

Increasing tax compliance with behavioral insights: evidence from a RCT in São Paulo, Brazil

Flora Finamor Pfeifer*, Tainá Souza Pacheco†

Resumo

Na cidade de São Paulo, o Imposto Predial Territorial Urbano (IPTU) representa quase um terço da arrecadação municipal. Como uma parcela significativa da população atrasa o pagamento do tributo, a prefeitura arrecada sempre menos do que o previsto. Ainda que largamente utilizado por outros países, a aplicação de ciências comportamentais à políticas públicas ainda é uma área pouco explorada por governos brasileiros. Esse artigo apresenta os resultados de um experimento aleatório controlado (N= 15.178) realizado em 2019 na cidade de São Paulo que alterou a comunicação municipal sobre dívidas com o IPTU a fim de aumentar o número de pagantes. Os devedores foram aleatoriamente alocados em grupos, um dos quais recebia a comunicação viegente na Prefeitura, e outros cinco, cada qual com uma variante específica da carta redesenhada. Os resultados mostram que todas as cartas redesenhadas produziram resultado positivo, e em especial a carta que deixa claro aos municípios as consequências de estar em débito com o IPTU foi a que apresentou melhor resultado, aumentando a chance de arrecadação em 44.07 pontos percentuais (52.53%, comparado a taxa de 48.46% do grupo de controle), um aumento de 8,4%. Desntro da amostra do experimento, o aumento de regularização gerado pelas cartas redesenhadas corresponde a um aumento de R\$ 0,95 milhões. Se a “carta das consequências” for implementada pela Secretaria de Finanças, estima-se que possa gerar um aumento de arrecadação da ordem de R\$ 60 milhões, ou 1.200 vezes o custo do experimento.

Palavras-chave: cumprimento fiscal, nudge, ciências comportamentais

Abstract

In the city of São Paulo, property tax represents almost one-third of municipal revenue. Nevertheless, the amount collected yearly is beyond the total due to collection: a significant part of taxpayers (12,67% in 2018) fails to pay their taxes on time despite receiving an official reminder letter. Therefore, improving tax compliance stands a priority for São Paulo’s municipal government and its Treasury Department. Behavioral insights or nudges have yielded great benefits for today’s public administrators by improving the quality of official messages and increasing revenue flows. However, this has never been rigorously tested in Brazil. This paper presents results from a large (N = 15,178) randomized controlled trial in São Paulo that redesigns the communication to promote tax compliance. The trial varied the letter received by taxpayers who had failed to pay their property tax for the 2019 tax year. Taxpayers were randomly allocated to receive either the letter originally used by the São Paulo’s Treasury Department or five letter variants adapted using behavioral design. The study finds that although all letters increased the rate of late payment, only two of the letters showed statistically significant results. The best performing treatments were an explanatory message informing taxpayers about the costs and consequences of non-compliance, and a deterrent message framing non-declaration as an intentional and deliberate choice. The “Consequences Letter” increased the regularization rate by 4.07 percentage points (52.53%, compared to 48.46% in the control group), which stands for an 8.4% increase. By increasing the regularization rate, the experiment itself has brought an increase of R\$ 0.95 million in revenue. The paper estimates that if sent to all taxpayers in the city, the “Consequences Letter” could generate additional tax revenues of approximately R\$ 60 million, which is 1,200 times the cost of redesigning the letter.

Keywords: tax-compliance, nudge, behavioral public policy

*Fundação Getulio Vargas, florafinamorpfeifer@gmail.com

†Affiliated Researcher at 011.lab, São Paulo City Hall, pacheco.taina@gmail.com

1 Introduction

Reducing tax evasion stands a priority for local governments around the globe. The policy tool kit cities can implement to improve tax compliance goes from traditional measures, such as tax legislation and tax administration reforms, to behavioral interventions. Those measures can be complementary to each other: behavioral insights can help tax authorities to align its strategy more accurately to taxpayer behavior. The Sao Paulo Tax Authority was interested in testing if applying behavioral insights to its communication with taxpayers could promote tax compliance.

An increasing body of evidence shows that behavioral science can improve compliance in the fiscal area (Bérgolo et al. 2017, Castro and Scartascini 2015, Chetty and Saez 2013, Coleman 2007, Del Carpio 2013, Dwenger et al. 2016, Hallsworth et al. 2017, Hernandez et al. 2019, 2017, Kettle et al. 2016, Kleven et al. 2011, Mascagni 2015, Ortega and Sanguinetti 2013, Pomeranz 2015, Wenzel and Taylor 2004). In particular, behavioral insights can be applied through simple, cheap, and scalable interventions, also known as 'nudges' (Thaler and Sunstein 2008). Fiscal departments around the world have redesigned their communication with taxpayers including behavioral aspects in the choice architecture, usually with cost-efficient results. Nevertheless, this has never been rigorously tested in Brazil.

In the city of São Paulo, the property tax (Imposto Predial Territorial e Urbano - IPTU) amounted to 30 billion Reais (R\$) in 2018, almost one-third of municipal revenue. Nevertheless, the amount collected yearly is beyond the total due to collection: a significant part of taxpayers (12.67% in 2018 according to São Paulo's Treasury Department) does not pay on time and acquire debts with the municipality, with a daily interest of 0.33%. According to the Treasury Department, the volume of taxes not paid in 2018 summed R\$ 1.67 billion (around US\$ 420 million). Therefore, reducing tax evasion stands a priority for São Paulo's municipal government and its Treasury Department.

This paper presents the result of a randomized controlled trial (RCT) that used letters to remind taxpayers in the city of São Paulo to pay their taxes. The trial varied the letter received by 15,178 debts from 12,310 taxpayers who had failed to pay their property tax for the 2019 tax year. Taxpayers were randomly allocated to receive either the letter originally used by the São Paulo's Treasury Department or five letter variants adapted using behavioral design.

While many authors and tax authorities use behavioral interventions to increase tax compliance, there is no consensus in the literature about what behavioral principle works the best in reaching that goal. Therefore, we designed a randomized control trial using behavioral principles from the literature to understand what works best for the city of São Paulo.

The results show that two of the letters significantly increased regularization rates among late taxpayers when compared to the control group. The best performing treatments were an explanatory message informing taxpayers about the costs and consequences of non-compliance, and a deterrent message framing non-declaration as an intentional and deliberate choice. The "Consequences Letter" increased the regularization rate by 4.07 percentage points (52.53%, compared to 48.46% in the control group), which stands for an 8.4% increase. By increasing the regularization rate, the experiment itself has brought an increase of R\$ 950 thousand in revenue. The paper estimates that if sent to all taxpayers in the city, the "Consequences Letter" could generate additional tax revenues of approximately R\$ 60 million, which is 1,200 times the cost of redesigning the letter.

This experiment provides several contributions to policy making and to the literature on behavioral interventions in tax compliance. It is the first randomized control trial with behavioral tax communications in Brazil. It indicates that some behavioral interventions can be effective in reducing property tax debts in the context of the city of São Paulo, thus increasing municipal revenue, with marginal additional costs. It also shows that some behavioral principles effective elsewhere, such as social norms messages, may not work in this context.

This paper proceeds as follows: section 2 reviews the literature that apply behavioral insights to tax compliance in similar contexts. Section 3 discusses current approaches on how to increase compliance based on the standard tax evasion model and recent developments in behavioral economics related to the effects of tax morale. Section 4 describes the São Paulo context and our field experiment. Section 5 presents our results and the cost-benefit analysis. Section 6 concludes.

2 Background

Since most traditional tax policies have been based on classic economic models of taxpayers as decision-makers, governments can benefit from applying behavioral science insights into fiscal policies (Weber, Fooker and Herrmann 2014). Experiments within governments aimed to test the empirical validity of such principles. Treasury departments around the world redesigned communications to taxpayers taking into account behavioral aspects and rigorously tested their impacts. A growing body of evidence from the last two decades indicates that behavioral insights are an useful tool to accelerate and/or increase tax compliance (Bérgolo et al. 2017, Castro and Scartascini 2015, Chetty and Saez 2013, Coleman 1996, Del Carpio 2013, Dwenger et al. 2016, Hallsworth et al. 2017, Hernandez et al. 2019, 2017, Kettle et al. 2016, Kleven et al. 2011, Mascagni 2015, Ortega and Sanguinetti 2013, Pomeranz 2015, Wenzel and Taylor 2004).

In this section, we will expose our findings from a literature review of field experiments with behavioral insights in government tax letter delivered by mail. We focused on this communication channel because it is the object of our study, although similar experiments have been conducted using text messages and emails as a way to communicate with taxpayers.

The first field experiment involving letters to taxpayers was in 1995, with the Minnesota Department of Revenue. On that experiment, the redesigned letters included messages appealing to the social norm of compliance (Coleman 1996). A similar experiment in the United Kingdom also showed that adding tax morale principles in the letters, such as social norms and public goods messages, could be effective in increasing payment rates of the income tax. The experiment tested a few varieties of social norms and public goods framing, and the best results came from the message that framed the late taxpayer as being part of the minority, with a 5.1% increase in compliance (Hallsworth et al. 2017).

An experiment in municipalities in Lima, Peru, also indicates that social norms message might be efficient in increasing property tax payment (Del Carpio 2013). While social norms messages increased payment in 20%, deadline reminders increased in 10%, and deterrence messages did not work. Similarly, Ariel (2012) tested only deterrence messages in Israelian communications with firms and did not find an impact. However, Kleven et al. (2011) tested the same in Denmark, but this time with individuals, and found positive results over the income tax payment rate.

Other randomized trials tested the efficacy of both tax morale and deterrent messages and got positive results for both approaches. Hallsworth et al. (2017), found that descriptive social norms work best than injunctive social norms, and that information about the pecuniary consequence (the fine value) might also work as a payment motivator for the income tax. In Guatemala, Kettle et al. (2016), got the best results in increasing the income tax payment rates with social norms and deliberate choice letters - both tax morale and deterrence approaches were efficient in that context. Simple reminders and national pride messages did not work that well. Mascagni, Nell and Monkam (2017) investigates the heterogeneous effects of such messages in Rwanda. Soft tones messages, such as reminders and public goods, work best in general, but deterrence approaches are more efficient with the lower-income population. In Poland, Hernandez et al. (2017) showed positive impacts with all versions tested, but the most efficient one combined omission (deliberate choice) with a deterrence message. In Kosovo, both reminders with a call to action and deliberate choice messages had positive results, with a 73% and 44% increase in the payment rate, respectively. Nevertheless, In Germany, neither did deterrence nor social norms messages worked to increase compliance with the church tax. Only a simplified

version and a version that contained messages over social and monetary benefits increased it (Dwenger et al. 2016).

Contrarily, a handful of experiments that tested both approaches showed that only deterrence messages work. In Australia, an experiment over the rental properties tax found that hard tones messages work, whether soft tone messages did not (Wenzel and Taylor 2004). Three experiments in South America, in Uruguay, Chile, and Sucre, Venezuela, tested the impact of such messages over business compliance with the service tax (VAT). While deterrence messages involving the probability of audit and information about the fine increased compliance, messages involving public goods, ethics, and only simplified versions did not work (Bérgolo et al. 2017, Ortega and Sanguinetti 2013, Pomeranz 2015). In Bangladesh, a similar trial with companies aimed to check whether some type of social recognition message worked: information about one's group, reward cards (gold, silver, and bronze), or sharing one's compliance info with one's peer group. The latter was the only one that worked, and it is a kind of deterrence (Chetty and Saez 2013). Finally, a trial in Junin, Argentina, with the property tax of individuals and businesses showed that only a deterrence letter worked, increasing payment in 10%, while messages eluding to social norms (equity) or public goods (justice) had no impact (Castro and Scartascini 2015).

As the paragraphs above illustrated, there is no consensus in the literature about what behavioral principle works the best in increasing tax compliance. The results vary a lot depending on the context the experiment was implemented. The trials reviewed vary on a multitude of dimensions, that may difficult our attempt to generalize conclusions. They deal with different tributes and with different target populations (individuals, companies, or both) and they vary on the scale (municipality, state, or federal level). Additionally, there is no sense of unity on how the behavioral principles are materialized, so the choice of sentence, words, and the connotation itself, even though it has a similar approach, might greatly differ. They also vary on the control group (old letters, no letters, simplified letters) and on the dependent variable being measured (payment rate on time, payment rate in general, declaration rate). Finally, although we only mentioned randomized control trials in this review, they might be designed and randomized differently, with methodology differences that may impair our comparisons efforts. Apart from that, differences in effect might arise due to societal and cultural differences among the countries analyzed. Kettle et al. (2016) emphasize that this variation can be due to the heterogeneity in contexts, such as recipients, tax type, trust in institutions, social norms, and tax culture. *"The effectiveness of a treatment may depend on seemingly insignificant elements of its presentation, highlighting the importance of isolating interventions and rigorous testing of small adaptations to how messages are framed in particular contexts"* (Kettle et al. 2016, p.5). Similar conclusions aroused from the World Bank Report, that gathered the findings from Kosovo, Poland, Guatemala, and Costa Rica: *"the evidence presented in this policy note confirms that context is key when it comes to motivating taxpayers. For example, reminders informed by social norms work well in one place, while punitive language performs better in others"* (World Bank 2018). Thus, it highlights the importance of testing by context.

3 Theory

Based on the expected utility theory, Allingham and Sandmo (1972) introduced a framework to analyse tax evasion as a risky choice. Their model assumes that taxpayers are homogeneous and act rationally, money maximising and selfish. Tax payers choose to pay (the safe choice) or to evade (the risk choice). Yitzhaki (1974) incorporates uncertainty to the Allingham-Sandmo model, making it possible to analyse the effect of changes in the probability of being detected, the fine rate and the tax rate.

Some authors developed modifications of the Yitzhaki version of the Allingham-Sandmo model, many of whom incorporate behavioral aspects to the traditional literature. Hashimzade, Myles and Tran-Nam (2013) identify two classes of contribution in the behavioural economics literature. The first uses the non-expected utility, which permits a broader range of possibilities for the structure of preferences and removes some of the constraints imposed by expected utility theory. The second class of contributions alter the payoff structure,

incorporating social interactions, such as the sense of justice, social norms, group effects, and prestige. This class of models understand that tax morale (what other taxpayers are doing and how the government uses public revenues) is an important motivator to tax compliance.

We follow the approach used by Kettle et al. (2016), as they incorporate both deterrence and non-deterrence approaches. According to the authors, incorporating the non-deterrence factors in the utility of the individual as “moral costs” allows “*preserving the theoretically tractable assumption of utility maximization while, we contend, increasing the real-world applicability*” (Kettle et al. 2016, p.6). Formally, we can model the non-payment decision as follows:

$$E(U) = \Pi(p)xc(f_1, m) + \Pi(1 - p)xc(f_0, m) \quad (1)$$

And the payment decision as follows:

$$E(U) = -xb(m) \quad (2)$$

where $\Pi(p)$ is the perceived probability of being punished; $\Pi(1 - p)$ is the perceived probability of getting away with some punishments; $c(f_1, m)$ is the utility function of the moral and monetary costs associated in the worst-case scenario; $c(f_0, m)$ is the utility function of the moral and monetary costs associated with the best-case scenario; $b(m)$ is the moral benefits of paying; and x is the amount owned.

The government could use the following strategies as policy tools:

- increase the perceived normative costs of punishment ($c(f_1)$);
- increase the moral costs of noncompliance ($c(m)$);
- increase the perceived likelihood of the severity of punishment $\Pi(p)$; and
- increase the moral benefit of compliance ($b(m)$).

Following Kettle et al. (2016), we will relax the assumption of full attention to the costs, which suggests that interventions such as reminders and instructions of how and when to pay can also be used. The redesigned letters will rely on this model.

4 Experimental Context and Design

4.1 Context description

The property tax (IPTU) stands as the second major source of revenue for São Paulo’s municipal government. It corresponds to almost one-third of the total amount collected. In 2018, this value was close to 13 billion reais, according to the Treasury Department. The IPTU is issued annually: the tax authority sends letters to taxpayers between the end of January and the end of February, notifying taxpayers about the amount charged. The taxpayer, this is, the property owner, can be a person or a company. The IPTU can be payed in monthly installments (from 2 to 10 installments), or to pay in a single installment, earning a 3% discount.

In 2018, the default rate calculated by the Treasury Department was 12.67%, which is equivalent to R\$ 1.67 billion (Dayrell 2018). The delay incurs a daily interest of 0.33%. Besides this direct pecuniary penalty, when a person owns a debt with the Treasury Department, he/she/it enters the Municipal Informative Register (Cadastro Informativo Municipal - CADIN).

As a consequence of being included in CADIN, the Municipal Administration bodies and entities are prevented from performing the following acts, concerning the individuals and firms to which it refers: (i) entering into agreements, adjustments or contracts involving the disbursement, under any heading, of financial resources;

(ii) transfers of agreement amounts or payments related to contracts; (iii) concession of aid and grants; (iv) granting tax and financial incentives. Besides, the records are subject to future legal actions and property attachment.

As mandated by law, before joining the CADIN, the late taxpayer receives a notification letter informing that if the debt is not regularized within 30 days, his/her/its name will be included in CADIN. This letter, issued weekly by the treasury department, contains details of the amount due, instructions on how to pay, and information about possible problems in the property registration. The property owner can regularize the situation by paying the debt or by correcting any problem in the property registration. Despite the warning, more than half of the taxpayer on debt does not regularize their status on time.

4.2 Experimental Treatments Design

We conducted qualitative research in order to collect general perceptions about the original letter and formulate hypotheses about the drivers of the non-compliance behavior. We conducted field research and semi-structured interviews with the Treasury Department staff (middle-level bureaucrats), attendants (street level bureaucrats), and citizens who were in the Treasury Department in-person service sector.

The original notice contains text written in a bureaucratic language and technical terms, which might discourage people to read it and to understand it. The instructions for payment were not clear. Additionally, a big part of the content was unnecessary to most taxpayers, once it described what to do in case one had sold the property, which accounted for a few cases. Based on that, we hypothesized that people might not understand what they have to do (because the call to action was not explicit), how they have to do it (because the instructions were not clear), and that they might procrastinate and leave it to another moment (because the sense of urgency was not emphasized). We also noticed that many people might not have a full perception of the consequences of being registered to CADIN, once many people do not even know what it stands for.

It is important to note that a great part of the taxpayers in debt might be due to structural causes, such as the lack of money or companies that deliberative opt for not to pay. Economic crisis, unemployment rate, and household indebtedness drive part of the debt behavior, and these would hardly be affected by our efforts. Even though, because we randomly assign subjects, we expected the non-behavioral mechanisms that are affecting the payment of duties to be, on average, the same among the control and the treatment groups.

The notice letter was redesigned based on these context perceptions from our field immersion and on the literature review, detailed in Section 3. After we designed initial prototypes, we tested them with fellow public servants from the Innovation and Technology Department and from the Treasury Department, using an open-question format. We also tested them with citizens, randomly approached in the streets, using a structured interview format. Both these tests aimed only to gather general impressions and to perform minor changes in the first prototype letters.

In the next subsection we explain the content of each letter, and the original Portuguese versions are included in the Appendix.

4.2.1 Simplified Letter

The original letter was redesigned to make it easier to read and understand. The new version has a headline with the objective call to action (*“Settle your debt with the City Hall”*) and simplified language. We excluded unnecessary parts of the text. Pieces of information are highlighted in bold to facilitate dynamic reading. The text explains the action needed (pay the due installments), the deadline (30 days), and how to do it. It also briefly enumerates the consequences of not taking action.

We moved the part that explained the required actions in exceptional cases, such as property sale to a separate box at the bottom of the letter, out of the attention focus. The general intention of this letter is to make the action easier and simpler.

4.2.2 Simplified + Social Norms Letter

This version explored the social norms of compliance, in addition to the changes made in the Simplified Letter. The Social Norms Letter's headline contains the following text: *“Be part of the majority that is up to date with São Paulo”*. The text reinforces this message (*“most people from São Paulo paid their property tax on time”*) and nudges the late taxpayer to follow the norm and regularize his/her situation. By showing the prevalence of desired behavior it aims to update one's beliefs about the frequency of this behavior in society. Framing the taxpayer as outside the group that he is part of leads to behavior change once he/she perceives himself/herself as deviant. The final paragraph is the same as the one in the Simplified Letter, with the instructions to payment and the consequences of not taking action.

4.2.3 Simplified + Deliberate Choice Letter

This letter is written in an alarming and deterrent tone. Its main purpose is to fight inaction by framing it as a deliberate choice. As our literature review showed, this strategy has increased compliance in similar contexts. In our experiment, the headline contains the following message: *“Warning: the deadline for not entering CADIN is in 30 days”*, which intends to create a sense of urgency. The text body starts with a sense of empathy, signaling comprehension to the flaw of non-payment. Nevertheless, it frames the subsequent failure to take action as a deliberate choice, enlisting the following consequences of such, as shown:

“So far, we consider the fact that you are not up to date with your property tax. However, if you do not pay within 30 days, we will understand that this was your choice, and you will enter CADIN and then the Municipality's Active Debt.”

Again, the last paragraph contained the instructions to payment and the consequences of not taking action, identical to the Simplified Letter.

4.2.4 Simplified + Consequences Letter

The main purpose of this letter is to inform taxpayers about the costs and consequences of non-compliance, which can usually be underestimated, thus increasing the perceived costs of non-compliance. According to the literature review, this has been effective in multiple contexts.

This version is written in an explanatory tone. As a headline, it contains *“You can still pay your property tax before your name is registered to CADIN”*. It details all the possible consequences of non-payment, exploring what can happen. This can be perceived as a deterrent component. It concludes with the following statement: *“Nobody wants this to happen: neither do you nor the City Hall”*, which frames inaction as a poor choice for both sides. It aims to establish an empathy bond between the tax authority and the taxpayer. The final paragraph contains the payment instructions identical to the Simplified Letter.

4.2.5 Simplified + Visual Illustration Letter

This letter redesigned the Simplified Letter's text content in a visual illustration. It brings the step by step required to payment, sequentially ordered, with icons: access the website; issue the billet; pay it. The intention was to make the content easier to read and more attractive. It also aimed to ease the perception of the action required, once it breaks it in smaller steps, reducing the cognitive load.

4.3 Sample selection and random treatment assignment

The trial focuses on late taxpayers and redesigned the CADIN letter. The sample was taken from a weekly mail at the end of June 2019 and all letters were delivered on the 3rd of July. Individuals were randomly assigned to one of six arms: one control and five treatment arms. Sample sizes are reported in Table 1. Randomization was made at property owner level¹. The tax authority was responsible for sending the letters to the properties' addresses.

The sample accounted for 15,178 debts and 12,310 individuals, because frequently a property owner has more than one property in debt. Besides that, a property owner can have more than one property and debt in only one of it. In this case, the owner will receive the CADIN notification for the property in debt in all properties he/she/it owns, as an effort to maximize the probability of one receiving the official letter. That is the reason why we sent more than 25 thousand letters². We programmed our randomization mechanism³ to make sure the same letter is sent to the same person in all the addresses registered to avoid contamination among groups. Randomization was made at property owner level.

Table 1: Letters, debts and property owner in each group

Group	Number of letters delivered	Number of debts	Number of property owners
Control	4,178	2,493	1,983
Simplified	4,225	2,536	2,016
Social Norms	4,524	2,772	2,208
Deliberate Choice	3,968	2,421	2,073
Consequences	3,922	2,376	1,968
Visual Illustration	4,240	2,580	2,062
Total	25,057	15,178	12,310

4.4 Dependent variables

The main outcome variable, *regularization*, takes the value 1 only if the taxpayer regularized his/her/its property liability within 30 days after receiving the letter, this is, before or at the due date. Regularization can happens when the property owner pays the debt, or when he/she/it corrects any register information that was wrong or missing. The latest accounts for less than 2% of all regularization.

4.5 Hypotheses

We can write the null hypothesis and the alternative hypothesis of this experiment as:

Null hypothesis: The mean regularization rate of the treatment and the control groups are equal;

Alternative hypothesis: The mean regularization rate of the treatment group is larger than the mean regularization of the control group.

¹An individual can own more than one property in São Paulo city. An individual is a property owner, that can be either a person or a company. If the individual is in debt in more than one property, then it/he/she will receive one letter for each property.

²As we were not aware of duplicated registers during the experiment design, initial power calculation were made considering the total number of letter, not the total number of debts, which was the real sample size for the test.

³Due to data privacy, we could not access the database prior the experiment, so we faced limitations in the experiment design and in the randomization process. We developed an algorithm that automatically randomized our sample and did the necessary checks to guarantee the database's integrity. Thus, it was not possible to analyse a similar database before conducting the experiment, to understand what individuals' characteristics could have been used to stratified our sample.

Following the analysis conducted in similar experiments, we expected a positive treatment effect, so we are performing a one-tailed test. The experiment has five types of treatment. These hypothesis are valid for each one of the treatment arms. The test was not intended to compare treatments with each other; only to compare each treatment with the control group. Thus, we are performing five different comparisons, and must account for that when analysing the significance of our results.

4.6 Descriptive Statistics

We did not have access to the database prior to the test to stratify the sample, so it was not possible to ensure balance by a block randomization strategy. Therefore, the first step of our analysis was to run balance checks. We conducted tests for the following variables: the characteristic of the property owner (person or company); the year of the debt origin (if it is from the year of 2019 or older); how many installments are due (if the property owner were in debt of only one installment or more than one); and the mean debt value.

Being a person or a company might affect how one responds to treatment. A binary variable was created to indicate whether the observation is an person (receives a value of 1 if so, receives the value of 0 if a company). Thus, the average of this variable will be the proportion of people in each group. We found that there is indeed a statistical distinct proportion of person and companies in three of the groups, when compared to the sample control group. The group that received the Social Norms Letter has a higher proportion of companies than the Control group; the groups that received the Deliberate Choice Letter and the Consequences Letter have a higher proportion of person than the Control group.

Our second balance check refers to the debt year of origin: older debts may be less likely to be paid than newer debts. The database has debts from the year 2012 to 2019, although the majority of debts are from 2019. We than compared the proportion of 2019's debts in each group and saw that this proportion is balanced among all treatment arms and control.

The number of installments in debt might be also an important variable in explaining payment, once it indicates a recurring behavior. Most observations have only one overdue installment, but about 40% have between 2 and 5 overdue installments. With a 95% significance level, only the Visual Illustration Letter group has a statistically different proportion of debts with only one installment, compared to the control group sample.

Following the same reasoning, the total debt value might also determinant to explain payment behavior. Number of installments and total value owned are two sides of the same variable: how relevant is the value of the debt to those who are in debt? To only analyse the gross amount can induce false conclusions: a two thousand Reais debt can be very relevant for one's financial situation and not so relevant for another. Thus, once the IPTU value is linked to the property value, and the latter is related to the person's purchasing power, looking at the number of installments may be more relevant than the total value of the debt. At the same time, for generalizations of the experiment it is necessary to know the average value due.

4.7 Empirical Strategy

To estimate the intention to treat (ITT) effects of the treatment letters on tax compliance, we employed an ordinary least squares (OLS) model. Formally, we estimated the following specification:

$$Y_i = \alpha + \beta_1 T_{1i} + \beta_2 T_{2i} + \beta_3 T_{3i} + \beta_4 T_{4i} + \beta_5 T_{5i} + u_i \quad (3)$$

where Y is the outcome variable, α is a constant; β_t is the parameter to be estimated, with $t = 1,2,3,4,5$, representing each treatment arm; T are binary variables representing the treatment groups ($T_1 =$ Simplified, $T_2 =$ Social Norms, $T_3 =$ Deliberate Choice, $T_4 =$ Consequences, $T_5 =$ Visual Illustration); and u_i is the error term. We performed our analysis using OLS with robust standard errors clustered at property owner level

Table 2: Balance check

	Difference relative to the control group mean					
	Control mean	Simplified	Social Norms	Deliberate Choice	Consequences	Visual Illustration
Person	0.8504 (0.010)	0.0050 (0.010)	-0.0300 (0.010)	0.0730 (0.009)	0.0330 (0.010)	0.0160 (0.010)
Debt from 2019	0.9795 (0.004)	0.0000 (0.004)	-0.0030 (0.004)	0.0040 (0.004)	-0.0010 (0.004)	0.0050 (0.004)
One installment	0.6117 (0.0140)	-0.0240 (0.0140)	-0.0120 (0.0130)	-0.0170 (0.0140)	-0.0040 (0.0140)	-0.0290 (0.0140)
Mean debt value (R\$)	3,733.10 (702.68)	-898.99 (600.29)	-1,047.13 (576.05)	-622.55 (619.08)	-682.00 (636.79)	-771.41 (603.40)
N	2,493	2,536	2,772	2,421	2,376	2,580

Notes: Each row shows a regression of the variable in question on treatment dummies and a constant term. The constant captures the value for the control group (no letter redesign). Columns (2)-(6) show the difference between the treatment groups and the control group. Robust standard errors are in parentheses. See a discussion about debt value in the Appendix.

We do not have information on whether the letters were actually opened. Thus, we are not able to precisely calculate the average treatment effect (ATE), only the ITT. The letters were redesigned only in the inside, so before opening it one cannot perceive any differences among them. Thus, even if we can not know the opening rate, we do not expect it to be different among groups. The regression uses the data collected on September 30th, 2019, three months after the letters were sent.

5 Results

Table 3 report ITT effects of the OLS estimation described in equation 3. The result that clearly emerges from the estimation is that all redesigned letters performed better than the control letter. Nevertheless, the size of the effect differs according to the letter design. In the control group, 48.46% of debt were regularized, while in the best performing treatment this percentage reaches 52.53%.

We must account for the fact that our treatment has multiple arms, so we make multiple comparisons. We performed Bonferroni, Holm and Benjamin - Hochberg corrections. Table 4 report the p-values for all corrections.

Considering our α level of 0.05, we can only reject the null hypothesis in all cases for the Consequences Letter, and we can reject the null hypothesis for Deliberate Choice Letter in two cases.

We calculated the power level for each arm, considering the estimated ITT. We also calculated what would be the minimum detectable effect (MDE) for the sample size, and the sample size needed to detect the estimated ITT. Results are reported at Table 5

We can observe that the only well powered treatment arm is the Consequences Letter, the treatment that showed a positive and non zero result. We were not aware of duplicated registers during the experiment design, so the power calculation considered the total of letter sent, not the total of debts (as showed in Table 1. The reduction of observations in our experiment reduced its power.

Table 3: Regression results

Variable	Coefficient	T-statistic
Intercept/ Control group mean	0.485*** [0.010]	48.398
Simplified	0.026* [0.014]	1.823
Social Norm	0.011 [0.014]	0.779
Deliberate Choice	0.031** [0.014]	2,140
Consequences	0.041*** [0.014]	2,839
Visual Illustration	0.024* [0.014]	1,680
Observations	15,178	

Notes: robust standard error clustered at individual property owner level.

*** p<0.01, ** p<0.05, * p<0.1.

Table 4: P-values

Letter	One-tailed p-value	Bonferroni	Holm	Benjamin
Simplified	0,034	0,171	0,102	0,057
Social Norm	0,218	1,000	0,218	0,218
Deliberate Choice	0,016	0,081	0,064	0,040
Consequences	0,002	0,011	0,010	0,010
Visual Illustration	0,047	0,233	0,102	0,059

Table 5: Power calculation

Letter	Effect (ITT)	Power (alpha = 0.05)	MDE	Sample needed to detected ITT
Simplified	0,0257	45%	0,0350	9.307
Social Norm	0,0108	12%	0,0342	53.286
Deliberate Choice	0,0305	57%	0,0354	6.598
Consequences	0,0407	81%	0,0355	3.712
Visual Illustration	0,0236	39%	0,0348	11.052

5.1 Results heterogeneity

Although the major purpose of our experiment was to observe what letter shows the best result on average, it is relevant to analyse whether different letters have heterogeneous impacts according to specific subject characteristics. It may be useful to inform future tax policies.

Overall, the regularization rate is different among individuals and legal entities, being lower for the second group. Even so, the observed impacts of the letters are greater for this group.

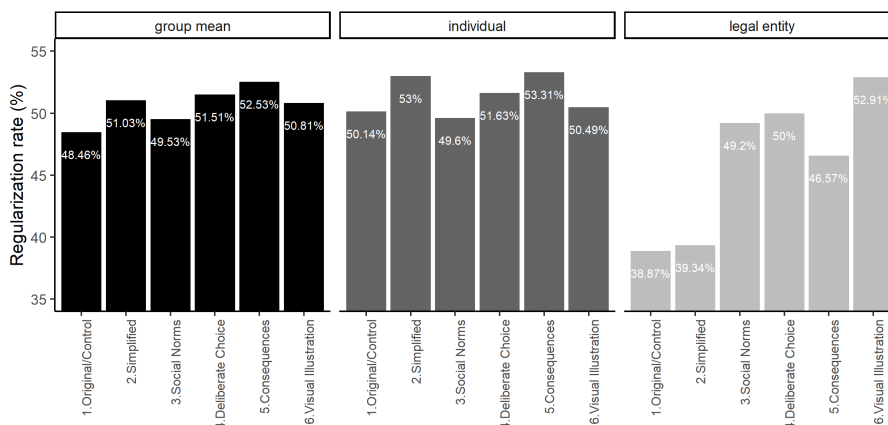


Figure 1: Results heterogeneity: individuals versus legal entities

The Social Norms Letter can be considered the most “neutral“, as it had very similar results for both individuals and companies. The control letter and the Simplified Letter (whose average is statistically equal to that of the control group) have the highest impact differential between legal entities and individuals.

We can not rule out the possibility that the overall positive result observed with the Consequences Letter is due to the difference in the proportion of individuals in the group: this letter works better for individuals and there is a higher proportion of individuals in this group, when comparing to the control group. It is possible to observe that this letter is not so effective for legal entities.

As for the number of installments, the letter effect seems to be larger for subjects that have only one installment due (small debt), than for those who have more than one installment due.

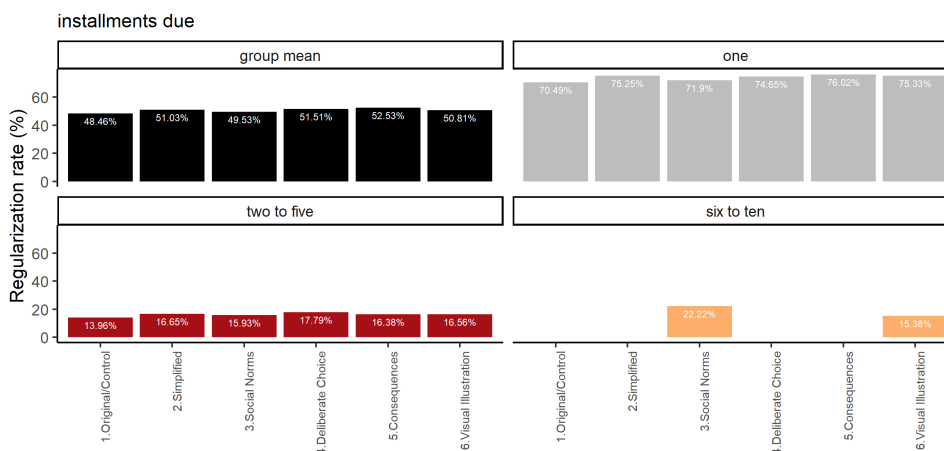


Figure 2: Regularization by number of installments due

5.2 Cost-Benefit Analysis

The first cost-benefit analysis aims to understand what happened to revenue collection only within the experiment universe. We compared what would have been collected with the regularization rate of the control group (if there was no intervention) and the estimated collection with the regularization rates of each group. Considering the debts that participated in the experiment, the intervention itself has already generated an increase in revenue of approximately R\$950 thousand. Since the operational cost of the experiment was close to R\$50 thousand, the trial itself covers the costs associated with it. Therefore, it indicates that this kind of endeavour is sustainable as a public policy and experiments should be done more broadly in the municipal administration.

Table 6: Revenue increase in the experiment

Mean debt (R\$)	Regularization rate	Expected debt value regularized (R\$)
3,733	regularization rate of each group	22,939,487
3,733	regularization rate of control group (48.46%)	21,986,328
Payment increase (R\$)		953,159

The second cost-benefit analysis aims at calculating how much municipal revenues could increase if the Consequences Letter were to be implemented to all late taxpayers. In 2019, 255,664 property owners had more than 460 thousand IPTU debts, which summed up to R\$2.5 billion. Thus, we calculate the mean debt value of R\$ 5,501, superior to the mean debt in the experiment (R\$ 3,733). In the control group, 48.5% of taxpayers regularize their situation, so 51.5% do not. If the percentage of taxpayers that pay their debts increases by 4.1 percentage points, revenues could increase around R\$ 60 millions (considering the mean debt of 3 thousand Reais).

6 Conclusion

This study presents the results of a randomized control trial conducted at the municipal level with 15,178 IPTU debts. It aimed to investigate whether behavioral principles added to the government communications to taxpayers are efficient in improving compliance, and, if so, what principles are more effective. The trial was run in the largest Brazilian city, São Paulo, and the property tax is the second main source of revenue in municipal budgets.

Two versions of the redesigned letter had positive and significant results in the regularization rate: the Deliberate Choice Letter and the Consequences Letter. The Simplified, the Social Norms, and the Visual Illustration versions showed no statistically significant impact in comparison to the original letter. Similar to previous studies, 'hard tone' messages were more effective than 'soft-tone' messages (Bérgolo et al. 2017, Castro and Scartascini 2015, Hernandez et al. 2019, 2017, Ortega and Sanguinetti 2013, Pomeranz 2015, Wenzel and Taylor 2004). Yet, the best version is not the most deterrent one, since the Consequences Letter has a softer, more explanatory tone, than the Deliberate Choice Letter's tone.

The Social Norms Letter was the one which increased payment rate the little, although it has been proven to increase compliance in different contexts. Whether it is due to cultural and societal traces, to the materialization of the principle, or to limitations in our experiment design, is something that still requires further investigation.

Our experiment was underpowered and had some balance issues among groups, so we can not be sure if

the non-significant versions would produce significant impact in further trials. This situation illustrates the importance of good research design in conducting experiments, and how difficult it is to implement perfect trials in the public administration, given its bureaucratic, operational, and legal constraints.

In terms of tax policy, we recommend the implementation of the Consequences Letter as the official CADIN communication to late taxpayers of IPTU. If the percentage of taxpayers that pay their debts increases by 4.1 percentage points, revenues could increase around R\$ 60 millions (considering the mean debt of 3 thousand reais) a year for the City of São Paulo.

Although we might not have detected the other letter's effects, the test indeed concludes that this version is the best among the versions tested, and best than the original letter version. It also offers strong suggestions to adopt similar approaches in other CADIN communications. This trial also indicates what could be an efficient approach to fiscal communications relative to other taxes, and other channels, such as emails and text messages. The increase of revenue calculated in the experiment alone was R\$950 thousand reais, which is 19 times higher than the costs needed to cover the operational costs required to implement the experiment (around R\$50 thousand reais). Thus, the execution of randomized control trials in the fiscal context seems to be financially sustainable in terms of budget, appearing as a good investment for fiscal authorities if there is available human resources and time to enter in such endeavour.

In terms of scientific contribution, this study is the first successful application of behavioral interventions rigorously tested in São Paulo municipality, and, to our knowledge, the first documented behavioral experiment with tax compliance in Brazil. Further research is needed in order to conclude which behavioral principles and tone are indeed the most effective in increasing tax compliance in Brazil.

Our study also indicates that there is indeed space for behavioral science applications in fiscal public policies in this context. The results suggest that adopting a behavioral design in taxpayer communications could produce considerable gains to municipal revenue. The efficacy of the communication pieces elaborated in this study could serve as insight for other governmental communications.

Bibliography

- Allingham, Michael G and Agnar Sandmo. 1972. "Income tax evasion: A theoretical analysis." *Journal of public economics* 1(3-4):323–338.
- Ariel, Barak. 2012. "Deterrence and moral persuasion effects on corporate tax compliance: findings from a randomized controlled trial." *Criminology* 50(1):27–69.
- Bérgolo, Marcelo L, Rodrigo Ceni, Guillermo Cruces, Matias Giacobasso and Ricardo Perez-Truglia. 2017. "Tax audits as scarecrows: Evidence from a large-scale field experiment."
- Castro, Lucio and Carlos Scartascini. 2015. "Tax compliance and enforcement in the pampas evidence from a field experiment." *Journal of Economic Behavior & Organization* 116:65–82.
- Chetty, Raj and Emmanuel Saez. 2013. "Teaching the tax code: Earnings responses to an experiment with EITC recipients." *American Economic Journal: Applied Economics* 5(1):1–31.
- Coleman, Stephen. 1996. "The Minnesota income tax compliance experiment: State tax results."
- Coleman, Stephen. 2007. "The Minnesota income tax compliance experiment: replication of the social norms experiment." *Available at SSRN 1393292* .
- Dayrell, Marina. 2018. "IPTU 2019: Entenda tudo sobre o imposto dos imóveis." . Available at <https://www.terra.com.br/economia/iptu-2019-entenda-tudo-sobre-o-imposto-dos-imoveis,df05bdb8d8d7a9206dc694198e6c8c40syyjbu2z.html>.

-
- Del Carpio, Lucia. 2013. "Are the neighbors cheating? Evidence from a social norm experiment on property taxes in Peru." *Princeton, NJ, Princeton University Working Paper* .
- Dwenger, Nadja, Henrik Kleven, Imran Rasul and Johannes Rincke. 2016. "Extrinsic and intrinsic motivations for tax compliance: Evidence from a field experiment in Germany." *American Economic Journal: Economic Policy* 8(3):203–32.
- Hallsworth, Michael, John A List, Robert D Metcalfe and Ivo Vlaev. 2017. "The behavioralist as tax collector: Using natural field experiments to enhance tax compliance." *Journal of public economics* 148:14–31.
- Hashimzade, Nigar, Gareth D Myles and Binh Tran-Nam. 2013. "Applications of behavioural economics to tax evasion." *Journal of Economic Surveys* 27(5):941–977. Available at <https://doi.org/10.1111/j.1467-6419.2012.00733.x>.
- Hernandez, Marco, Jonathan Karver, Mario Negre and Julie Perng. 2019. "Promoting Tax Compliance in Kosovo with Behavioral Insights."
- Hernandez, Marco, Julian Jamison, Ewa Korczyk, Nina Mazar and Roberto Sormani. 2017. "Applying behavioral insights to improve tax collection: experimental evidence from Poland."
- Kettle, Stewart, Marco Hernandez, Simon Ruda and Michael Sanders. 2016. "Behavioral interventions in tax compliance: Evidence from Guatemala." Available at <https://elibrary.worldbank.org/doi/abs/10.1596/1813-9450-7690>, eprint = <https://elibrary.worldbank.org/doi/pdf/10.1596/1813-9450-7690>.
- Kleven, Henrik Jacobsen, Martin B Knudsen, Claus Thustrup Kreiner, Søren Pedersen and Emmanuel Saez. 2011. "Unwilling or unable to cheat? Evidence from a tax audit experiment in Denmark." *Econometrica* 79(3):651–692.
- Mascagni, Giulia. 2015. "Tax Experiments in Developing Countries: A Critical Review and Reflections on Feasibility."
- Mascagni, Giulia, Christopher Nell and Nara Monkam. 2017. "One size does not fit all: a field experiment on the drivers of tax compliance and delivery methods in Rwanda."
- Ortega, Daniel and Pablo Sanguinetti. 2013. "Deterrence and reciprocity effects on tax compliance: experimental evidence from Venezuela."
- Pomeranz, Dina. 2015. "No taxation without information: Deterrence and self-enforcement in the value added tax." *American Economic Review* 105(8):2539–69.
- Thaler, Richard H and Cass R Sunstein. 2008. "Nudge: improving decisions about health." *Wealth, and Happiness* 6.
- Weber, Till Olaf, Jonas Fooker and Benedikt Herrmann. 2014. "Behavioural economics and taxation." *European Commission Taxation Papers* .
- Wenzel, Michael and Natalie Taylor. 2004. "An experimental evaluation of tax-reporting schedules: a case of evidence-based tax administration." *Journal of Public Economics* 88(12):2785–2799.
- World Bank. 2018. Mind, Behavior, and Development Unit : Applying Behavioral Science to End Poverty and Enhance Equity (English). Technical report World Bank. Available at <http://documents.worldbank.org/curated/en/744191532458732002/Mind-Behavior-and-Development-Unit-Applying-Behavioral-Science-to-End-Poverty-and-Enhance-Equity>.
- Yitzhaki, Shlomo. 1974. "A Note on Income Tax Evasion: A Theoretical Approach." *Journal of Public Economics* 3(2):2011–202.