above, but we vary the time frame that we examine. Specifically, we analyze the cumulative treatment effect across a period of 2-week intervals from when the SMS messages were sent (i.e., 0–2 weeks, 0–4 weeks, 0–6 weeks, and so on).²⁷ Finally, to examine whether firms in Port Moresby, the capital city of PNG, respond differently than firms elsewhere in the country, we also explore heterogeneous treatment effects on the basis of the locations of taxpayers.

In appendix B, we present robustness checks of the treatment effects on the amount of tax paid (see table B4). Specifically, we winsorize this variable (at the 5th and the 95th percentiles) to reduce the influence of outliers. The results we present in the body of the paper are qualitatively similar to the results of these robustness checks.

V. Results

A. Treatment Effects on Previously Nonfiling and Previously Filing Firms

Table 1 shows the treatment effects on the number of SWT declarations and the amount of SWT paid by firms observed in the first trial (cols. 1 and 2), the effects on the number of SWT declarations and the amount of SWT paid by firms observed in the second trial (cols. 3 and 4), and the effects on the number of VAT declarations and the amount of VAT paid by firms observed in the second trial (cols. 5 and 6). In each case, we estimate the treatment effect for the full sample of all firms (panel A) and for the subsamples of previously filing and previously nonfiling firms (panels B and C). The numbers in panel A indicate that on average, the treatments had little effect on whether firms filed a declaration and on the amount of tax paid. We observe a statistically significant effect on SWT and VAT declarations made by all taxpayers only in the flyer trial, and the effect size is between 4.1 and 6.3 percentage points (between 21.4% and 27.2%).

The results in panel B of table 1 reveal that the treatments had a significantly positive effect on declarations of previously filing firms in both trials and across

²⁷ Our main results for the first trial examine only on-time payments, as this additional analysis captures the effect of the treatment over various time frames.

TABLE 1
TREATMENT EFFECTS BY TYPE OF FIRM

	SMS Trial		Flyer Trial			
	SWT Declarations (1)	SWT Paid (PGK) (2)	SWT Declarations (3)	SWT Paid (PGK) (4)	VAT Declarations (5)	VAT Paid (PGK) (6)
A. All Firms						
Any treatment	.030 (.019)	1,185 (1,964)	.041*** (.010)	4679 (5,055)	.063*** (.011)	−979 (1,424)
<i>p</i> -value	.120	.546	.000	.355	.000	.492
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Mean of dependent variable	.167	7,151	.192	14,849	.232	2,546
Observations	23,489	23,489	5,157	5,157	4,552	4,552
B. Previously Filing Firms						
Any treatment	.093** (.037)	3,986 (7,243)	.083*** (.020)	6,230 (11,116)	.111*** (.021)	−1,166 (1,394)
<i>p</i> -value	.013	.582	.000	.575	.000	.403
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Mean of dependent variable	.434	27,125	.370	30,851	.422	3,147
Observations	6,169	6,169	2,249	2,249	2,045	2,045
C. Previously Nonfiling Firms						
Any treatment	.006 (.022)	−72 (60)	.000 (.008)	2,530 (2,896)	.022** (.011)	−751 (2,348)
<i>p</i> -value	.782	.227	.968	.383	.036	.749
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Mean of dependent variable	.072	37	.054	2,474	.077	2,056
Observations	17,320	17,320	2,908	2,908	2,507	2,507
<i>F</i> -statistic (<i>p</i> -value)	.040	.552	.000	.567	.000	.837

Note. Shown are the treatment effects for all firms (panel A) and for the subsamples of previously filing firms (panel B) and previously nonfiling firms (panel C) of the first trial (cols. 1 and 2) and the second trial (cols. 3–6). The second trial represents a comparison of a deterrence letter and public-benefit flyer (treatment) to a deterrence letter (control). “Previously filing firms” is defined as having filed a declaration between January 1, 2018, and March 27, 2019. “Previously nonfiling firms” is defined as having not filed a declaration between January 1, 2018, and March 27, 2019. Robust standard errors are shown in parentheses. The *p*-value of the *F*-statistic refers to the difference in treatment effects between previously filing firms and previously nonfiling firms. PGK = Papua New Guinea kina (US\$1 was approximately equal to K3.4 on August 29, 2019).

** $p < .05$.

*** $p < .01$.

both tax types. The size of the treatment effect ranges from 8.3 to 11.1 percentage points (between 21.4% and 26.3%). However, we find no effect on the amount of tax collected. There was also no effect on previously nonfiling taxpayers, with the exception of a slight increase in VAT declarations made in the second trial (see panel C of table 1). In table 1, we also present the *F*-statistic that tests the statistical significance of the difference in treatment effects between

previously nonfiling and previously filing firms. In both trials, the difference in declarations is significant at conventional levels.

B. Treatment Effects on Previously Filing Firms Based on Whether They Previously Reported a Positive Tax Liability

Table 2 presents separate treatment-effect estimates for previously filing firms on the basis of whether they recently reported a nil tax liability (zero filers) or recently reported a positive tax liability (nonzero filers). We find a positive treatment effect on SWT declarations made by zero filers in both trials. The corresponding effects on nonzero filers are not statistically significant. In the first trial, the treatment increases the average number of declarations made by 17.7 percentage points (an increase of 55.7%). In the second trial, the treatment increases the number of declarations made by 3.3–4.3 percentage points

TABLE 2
TREATMENT EFFECTS ON ZERO FILERS AND NONZERO FILERS

	SMS Trial		Flyer Trial			
	SWT Declarations (1)	SWT Paid (PGK) (2)	SWT Declarations (3)	SWT Paid (PGK) (4)	VAT Declarations (5)	VAT Paid (PGK) (6)
A. Zero Filers						
Any treatment	.177*** (.060)		.043** (.022)		.033 (.023)	
p-value	.003		.050		.155	
Controls	Yes		Yes		Yes	
Mean of dependent variable	.318		.216		.256	
Observations	3,150		1,406		1,421	
B. Nonzero Filers						
Any treatment	.017 (.042)	12,680 (15,085)	.036 (.035)	−9,328 (28,518)	−.022 (.034)	−11,776* (6,386)
p-value	.684	.401	.293	.744	.515	.066
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Mean of dependent variable	.555	55,426	.626	82,306	.798	10,314
Observations	3,019	3,019	843	843	624	624
F-statistic (p-value)	.029		.937		.223	

Note. Shown are the treatment effects on zero filers (panel A) and nonzero filers (panel B) of the first trial (cols. 1 and 2) and the second trial (cols. 3–6). The second trial represents a comparison of a deterrence letter and public-benefit flyer (treatment) to a deterrence letter (control). “Zero filers” is defined as having filed a declaration between January 1, 2018, and March 27, 2019, but claimed to have a nil tax liability. “Nonzero filers” is defined as having filed a declaration between January 1, 2018, and March 27, 2019, and paid some tax. Robust standard errors are shown in parentheses. The p-value of the F-statistic refers to the difference in treatment effects between zero filers and nonzero filers. PGK = Papua New Guinea kina (US\$1 was approximately equal to K3.4 on August 29, 2019).

* $p < .10$.

** $p < .05$.

*** $p < .01$.

(an increase of 12.9%–20.0%). In the case of VAT declarations, the effect is not significant (the p -value is .155), but it is clear from examining the direction of the point estimates that the overall positive effect across previously filing firms is driven by zero filers. The negative effect on the amount of VAT paid is not robust to alternative specifications of this variable that we show in table B4. In table 2, we also present the F -statistic that tests the statistical significance of the difference in treatment effects between zero filers and nonzero filers. In the first trial, the difference is significant at conventional levels, but this is not the case in the second trial. The lack of a statistically significant difference could partly be a consequence of the substantially smaller sample size in the second trial.

C. Effects of Different Treatments in the First Trial

Table 3 presents separate treatment effects for the two nudges that were used in the first trial. We find that both treatments increase the number of on-time SWT declarations among previously filing firms, although the effect is only significant for the reminder treatment (the p -value for the public treatment is .167). The two treatments do not lead to a significant increase in the amount of SWT paid. The size of the effect on SWT declarations is between 7.4 and 11.1 percentage points (ranging from 18.0% to 26.1%), and the difference between the two treatments is not statistically significant. There are no effects from the treatments on previously nonfiling taxpayers. In table 3, we also present the F -statistic that tests the statistical significance of the difference in treatment effects between previously filing and previously nonfiling firms. For the reminder treatment, the difference in declarations is significant at conventional levels, but this is not the case for the public-benefit treatment.

D. Treatment Effects in the First Trial over Time

To study how the effect of the SMS messages evolves over time, we obtain the treatment effect observed after 2 weeks and then gradually add 2-week intervals to our sample to reestimate the effect. We also examine whether any differences existed between the treatment and control groups in the 4 weeks prior to the SMS messages being sent. The estimates presented in figure 5 reveal that the treatment effect of our first trial was short lived. We observe a significant and economically meaningful effect on on-time declarations of previously filing firms, which is 25% higher than the effect on the control group. However, within 4 months the difference between the treatment and control groups disappears entirely. This can be seen in figure 5, which shows the cumulative treatment effect at 2-week intervals from when the SMS messages were sent.²⁸

²⁸ Figure B1 shows the treatment effect within each 2-week period. The estimates indicate that the observed point estimate turns negative after 6 weeks.

TABLE 3
TREATMENT EFFECTS OF TRIAL 1 BY TYPE OF TREATMENT

	SWT Declarations (1)	SWT Paid (PGK) (2)
A. Reminder		
Previously filing firms:	.111*** (.043)	2,515 (6,944)
<i>p</i> -value	.010	.717
Controls	Yes	Yes
Mean of dependent variable	.426	25,234
Observations	4,145	4,145
Previously nonfiling firms:	.006 (.026)	-72 (61)
<i>p</i> -value	.810	.237
Controls	Yes	Yes
Mean of dependent variable	.071	49
Observations	11,514	11,514
<i>F</i> -statistic (<i>p</i> -value)	.036	.734
B. Public Benefit		
Previously filing firms:	.074 (.054)	5,097 (10,813)
<i>p</i> -value	.167	.637
Controls	Yes	Yes
Mean of dependent variable	.410	27,524
Observations	4,111	4,111
Previously nonfiling firms:	.006 (.026)	-73 (60)
<i>p</i> -value	.826	.222
Controls	Yes	Yes
Mean of dependent variable	.071	48
Observations	11,550	11,550
<i>F</i> -statistic (<i>p</i> -value)	.223	.576

Note. Shown are the effects of the reminder treatment on previously filing and previously nonfiling firms in trial 1 (panel A) and the effects of the public-benefit treatment on previously filing and previously nonfiling firms in trial 1 (panel B). "Previously filing firms" is defined as having filed a declaration between January 1, 2018, and March 27, 2019. "Previously nonfiling firms" is defined as having not filed a declaration between January 1, 2018, and March 27, 2019. Robust standard errors are shown in parentheses. The *p*-value of the *F*-statistic refers to the difference in treatment effects between previously filing firms and previously nonfiling firms. PGK = Papua New Guinea kina (US\$1 was approximately equal to K3.4 on August 29, 2019).

*** $p < .01$.

E. Differences in Treatment Effects across Papua New Guinea

Table 4 shows the treatment effects on previously filing taxpayers disaggregated by whether they are based in Port Moresby. We present the effects on the number of SWT declarations and SWT paid in the first trial (panel A), the number of SWT declarations and SWT paid in the second trial (panel B), and the number of VAT declarations and VAT paid in the second trial (panel C). We find that the treatment effect on SWT declarations is more than twice as large for previously filing firms outside Port Moresby in both trials. The effects of the

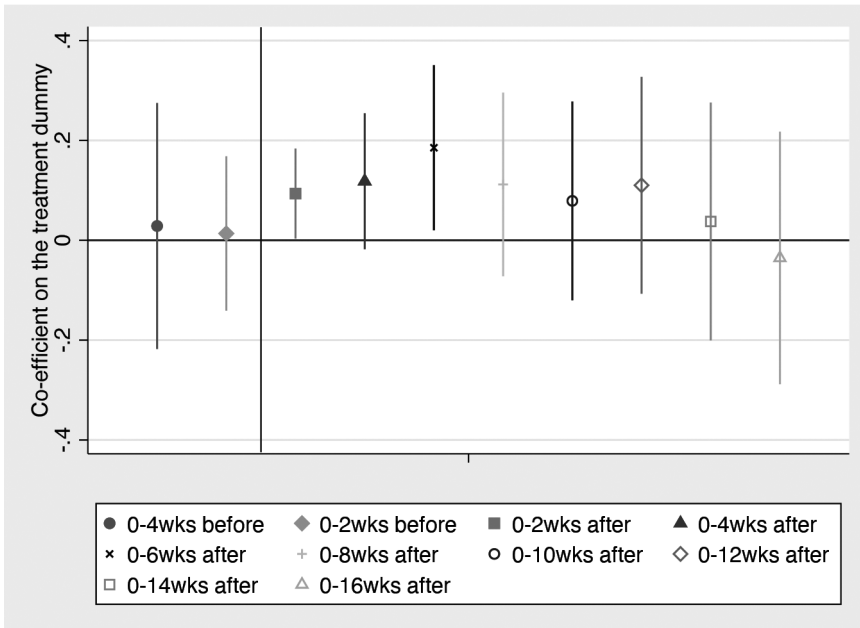


Figure 5. Treatment effects over time. Shown is the cumulative treatment effect at 2-week intervals in the 4 weeks before and 16 weeks after the SMS messages were sent. The coefficient on the treatment dummy of equation (1) (presented in sec. IV.C) is plotted on the y-axis. The vertical line indicates when the SMS messages were sent.

nudges as a proportion of the control group are particularly large because tax compliance is considerably lower outside Port Moresby. We observe treatment effects between 32.6% and 48.6% for previously filing firms outside Port Moresby and between 12.1% and 11.6% for previously filing firms in Port Moresby.

The F -statistics of the differences between treatment effects across samples presented in table 4 indicate that these differences are largely insignificant, with exception of treatment effects on SWT declarations in trial 2.²⁹ Despite this finding, we do observe a sizeable gap in treatment effects across samples.

VI. Discussion and Conclusion

A. Channel Driving Our Effects

Our results indicate firms that face the lowest cost from tax compliance are the most likely to respond to credible, nondeterrence nudges. These findings support the predictions of our model in section II.A where we build on standard theory by introducing the idea that taxpayers face transaction costs from

²⁹ We also perform regressions without control variables and find that the significant difference in treatment effects on SWT declarations in trial 2 may be attributed to the inclusion of control variables, suggesting that the slight imbalances between treatment and control groups observed for trial 2 may be responsible for the differences observed in cols. 1 and 2 of table 4.

TABLE 4
HETEROGENEOUS TREATMENT EFFECTS BY LOCATIONS OF PREVIOUSLY FILING FIRMS

	Declarations		Tax Paid (PGK)	
	Outside Port Moresby (1)	Port Moresby (2)	Outside Port Moresby (3)	Port Moresby (4)
A. Trial 1 (SWT)				
Treatment	.140** (.069)	.064 (.043)	7,607* (4,222)	6,338 (12,759)
p-value	.042	.136	.072	.619
Controls	Yes	Yes	Yes	Yes
Mean of dependent variable	.430	.437	11,559	37,303
F-statistic (p-value) ^a	.348		.925	
Observations	2,439	3,730	2,439	3,730
B. Trial 2 (SWT)				
Treatment	.138*** (.030)	.051** (.026)	2,889 (13,794)	8,813 (16,178)
p-value	.000	.047	.834	.586
Controls	Yes	Yes	Yes	Yes
Mean of dependent variable	.284	.423	16,968	39,334
F-statistic (p-value) ^a	.028		.129	
Observations	853	1,396	853	1,396
C. Trial 2 (VAT)				
Treatment	.125*** (.035)	.106*** (.027)	−30 (429)	−1,743 (2,114)
p-value	.000	.000	.944	.410
Controls	Yes	Yes	Yes	Yes
Mean of dependent variable	.331	.469	763	4,412
F-statistic (p-value) ^a	.664		.592	
Observations	709	1,336	709	1,336

Note. Robust standard errors are shown in parentheses. PGK = Papua New Guinea kina (US\$1 was approximately equal to K3.4 on August 29, 2019).

^a F-test of difference between treatment effects across samples.

* $p < .10$.

** $p < .05$.

*** $p < .01$.

complying, in addition to financial costs. In line with our first prediction, we find that previously filing taxpayers are more responsive to nudges than are previously nonfiling taxpayers. In both trials, the treatments had a positive effect on firms that had filed a declaration at least once in the previous 15 months but were not potent enough to change the behavior of firms that had an entrenched behavior of not paying tax. In support of our second prediction, we show that zero filers (i.e., taxpayers who reported no tax liability) are more responsive to nudges than are nonzero filers (i.e., taxpayers who had a positive tax liability). In both trials, there was no effect on the total amount of revenue collected because the increase in the number of declarations made was primarily driven by firms that reported a nil tax liability.

We provide evidence suggesting that the channel through which the treatments affect taxpayer behavior could be by leading them to perceive they are under greater scrutiny from the tax authority. We find very similar effects on filing on-time declarations and on the amount of tax paid from the two types of SMS messages in the first trial (see sec. V.C), even though one message explicitly refers to the specific due date for SWT (the reminder nudge), and the other message does not refer to dates or types of taxes (the public-benefit nudge). We also show that the differences in declaration behavior between taxpayers in the treatment group and taxpayers in the control group in the first trial appear to be purely an immediate response to having received a nudge as opposed to a sustained change in behavior. Collectively, these findings point toward the idea that taxpayers may well perceive nudges as a sign that they are under greater scrutiny.

Relating this proposed mechanism to our model in section II.A, by generating novel interactions between taxpayers and the IRC, the treatments we provided potentially made taxpayers believe they were under increased scrutiny from the IRC and/or the IRC was enhancing their compliance processes. This may have increased taxpayers' belief that they may receive punishment for noncompliance in the short term (i.e., the nudges may have increased γ_b).³⁰ This could explain why, in the first trial, we observe the same effect regardless of whether we attempt to decrease the transaction costs taxpayers face from compliance by use of the reminder message (r) or whether we attempt to increase the utility taxpayers gain from paying tax by use of the public-benefit message (a). This mechanism could also explain why the treatment effect disappears over time without further correspondence from the IRC: taxpayers may gradually reduce their perceived likelihood of receiving punishment resulting from noncompliance to a level similar to what it was previously. At a minimum, our results indicate that the absence of on-time filing of declarations is not due to taxpayers being unaware of due dates. If this were the case, we would have found a substantially larger effect from the reminder treatment that mentioned the due date than from the public-benefit treatment in the first trial.

B. Alternative Explanations

While the channel described in section VI.A provides a highly plausible explanation for the results, there are other alternative possible explanations.³¹ As

³⁰ While the content of the message is slightly different, Carpio (2013) provides evidence of non-deterrence messages increasing taxpayers' beliefs about the likelihood of being caught for noncompliance.

³¹ We do not observe any interaction between the treatments across the two trials. At first glance, this may seem counter to the explanation that the nudges led taxpayers to perceive they were under greater scrutiny, because one might expect that taxpayers who received treatments in both trials would be the

well as (or instead of) the nudges causing taxpayers to perceive they are under greater scrutiny, some of the messages could be lowering transaction costs or increasing the utility from being compliant.

For instance, while both SMS messages represented a novel interaction with the IRC and could have led taxpayers to believe enforcement effects were increasing, this may not be the case for the flyer trial, which simply involved adding to an existing piece of correspondence. This suggests the content of the flyer (a message about the public benefits from taxes) drove increased filing rates. This could be due to an increase in the utility taxpayers gained from paying tax by making them feel like they are contributing to public goods (a in our model).

Similarly, to explain the fact that the two SMS messages had similar results, despite one containing an explicit public-benefit message, it is possible that both messages, which included a link to the IRC's website, served as a reminder and source of more information. Both messages could have decreased the transaction cost taxpayers face for compliance (r in our model) by making it easier to access the necessary information.

Further, part of the reason why the public-benefit message may not have been more successful in improving tax collections in the SMS trial is that it did not provide evidence of actual benefits. Taxpayers may have assessed the promised benefits as unlikely to be realized, whereas in the flyer trial, specific instances of government spending were provided. This is consistent with a model presented by Krause (2020), who outlined that the motivation for an individual to comply depends on (among other factors) the level of enforcement and the value of the public goods being provided. In this model, in the absence of public goods or credible fines, increasing the enforcement effort increases evasion. Similar effects could be present in our trials whereby individuals were discouraged that their actual tax liability exceeded the level of government services they were receiving. In fact, this may be a cause of the large number of individuals claiming no liability. Ultimately, the exact mechanisms are very hard to pin down because of data limitations, and it is difficult to be able to rigorously compare the results of the two trials.

C. How Our Findings Relate to Previous Studies

Our analysis extends the frontier of knowledge on the use of nudges to improve tax compliance in low-income and lower-middle-income countries in several

most likely to respond. However, for the median taxpayer that received a treatment in both trials, there was an 8-month period between receiving the SMS and receiving the flyer. Given this is the case, it is extremely unlikely that there would be substantial interaction between the trials.

directions. For example, this is, to our knowledge, the first population-wide tax-compliance trial in a country with a low level of development, which allows for a thorough examination of how the effect of nudges varies between different types of taxpayers on the basis of their history of tax compliance and location. However, there are still important parallels between the findings of our research and those of previous studies.

First, similar to the findings of Santoro et al. (2020), we find significant variation in the way different types of taxpayers respond to nudges. However, our findings differ in important ways from those of Santoro et al. (2020). For instance, whereas we find zero filers to be most responsive to nudges, Santoro et al. (2020) find that nonfilers significantly respond to nudges, while nil and zero filers do not. A potential reason for this finding is the definition of “nonfilers.” Nonfilers in the study by Santoro et al. (2020) are taxpayers who did not file last year’s income tax return; in our study, nonfilers did not file in each of the past 15 months—a strong indication of persistent noncompliance, which is less likely to be influenced by a nudge. However, noticeably we find nudges to be ineffective at increasing the amount of tax paid. This is consistent with the study by Santoro et al. (2020), who find that their messages do not increase the payment probability of zero filers and the amount of tax paid by nonzero filers.

Second, our finding that correspondence from the tax authority can change filing behavior in the short run but not revenue has also been observed in previous studies (Kettle et al. 2016; Carrillo, Pomeranz, and Singhal 2017; Perez-Truglia and Troiano 2018). Carrillo, Pomeranz, and Singhal (2017) illustrate this occurred in Ecuador where in response to letters drawing on third-party information, firms increased their reported costs to largely offset the higher revenues that had been detected. Kettle et al. (2016) find that a simple reminder message increases declarations without increasing tax paid in Guatemala because the effect on declarations is driven by noncompliant firms that subsequently report they are exempt from paying tax. These findings are consistent with the explanation in our paper, whereby taxpayers facing lower total costs for compliance are the most likely to respond to nudges.

Third, the finding of similar effects of different types of nudges is consistent with the findings of some earlier research (Mascagni 2018; Slemrod 2019). In Tanzania, Collin et al. (2021) find simple reminder and shared-benefit messages to have similar effects, just as we did, although their messages increased tax payment. Further, in Ethiopia and Rwanda, Shimeles, Gurara, and Woldeyes (2017) and Mascagni, Nell, and Monkam (2017), respectively, find that deterrence and nondeterrence messages have similar effects on the likelihood of taxpayers filing a declaration and on the amount of tax paid. They note that in these settings, most

registered taxpayers do not receive any correspondence from their tax authority—in the case of Shimeles, Gurara, and Woldeyes (2017), their sample consisted of taxpayers that had never been contacted before. All these researchers argue that any nudge may make the risk of punishment for noncompliance more salient for taxpayers. We provide suggestive evidence that this is also the case for the trials we conducted. Future research could explore this issue in a low-income or lower-middle-income country setting in more detail, such as by conducting population-wide trials that directly compare the effects of deterrence and nondeterrence nudges.

D. Implications of Our Findings

Our results draw into question the value for money of credible, nondeterrence nudges aiming to generate substantial amounts of tax revenue across the population of taxpayers in low-income and lower-middle-income countries. While we do detect a sizeable effect on the likelihood of firms filing a declaration, this does not translate into greater tax revenue being collected. However, there are potentially other benefits that are difficult to quantify from taxpayers filing declarations. These include providing the government up-to-date information about economic activity and helping taxpayers to form habits so that if in the future they have a positive tax liability, they will continue to file declarations (and pay tax). Although the SMS messages and printing of flyers is relatively low cost compared with traditional methods of hiring and training more staff to audit taxpayers, they are not free. In total, the SMS messages sent as part of the first trial cost about US\$1,625 (K5,500), and the printing of flyers for the second trial cost about US\$500 (K1,700). Thus, it is highly unlikely the interventions we provided would pass a cost-benefit test.

By introducing the idea that taxpayers face transaction costs for compliance and by providing extensive experimental findings to support this, we illustrate that substantially different approaches are required to improve the compliance of various subsets of taxpayers depending on their previous tax-compliance behavior. Firms that are either nonfilers or nonzero filers did not respond to the treatments we provided. To be able to improve the compliance of these types of firms, tax authorities may need to rely on traditional methods, such as increasing punishment for noncompliance (this is captured as p in our model in sec. II.A) or increasing the likelihood that firms will be caught if they do not comply (this is illustrated by γ in our model in sec. II.A). Alternatively, governments in low-income and lower-middle-income countries could consider investing in raising the quality and quantity of third-party reporting and using personalized information to nudge taxpayers (Kleven, Kreiner, and Saez 2018; Brockmeyer et al. 2019).

References

- Allingham, M., and A. Sandmo. 1972. "Income Tax Evasion: A Theoretical Analysis." *Journal of Public Economics* 1, nos. 3–4:323–38.
- Antinyan, A., and Z. Asatryan. 2019. "Nudging for Tax Compliance: A Meta-analysis." ZEW Discussion Paper no. 19-055, ZEW–Leibniz Center for European Economic Research, Mannheim.
- Becker, G. 1968. "Crime and Punishment: An Economic Approach." *Journal of Political Economy* 76, no. 2:169–217.
- Bergeron, A., G. Tourek, and J. L. Weigel. 2023. "The State Capacity Ceiling on Tax Rates: Evidence from Randomized Tax Abatements in the DRC." NBER Working Paper no. 31685 (September), National Bureau of Economic Research, Cambridge, MA.
- Bergolo, M., R. Ceni, G. Cruces, M. Giacobasso, and R. Perez-Truglia. 2019. "Tax Audits as Scarecrows. Evidence from a Large-Scale Field Experiment." IZA Discussion Paper no. 12335, Institute of Labor Economics, Bonn.
- Beasley, T., and T. Persson. 2014. "Why Do Developing Countries Tax So Little." *Journal of Economic Perspectives* 28, no. 4:99–120.
- Brockmeyer, A., A. Estefan, K. R. Arras, and J. C. S. Serrato. 2021. "Taxing Property in Developing Countries: Theory and Evidence from Mexico." NBER Working Paper no. 28637 (May), National Bureau of Economic Research, Cambridge, MA.
- Brockmeyer, A., S. Smith, M. Hernandez, and S. Kettle. 2019. "Casting a Wider Tax Net: Experimental Evidence from Costa Rica." *American Economic Journal: Economic Policy* 11, no. 3:55–87.
- Carpio, L. Del. 2013. "Are the Neighbors Cheating? Evidence from a Social Norm Experiment on Property Taxes in Peru." Unpublished manuscript, Department of Economics, Princeton University.
- Carrillo, P., D. Pomeranz, and M. Singhal. 2017. "Dodging the Taxman: Firm Misreporting and Limits to Tax Enforcement." *American Economic Journal: Applied Economics* 9, no. 2:144–64.
- Castro, L., and C. Scartascini. 2015. "Tax Compliance and Enforcement in the Pampas: Evidence from a Field Experiment." IDB Working Paper no. IDB-WP-472, Inter-American Development Bank, Washington, DC.
- Chetty, R., M. Mobarak, and M. Singhal. 2014. "Increasing Tax Compliance through Social Recognition." IGC Policy Brief no. 31101, International Growth Centre, London.
- Collin, M., V. Di Maro, D. K. Evans, and F. Manang. 2021. "Property Tax Compliance in Tanzania: Can Nudges Help?" Global Working Paper no. 161, Brookings Institution, Washington, DC.
- DellaVigna, S., and E. Linos. 2020. "RCTs to Scale: Comprehensive Evidence from Two Nudge Units." NBER Working Paper no. 27594 (July), National Bureau of Economic Research, Cambridge, MA.
- De Neve, J., C. Imbert, J. Spinnewijn, T. Tsankova, and M. Luts. 2019. "How to Improve Tax Compliance? Evidence from Population-wide Experiments in Belgium." Working Paper no. 2019-07, Saïd Business School, Oxford.
- Gordon, R., and W. Li. 2009. "Tax Structures in Developing Countries: Many Puzzles and a Possible Explanation." *Journal of Public Economics* 93, nos. 7–8:855–66.

- Hallsworth, M. 2014. "The Use of Field Experiments to Increase Tax Compliance." *Oxford Review of Economic Policy* 30, no. 4:658–79.
- Holz, J. E., J. A. List, A. Zentner, M. Cardoso, and J. Zentner. 2020. "The \$100 Million Nudge: Increasing Tax Compliance of Businesses and the Self-Employed Using a Natural Field Experiment." NBER Working Paper no. 27666 (August), National Bureau of Economic Research, Cambridge, MA.
- Holzinger, L. Arcos, and N. Biddle. 2016. "Behavioural Insights of Tax Compliance: An Overview of Recent Conceptual and Empirical Approaches." TTPI Working Paper no. 8/2016, Tax and Transfer Policy Institute, Canberra.
- IMF (International Monetary Fund). 2020. "Revenue Administration Fiscal Informal Tool (RA-FIT)." Washington, DC: IMF. <https://data.imf.org/?sk=BA91013D-3261-42F8-A931-A829A78CB1EC>.
- Kettle, S., M. Hernandez, S. Ruda, and M. Sanders. 2016. "Behavioral Interventions in Tax Compliance: Evidence from Guatemala." Policy Research Working Paper no. 7690, World Bank, Washington, DC.
- Kleven, H., C. Kreiner, and E. Saez. 2018. "Why Can Modern Governments Tax So Much? An Agency Model of Firms as Fiscal Intermediaries." *Economica* 83:219–46.
- Krause, B. 2020. "Balancing Purse and Peace: Tax Collection, Public Goods and Protests." Unpublished manuscript, University of California, Berkeley.
- Mascagni, G. 2018. "From the Lab to the Field: A Review of Tax Experiments." *Journal of Economic Surveys* 32, no. 2:273–301.
- Mascagni, G., C. Nell, and N. Monkam. 2017. "One Size Does Not Fit All: A Field Experiment on the Drivers of Tax Compliance and Delivery Methods in Rwanda." ICTD Working Paper no. 58, International Centre for Tax and Development, Brighton.
- McKenzie, D., and A. Paffhausen. 2019. "Small Firm Death in Developing Countries." *Review of Economics and Statistics* 101, no. 4:645–57.
- Meiselman, B. S. 2018. "Ghostbusting in Detroit: Evidence on Nonfilers from a Controlled Field Experiment." *Journal of Public Economics* 158:180–93.
- Muralidharan, K., and P. Niehaus. 2017. "Experimentation at Scale." *Journal of Economic Perspectives* 31, no. 4:103–24.
- Naritomi, J. 2019. "Consumers as Tax Auditors." *American Economic Review* 109, no. 9:3031–72.
- OECD (Organization for Economic Co-operation and Development). 2019. "Revenue Statistics Asia and Pacific - Papua New Guinea." Paris: OECD. <https://www.oecd.org/countries/papuanewguinea/revenue-statistics-asia-and-pacific-papua-new-guinea.pdf>.
- Okunogbe, O. 2021. "Becoming Legible to the State: The Role of Detection and Enforcement Capacity in Tax Compliance." Policy Research Working Paper no. 9852, World Bank, Washington, DC.
- Ortega, D., and C. Scartascini. 2020. "Don't Blame the Messenger. The Delivery Method of a Message Matters." *Journal of Economic Behavior and Organisation* 170:286–300.
- Ortega, D., C. Scartascini, and M. Mogollon. 2020. "Who's Calling?: The Effect of Phone Calls and Personal Interaction on Tax Compliance." IDB Working Paper no. 10057, Inter-American Development Bank, Washington, DC.

- Perez-Truglia, R., and U. Troiano. 2018. "Shaming Tax Delinquents." *Journal of Public Economics* 167:120–37.
- Pomeranz, D. 2015. "No Taxation without Information: Deterrence and Self-Enforcement in the Value Added Tax." *American Economic Review* 105, no. 8:2539–69.
- Santoro, F., E. Groening, W. Mdluli, and M. Shongwe. 2020. "How to Best Nudge Taxpayers? The Impact of a Tailored Letter Experiment in Eswatini." ICTD Working Paper no. 112, International Centre for Tax and Development, Brighton.
- Santoro, S., and W. Mdluli. 2021. "Nil-Filing in Eswatini: Should the Revenue Administration Be Concerned?" African Tax Administration Paper no. 6, International Centre for Tax and Development, Brighton.
- Shimeles, A., D. Z. Gurara, and F. Woldeyes. 2017. "Taxman's Dilemma: Coercion or Persuasion? Evidence from a Randomized Field Experiment in Ethiopia." *American Economic Review* 107, no. 5:420–24.
- Slemrod, J. 2019. "Tax Compliance and Enforcement." *Journal of Economic Literature* 57, no. 4:904–54.
- Slemrod, J., O. U. Rehman, and M. Waseem. 2022. "How Do Taxpayers Respond to Public Disclosure and Social Recognition Programs? Evidence from Pakistan." *Review of Economics and Statistics* 104, no. 1:116–132.
- UNDP (United Nations Development Program). 2019. "Human Development Index." New York: United Nations.
- World Bank. 2019. "World Bank Country and Lending Groups." Washington DC: World Bank.