

*Journal of Development Economics*

**Pre-Results Review (Registered Reports)**

**Optional Stage 1 Submission Template**

**Contents**

Contents .....	1
Cover page.....	2
Abstract.....	2
Proposed timeline.....	2
1. Introduction .....	3
2. Research Design .....	7
Conceptual framework.....	7
Hypotheses.....	8
Treatments.....	11
Outcome variables.....	13
Variables measuring people's prior beliefs and preferences .....	13
3. Data.....	14
Selection of countries .....	14
Approach to data collection .....	15
Details about data collection.....	16
Sample and statistical power.....	17
Pilot data.....	18
4. Analysis.....	20
Main statistical analysis .....	23
Robustness Checks.....	24
5. Interpreting Results.....	26
6. Bibliography .....	28
7. Administrative information .....	30

## Cover page

# Journal of Development Economics

## Registered Report Stage 1: Proposal

### How does the progressivity of taxes and government transfers impact people's willingness to pay tax? Experimental evidence across developing countries

**Author and institutional affiliation:** Christopher Hoy (World Bank)

**Date of latest draft:** 20<sup>th</sup> of January 2022

#### Abstract

This registered report describes a study that will examine how the progressivity of taxes and government transfers impacts people's willingness to pay tax through a large-scale randomized survey experiment across eight developing countries (Colombia, Ghana, Indonesia, Jordan, Mexico, Sri Lanka, South Africa and Tanzania). Respondents will be randomly allocated to either receive information about the progressivity of taxes and/or government transfers in their country or to receive no information. The treatments are based on a recently released, cross-country database that measures the Sustainable Development Goal target to improve the redistributive impact of fiscal policy. The impact of these treatments on people's willingness to pay tax will be measured using standardized questions from the existing literature. By examining heterogeneous treatment effects based on people's prior beliefs and preferences as well as differences across countries it will be possible to identify how equity and reciprocity considerations influence people's willingness to pay tax.

**Keywords:** Redistribution, Tax compliance, Randomized Experiment, Political Economy

**JEL codes:** D31, D91, H22, H23, H24, H26

**Study pre-registration:** This study has been pre-registered on the American Economic Association RCT Registry (ID number AEARCTR-0008847).

#### Proposed timeline

This study is due to be launched on the 24<sup>th</sup> of January 2022 and all fieldwork should be completed by early April 2022. In December 2021, piloting of the survey instrument took place.

# 1. Introduction

There is little to no government-led redistribution from rich to poor households in low- and middle-income countries, which has been identified as an issue for policy makers to address in the Sustainable Development Goal (SDG) Target 10.4. Figure 1 shows that the difference between the gross (i.e., pre-taxes and government transfers) and net (i.e., post-taxes and government transfers) GINI index (the indicator for tracking SDG target 10.4) is negligible in almost all low- and middle-income countries for which comparable data is available (CEQ, 2021). This is because taxes and government transfers, particularly indirect taxes (e.g., value added tax) and indirect transfers (e.g., subsidies), are typically not progressive (i.e., the rich do not pay a relatively higher share of their income in tax and the poor do not receive a relatively higher share of their income in government transfers). Even though most households in low- and middle-income countries either directly or indirectly pay tax and/or receive government transfers, this typically is only equivalent to a very small share of their household income (CEQ, 2021). Immediately prior to the COVID-19 pandemic, only 10 percent of Gross Domestic Product (GDP) was collected in tax revenue and 1.5 percent of GDP was provided in government transfers, on average, across low- and middle-income countries (World Bank, 2021). These patterns raise the question as to how a lack of equity and reciprocity in tax and transfer systems impacts people's willingness to pay tax in low- and middle-income countries.

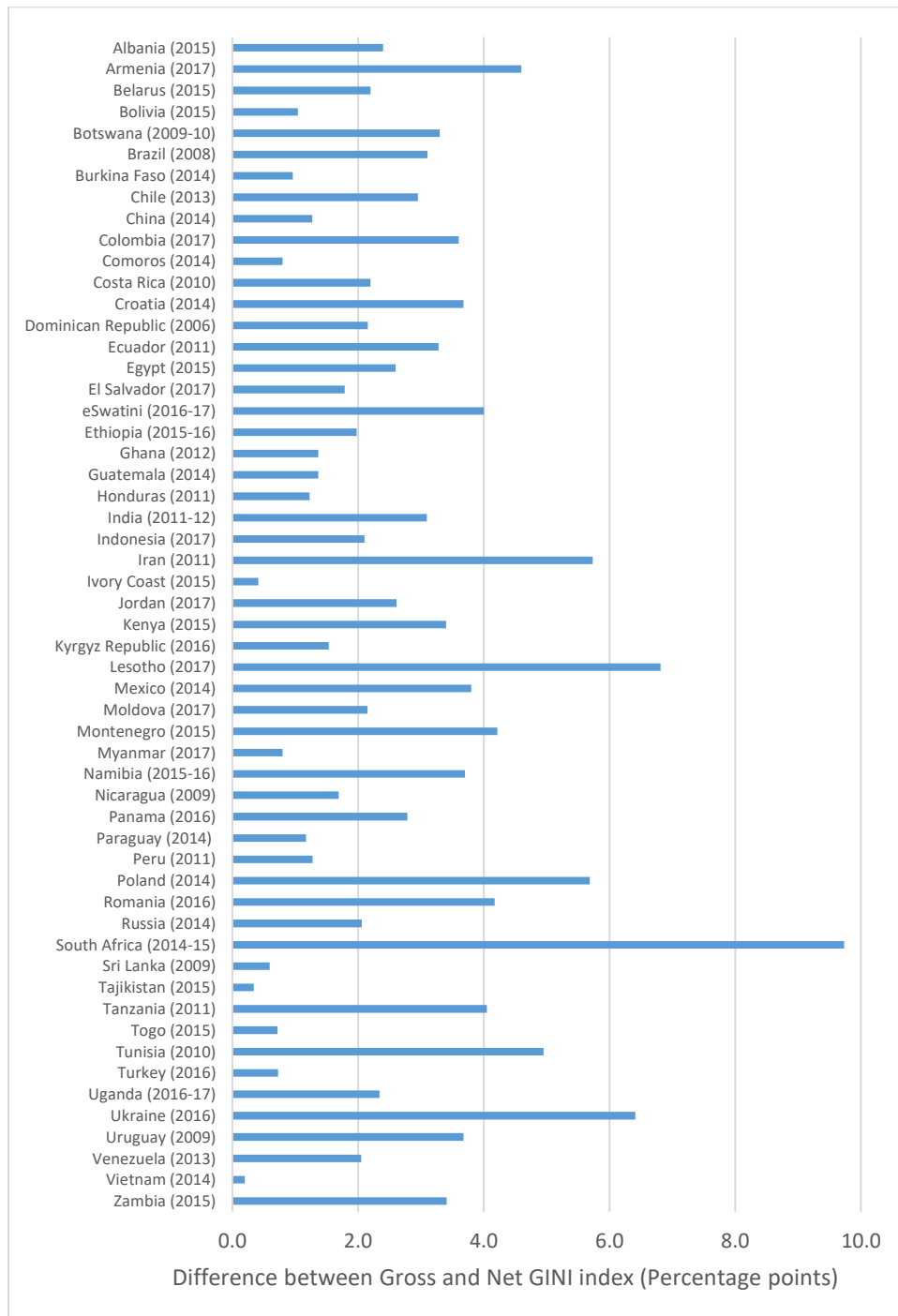
We answer this question by conducting a large-scale, online randomized survey experiment that is broadly representative of the population with internet access in eight low- and middle-income countries (Colombia, Ghana, Indonesia, Jordan, Mexico, Sri Lanka, South Africa and Tanzania). This is a diverse set of countries that make up around ten percent of the developing world's population, are spread across Latin America, Africa, Asia and the Middle East, and had GNI per capita ranging from \$US1080 to \$US8480 (Atlas Method) in 2020 (World Bank, 2021). We exploit a newly released database that uses the Commitment to Equity Institute methodology<sup>1</sup> to measure how taxes and government transfers are distributed in these countries based on recent Household Income and Expenditure Surveys (HIES) (CEQ, 2021). Specifically, we test how accurate information about the progressivity of taxes, government transfers or both impacts people's willingness to pay taxes, which is measured using standardized questions from cross-country surveys (e.g., from the Afrobarometer). We also capture people's prior beliefs and preferences about the distribution of taxes and government transfers, levels of inequality, their perceived place in the national income distribution, and beliefs about whether their household is a net contributor to or beneficiary from the tax and transfer system.

Equity and reciprocity in the tax and transfer system are expected to impact people's willingness to pay tax as they are "quasi-voluntary" reasons for compliance. These motivations for paying tax are particularly important in low- and middle-income countries as there is typically weak capacity to enforce tax legislation (Prichard et al., 2019). Equity is thought to be a central driver for why people pay tax because most people prefer to live in societies with lower levels of inequality than what they perceive to exist (WVS, 2020) and governments can play a key role in bringing about greater equality through redistribution using tax and transfer systems. Reciprocity is deemed to be a significant factor as people's willingness to pay taxes is influenced by the extent to which they believe their household benefits from how taxes are spent; most directly this is in the form of receiving government transfers (including subsidies).

---

<sup>1</sup> The CEQ approach was developed by the Commitment to Equity Institute (CEQ Institute) at Tulane University. The methodology, implementation guidelines, applications, and software of the CEQ approach can be found in N. Lustig (Ed.). 2018. "Commitment to Equity Handbook. Estimating The Impact of Fiscal Policy on Inequality and Poverty." Brookings Institution Press.

**Figure 1 – Difference between the Gross and Net GINI index in low- and middle-income countries**



*Note: The year shown in brackets next to each country is the year in which the survey took place that the GINI index is based on.*

*This figure shows that the difference between the gross (i.e., pre-taxes and government transfers) and net (i.e., post-taxes and government transfers) GINI index is negligible in almost all low- and middle-income countries. In comparison the difference between the gross and net GINI index in Spain (2017) and the USA (2011) was 11.9 and 7.0 percentage points respectively (CEQ, 2021).*

This study will examine how people's perceptions about and preferences for the degree of equity and reciprocity in the tax and transfer system influence their willingness to pay tax. Firstly, we will identify stylized facts about what relationships exist between people's beliefs and preferences about how taxes and government transfers are distributed in their country and their willingness to pay tax (this will be discussed in the data section of the paper). By doing so we will quantify the extent to which people's (mis)perceptions about equity and reciprocity in the tax and transfer system may be making them more or less likely to pay tax. Secondly, we will test how accurate information about the distribution of taxes and/or transfers influences people's willingness to pay tax (this will be discussed in the results section of the paper). By examining the heterogeneous treatment effects based on people's prior beliefs as well as differences between the treatments and across countries we will identify the channels through which people's tax compliance is influenced by the distribution of taxes and/or government transfers in their country.

The findings from this study will shed light on how policy reforms that alter equity and reciprocity in the tax and transfer system may influence people's willingness to pay taxes, as well as identifying when competing channels exist between equity and reciprocity. Specifically, we will show the extent to which efforts to improve a country's fiscal position by changing the equity and reciprocity in the tax and transfer system may *backfire* by reducing people's willingness to pay taxes. This can be illustrated using the example of a reform requiring the richest households to pay more tax (e.g., an increase in the top marginal income tax rate), which would make the tax and transfer system more progressive. In this case, people may respond by being more likely to comply if they believe the tax and transfer system was not progressive enough; however, it is possible that the opposite may be the case if they believe the tax and transfer system was already progressive enough. Further, if a tax increase along these lines was intended to fund government spending that primarily benefits the rich (poor) this could increase (decrease) compliance as there would be more (less) reciprocity in the tax and transfer system. Another illustrative example is a reform that improves the progressivity of an existing government transfer (e.g., a cash transfer) by restricting eligibility so that benefits are only provided to the very poorest households. This reform could undermine reciprocity between middle class and richer households and the government, which could consequently decrease people's willingness to pay taxes. The net impact of a reform along these lines on people's willingness to pay tax will also be influenced by whether most people believe the tax and transfer system is already progressive enough. Both of these stylized examples illustrate how understanding the ways equity and reciprocity impact people's willingness to pay taxes has important implications for policy makers looking to reform the tax and transfer system. Inadequate consideration of these issues could result in reforms of the tax and transfer system, which were intended to improve the fiscal position of a country, leading to lower levels of revenue due to worsening tax compliance.

This study makes several contributions to two broad strands of the existing literature. The first strand our study contributes to is in relation to how people's perceptions shape their preferences regarding tax and transfer policies (Gimpelson and Treisman, 2018; Hauser and Norton, 2017). Seminal work on this topic has been conducted in recent years using large-scale, randomized survey experiments in the United States and Western Europe examining a range of topics, such as inequality (Kuziemko et al., 2015), social mobility (Alesina et al., 2018) and immigration (Alesina et al., 2021). A common thread in these studies is that, in general, most people have a poor understanding about the economic circumstances in their country (e.g., about the level of inequality (Norton and Ariely, 2011)) and they have tested what happens to people's general policy preferences when they are provided accurate information. We extend this literature in three ways. Firstly, we test how accurate information about existing *policies* (specifically the progressivity of taxes and government transfers), as opposed to existing *circumstances* (e.g., the level of inequality), shifts

people's preferences. In other words, we directly alter people's beliefs about the role the government currently plays in distributing resources in their country, and see how this shifts their preferences, as opposed to examining how people's views change about what the role of the government should be once they are aware of the actual circumstances in their country. Secondly, we focus our randomized survey experiments on measuring a specific intention (people's willingness to pay), which is a key way people engage with the government, as opposed to general preferences. This allows us to infer clearer policy implications from our work. Thirdly, we conduct one of the first and by far the largest randomized survey experiments in this literature in low- and middle-income countries (the previously largest study was in five middle-income countries by Hoy and Mager, 2021). We incorporate best practices into the design of the randomized survey experiments from cross-country studies in high-income countries and also utilize a novel sampling methodology that allows us to collect a more representative sample of the internet population than is typically captured in online surveys in low- and middle-income countries.

The second strand of literature is in relation to a growing body of research about "quasi-voluntary" motivations for tax compliance. Many studies suggest that a more multifaceted model of the drivers of tax compliance is required, beyond the seminal work of Allingham and Sandmo (1972) that shows taxpayers make an assessment about whether to comply based on the likelihood of enforcement and punishment for non-compliance. For example, there has been a proliferation of studies in high-income countries trialing how non-deterrence messages, such as referring to social norms, can increase tax compliance (Hallsworth, 2014; De Neve et al., 2019). The extent to which "quasi-voluntary" motivations for tax compliance exist in low- and middle-income countries that are typically characterized by weaker enforcement is still unclear. Outside of Latin America, there has been only a small number of randomized field experiments testing alternative motivations for people's willingness to pay tax in low- and middle-income countries, such as in Ethiopia (Shimeles et al., 2017), Rwanda (Mascagni et al., 2017) and Papua New Guinea (Hoy, McKenzie and Sinning, 2021). We contribute to this field by going well beyond existing work that has focused on providing subsets of taxpayers with generic messages about "public goods" in three ways. Firstly, our randomized survey experiment has been designed to specifically identify how the distribution of taxes and government transfers impacts people's willingness to pay tax. By capturing people's prior perceptions and preferences we can isolate the channels that impact people's intentions. Secondly, we use standardized questions from cross-country surveys that measure different aspects of people's willingness to pay tax (e.g., from the Afrobarometer and World Values Survey) based on best practice in the literature. Our approach will allow us to identify the robustness of our results by examining the consistency (or lack thereof) of alternative ways of measuring people's willingness to pay tax. Thirdly, we collect data that is representative of the internet population within each country and is comparable across a diverse set of low- and middle-income countries. Consequently, our results provide rigorous insights for a much wider population and arguably have far greater external validity than previous work in this field.

This registered report is structured as follows. Section 2 provides a conceptual framework and hypotheses for this study as well as details about the treatments and outcome variables in the randomized survey experiment. Section 3 describes the data collection process in detail, including how the countries were selected and sample sizes were determined as well as summarizing the lessons learnt from the piloting process. Section 4 outlines how the data will be analyzed and presented in the Stage 2 version of the paper, including the specific statistical models that will be run and the robustness checks that will be conducted. Section 5 discusses how the results will be interpreted, as well as potential policy implications.

## 2. Research Design

### Conceptual framework

Traditionally, people's willingness to pay taxes was conceptualized as a trade-off between the punishment they face from being caught for non-compliance compared to the cost of complying (Allingham and Sandmo, 1972)<sup>2</sup>. This is shown formally in the utility functions below whereby  $y_i$  is an individual's household income before tax,  $d$  is the probability of being detected as non-compliant,  $p_i$  is a fixed amount that represents the punishment a taxpayer will face if found to be non-compliant and  $t_i$  is a fixed amount that represents a taxpayer's tax obligation. However, in recent years this model of tax compliance has been extended to include other factors that drive compliance, such as people's desire to keep in line with social norms (Hallsworth 2014; Slemrod, 2019; Antinyan and Asatryan, 2019). As such, the traditional model of tax compliance has been broadened to incorporate what is often referred to as "quasi-voluntary" motivations for tax compliance (Prichard et al., 2019). The utility gain an individual receives from these additional reasons for paying tax is represented formally as  $a_i$  in the model below. As such, for a single point in time an individual's utility from complying with taxes ( $U_{ci}$ ) and from not complying ( $U_{ni}$ ) can be expressed as:

$$U_{ci} = y_i - t_i + a_i \quad (1)$$

and

$$U_{ni} = y_i - dp_i \quad (2)$$

Accordingly, a taxpayer would comply if  $U_{ci} > U_{ni}$ , which requires that:

$$t_i < dp_i + a_i \quad (3)$$

We extend this basic model by decomposing quasi-voluntary motivations for paying tax (shown as  $a_i$  in the model) to specifically identify how "reciprocity" and "equity" play a role in driving people's willingness to pay tax (Prichard et al., 2019). By doing so we separate these reasons from other quasi-voluntary motivations (shown as  $b_i$  in the revised model below). Reciprocity is deemed to be important as people pay tax partly because they also benefit from how taxes are spent, most directly in the form of government transfers (including subsidies). This is formally represented in the revised model below as  $r_i$ , which is the fraction of tax paid that an individual believes their household will receive in return from the government, both directly through transfers and indirectly through the provision of goods and services. Equity, more precisely articulated as vertical equity by Prichard et al. (2019), is considered to be a driver of tax compliance because many people would prefer lower levels of inequality in their country and consequently are supportive of the role taxes and transfers can play in redistributing resources from rich to poor. This is formally integrated into our model by drawing on the "workhorse" utility function by Alesina and Giuliano (2011) that shows how differences between actual and preferred levels of inequality ( $Q - Q_i^*$ ) impact people's utility (the weight an individual places on deviations from their ideal level of inequality is captured in the term  $\delta_i$ ). The revised model of people's utility from paying tax can be expressed as follows:

---

<sup>2</sup> In some respects, this work involved applying Becker's (1965) seminal work on crime and punishment to tax compliance.

$$U_{ci} = y_i - (1 - r_i)t_i + b_i - \delta_i(Q - Q_i^*)^2 \quad (4)$$

We disaggregate this utility function further by continuing to draw on Alesina and Giuliano's (2011) seminal work as they argue that people's utility is largely (if not exclusively) influenced by differences between actual and preferred levels of inequality that are due to factors outside an individual's control ( $Q^l - Q_i^{l*}$ ), as opposed to overall levels of inequality ( $Q - Q_i^*$ ). We identify that one of the key determinants of inequality outside an individual's control in a country is the degree of progressivity in the tax and transfer system. We reflect this in the model with the term ( $Q^{tt} - Q_i^{tt*}$ ), whereby  $Q^{tt}$  is the level of progressivity in the tax and transfer system,  $Q_i^{tt*}$  is people's preferred levels of progressivity in the tax and transfer system and  $\delta_i^{tt}$  reflects the weighting people place on this (all other differences in inequality are captured in the terms denoted with  $o$ ). As such, holding everything else constant, people who prefer the existing level of progressivity in the tax and transfer system will be more willing to comply with taxes than those who do not. Consequently, the revised model of people's utility from paying tax can be expressed as follows:

$$U_{ci} = y_i - (1 - r_i)t_i + b_i - \delta_i^{tt}(Q^{tt} - Q_i^{tt*})^2 - \delta_i^o(Q^o - Q_i^{o*})^2 \quad (5)$$

The final substantive modification we make is to incorporate the fact that it is people's perceptions of how taxes and transfers are distributed, as opposed to what is actually the case, that will influence their willingness to pay tax. Previous research has shown that people tend to have a poor understanding of both the level of inequality in their country as well as their position in the national income distribution (see, for example, Hoy and Mager, 2021) and there is evidence from the United States to suggest these misperceptions also extend to tax and transfer policies (Stantcheva, 2021). Consequently, we rewrite the utility function to factor in that people's willingness to pay tax will be influenced by their perception of the extent to which they benefit from how taxes are spent ( $r_{pi}$ ) and the extent to which they believe the tax and transfer system is progressive ( $Q_{pi}^{tt}$ ).

$$U_{ci} = y_i - (1 - r_{pi})t_i + b_i - \delta_i^{tt}(Q_{pi}^{tt} - Q_i^{tt*})^2 - \delta_i^o(Q_{pi}^o - Q_i^{o*})^2 \quad (6)$$

This utility function provides guidance as to how people's willingness to pay tax ( $U_{ci}$ ) will be influenced by accurate information ( $I$ ) about the distribution of taxes and/or transfers in their country ( $Q^{tt}$ ). In other words, it is possible to make predictions about how people's utility from paying taxes varies when they have accurate information (i.e.,  $U_{ci}|I$ ) based on their priors (i.e., the difference between  $Q_{pi}^{tt} - Q^{tt}$  as well as the difference between  $r_{pi} - r_i$ ). The two main dimensions in which heterogeneity is expected are in terms of people's prior beliefs and preferences about the progressivity of tax and transfer policies (captured formally as  $(Q^{tt} - Q_i^{tt*})$  vs  $(Q_{pi}^{tt} - Q_i^{tt*})$ ) and people's prior beliefs about whether they are relatively rich or poor, which in turn determines how much they benefit from progressivity (captured formally as  $r_{pi}$  vs  $r_i$ ).

## Hypotheses

Operationalizing the randomized survey experiment means that inevitably the hypotheses discussed in this subsection may not fully capture all the nuance of the theory outlined in the preceding subsection. For example, it would be rather challenging to have a large enough sample size in each country to have adequate statistical power to examine the interaction of a range of different aspects of the conceptual framework, such as how different prior beliefs and preferences about the distribution of taxes and transfers vary based on people's beliefs about whether they are relatively rich or poor. As such, the focus of the



groups of hypotheses outlined below is to examine each of the aspects of the conceptual framework independently from one another.

There are four groups of primary and two groups of secondary hypotheses that emerge from the conceptual framework. Group A of Hypotheses are based on a key implication from the theory that progressivity in the tax and transfer system will on average make people more willing to pay tax. Group B of Hypotheses are built on the theoretical conjecture that poorer people should be more in favor of a progressive tax and transfer system than richer people. Group C of Hypotheses summarizes how people's willingness to pay tax is likely to vary by their prior beliefs about the progressivity of the tax and transfer system. Group D of Hypotheses outlines how people's willingness to pay tax is expected to vary by their preferences for progressivity in the tax and transfer system. In addition, the conceptual framework discussed above indirectly implies that there may be variation between respondents depending on their prior beliefs and preferences about the level of inequality in general and their prior beliefs about whether they are a net contributor to or beneficiary from the tax and transfer system. These aspects form the basis of the secondary hypotheses (Groups E and F). Furthermore, heterogeneous treatment effects will be explored based on whether respondents believed their household paid a large share of income in taxes and/or whether they were working (and consequently would be required to pay tax directly). However, ex-ante the direction of the impact of the treatments on these dimensions is hard to estimate.

### **Primary Hypotheses**

#### Group A – People's willingness to pay tax varies by the degree of progressivity in the tax and transfer system

Hypothesis A1: Informing people that the distribution of taxes is progressive (not progressive), will increase (decrease) their willingness to pay tax

Hypothesis A2: Informing people that the distribution of transfers is progressive (not progressive), will increase (decrease) their willingness to pay tax

Hypothesis A3: Informing people that the net impact of taxes and transfers is progressive (not progressive), will increase (decrease) their willingness to pay tax

#### Group B – People's willingness to pay tax varies based on their perceived place in the income distribution

Hypothesis B1: Informing people that the distribution of taxes is progressive (not progressive), will increase (decrease) their willingness to pay tax if they perceive they are relatively poor (rich)

Hypothesis B2: Informing people that the distribution of transfers is progressive (not progressive) will increase (decrease) their willingness to pay tax if they perceive they are relatively poor (rich)

Hypothesis B3: Informing people that the net impact of taxes and transfers is progressive (not progressive), will increase (decrease) their willingness to pay tax if they perceive they are relatively poor (rich)

#### Group C – People's willingness to pay tax varies by their prior beliefs about the progressivity of tax and transfer system

Hypothesis C1: Informing people that the distribution of taxes is progressive (not progressive) when they thought it was not progressive (progressive), will increase (decrease) their willingness to pay tax

Hypothesis C2: Informing people that the distribution of transfers is progressive (not progressive) when they thought it was not progressive (progressive), will increase (decrease) their willingness to pay tax

Hypothesis C3: Informing people that the net impact of taxes and transfers is progressive (not progressive) when they thought taxes were not progressive (progressive), will increase (decrease) their willingness to pay tax

Hypothesis C4: Informing people that the net impact of taxes and transfers is progressive (not progressive) when they thought transfers were not progressive (progressive), will increase (decrease) their willingness to pay tax

Group D – People’s willingness to pay tax varies by their preferences for the progressivity of tax and transfer system

Hypothesis D1: Informing people that the distribution of taxes is progressive (not progressive) when they prefer it to be progressive (not progressive), will increase (decrease) their willingness to pay tax

Hypothesis D2: Informing people that the distribution of transfers is progressive (not progressive) when they prefer it to be progressive (not progressive), will increase (decrease) their willingness to pay tax

Hypothesis D3: Informing people that the net impact of taxes and transfers is progressive (not progressive) when they prefer taxes to be progressive (not progressive), will increase (decrease) their willingness to pay tax

Hypothesis D4: Informing people that the net impact of taxes and transfers is progressive (not progressive) when they prefer transfers to be progressive (not progressive), will increase (decrease) their willingness to pay tax

**Secondary Hypotheses**

Group E – People’s willingness to pay tax varies by their preferences for the level of inequality

Hypothesis E1: Informing people that the distribution of taxes is progressive (not progressive), will increase (decrease) their willingness to pay tax if they would prefer (do not prefer) lower levels of inequality

Hypothesis E2: Informing people that the distribution of transfers is progressive (not progressive), will increase (decrease) their willingness to pay tax if they would prefer (do not prefer) lower levels of inequality

Hypothesis E3: Informing people that the net impact of taxes and transfers is progressive (not progressive), will increase (decrease) their willingness to pay tax if they would prefer (do not prefer) lower levels of inequality

Group F – People’s willingness to pay tax varies by their perceptions about whether their household is a net contributor to or beneficiary from the tax and transfer system

Hypothesis F1: Informing people that the distribution of taxes is progressive (not progressive), will increase (decrease) their willingness to pay tax if they perceive their household is a net beneficiary (contributor)

Hypothesis F2: Informing people that the distribution of transfers is progressive (not progressive), will increase (decrease) their willingness to pay tax if they perceive their household is a net beneficiary (contributor)

Hypothesis F3: Informing people that the net impact of taxes and transfers is progressive (not progressive), will increase (decrease) their willingness to pay tax if they perceive their household is a net beneficiary (contributor)

## Treatments

The treatments have been designed to provide people with accurate information about the distribution of taxes and/or government transfers in each country. The treatments draw on data provided by CEQ (2021) and for simplicity they present the distribution of taxes and/or transfers across quintiles (whereas the CEQ database focuses on deciles). Table 1 summarizes variations in the degree of progressivity in the tax and transfer systems (taking into account both direct and indirect taxes and government transfers) across the eight countries in this study according to the CEQ database. The tax system is progressive in four of the eight countries (Colombia, Mexico, Ghana and Tanzania) and the transfer system is progressive in six of the eight countries (Colombia, Mexico, Indonesia, Jordan, Sri Lanka, South Africa). The net impact of taxes and transfers is progressive in all countries, however in Ghana and Tanzania the net impact is negative across all quintiles (i.e., all households pay more in tax than they receive in transfers). The exact treatments (in English) that will be provided in each country are contained in Appendix A.

The treatments were designed to provide an indication of whether taxes and/or transfers were progressive in their country, but not provide information about the level of taxes and transfers as a percentage of household income. This is because we would not be able to as clearly isolate the channels through which the treatments were impacting people's willingness to pay tax if we combined information about both the distribution and level of taxes and transfers. For example, if both aspects were included it is possible some respondents may be reacting to the degree of progressivity, while others were reacting to the level of taxes and transfers, and we would not be able to differentiate. Efforts were also made to make sure that respondents were more likely to trust the content of the treatments by following a similar approach to seminal work by Alesina et al. (2018). For example, respondents were informed that the information they were provided with recently became publicly available online through a collaboration between universities, civil society and international organisations (see Lustig, Mariotti and Sánchez-Páramo (2020)).

Survey respondents in each country will be randomly allocated either to one of three treatment groups or to a control group that receives no information (i.e., the multiple treatment arms are exclusive of one another). The first treatment (hereafter the "taxes treatment") will provide information from the CEQ database about the distribution of taxes (both direct and indirect taxes, such as income tax and value added tax) in their country. The second treatment (hereafter the "transfers treatment") will provide information from the CEQ database about the distribution of government transfers (both direct and indirect transfers, such as cash payments and energy subsidies). The third treatment (hereafter the "taxes and transfers treatment") will provide information from the CEQ database about the net effect of the distribution of taxes and government transfers. Hypotheses A1, B1, C1, D1, E1 and F1 will be tested by the taxes treatment. Hypotheses A2, B2, C2, D2, E2 and F2 will be tested by the transfers treatment. Hypotheses A3, B3, C3, C4, D3, D4, E3 and F3 will be tested by the taxes and transfers treatment.

**Table 1 – Progressivity of the tax and transfer system in each of the countries in this study**

	Quintiles	Share of household income		
		Taxes	Transfers	Net
Colombia (2014)	1	3.9%	13.5%	9.6%
	2	5.2%	4.6%	-0.6%
	3	6.5%	2.4%	-4.1%
	4	7.8%	1.2%	-6.6%
	5	8.7%	0.3%	-8.4%
Ghana (2012)	1	6.9%	3.0%	-3.9%
	2	7.4%	2.3%	-5.0%
	3	8.1%	2.3%	-5.8%
	4	9.4%	2.5%	-6.9%
	5	13.4%	2.1%	-11.2%
Indonesia (2017)	1	7.1%	12.1%	5.0%
	2	7.8%	11.5%	3.8%
	3	7.7%	10.7%	3.0%
	4	7.6%	9.4%	1.8%
	5	9.5%	7.3%	-2.2%
Jordan (2017)	1	11.8%	26.5%	14.7%
	2	10.9%	13.9%	3.0%
	3	10.6%	10.9%	0.3%
	4	10.2%	8.2%	-2.0%
	5	10.5%	4.2%	-6.3%
Mexico (2014)	1	6.8%	32.3%	25.5%
	2	8.8%	9.8%	1.0%
	3	10.7%	5.8%	-4.8%
	4	12.4%	3.8%	-8.6%
	5	15.7%	1.9%	-13.9%
Sri Lanka (2009)	1	10.1%	20.9%	10.8%
	2	9.1%	7.1%	-1.9%
	3	9.0%	5.9%	-3.1%
	4	8.9%	4.7%	-4.2%
	5	8.5%	2.9%	-5.6%
Tanzania (2011)	1	9.6%	1.3%	-8.2%
	2	10.2%	1.2%	-9.1%
	3	11.2%	1.1%	-10.1%
	4	13.5%	1.3%	-12.2%
	5	25.0%	1.4%	-23.6%
South Africa (2010)	1	104.9%	567.1%	462.2%
	2	35.1%	120.2%	85.1%
	3	21.0%	28.0%	7.0%
	4	19.8%	5.7%	-14.2%
	5	30.6%	0.4%	-30.2%

*Note: The year shown in brackets next to each country is the year in which the survey took place that the GINI index is based on.*

*This table shows the average share of household income for each quintile that is directly or indirectly paid in taxes or received in government transfers as well as the net impact of taxes and transfers on household income (CEQ, 2021).*

## **Outcome variables**

All of the groups of hypotheses outlined above have the same five outcome variables (see full survey instrument in Appendix B). The first outcome variable (based on Q14) is a straightforward measure of people's willingness to pay tax that asks respondents whether they would pay tax if they knew that they would not be caught for non-compliance (i.e., this question directly measures respondents' willingness to pay tax in the absence of punishment, which traditionally was conceptualized as the only reason to pay). This question is based on the experience of the World Bank Innovations for Tax Compliance program, which has extensively used surveys to measure willingness to pay tax. A potential shortcoming of this simple measure of "quasi-voluntary" tax compliance is that people may be unwilling to provide honest answers and as a result the share of respondents claiming they will pay tax in the control group (i.e., in the absence of additional information provided through the treatment) will almost certainly be higher than what will actually be the case. Therefore, at best this outcome variable will provide a lower bound estimate of the impact of the treatments of people's willingness to pay tax.

The remaining four outcome variables capture slightly different aspects of indirect measures of people's willingness to pay tax (sometimes referred to as "tax morale"), which have been sourced from existing studies in low- and middle-income countries on this topic. The second outcome variable (based on Q15) measures the degree to which respondents believe people not paying is understandable; it was used in the Afrobarometer (2012; 2013; 2015) as well as by Ali, Fjeldstad and Sjursen (2014). The third outcome variable (based on Q16) measures whether people believe paying taxes is important and was used by Khwaja et al. (2020). The fourth outcome variable (based on Q17) measures people's unconditional beliefs about the extent to which the government has the right to make people pay taxes and it has been included in many rounds of the Afrobarometer (2002; 2003; 2004; 2008; 2012; 2013; 2014; 2015; 2017). The fifth and final outcome variable measures the degree to which people believe that paying tax should be conditional on what the government spends tax on and is a slightly modified version of what was used by Prichard, Jibao, and Orgeira (forthcoming). How each of the outcome variables are coded and analyzed based on answers to the corresponding questions is discussed in Section 4.

## **Variables measuring people's prior beliefs and preferences**

Prior to the treatments, respondents will be asked to provide information about their perception of their household's position in the income distribution in their country (Hypotheses Group B), their perceptions of and their preferences for the distribution of taxes and government transfers in their country (Hypotheses Groups C and D), prior perceptions of and preferences for the level of inequality in their country (Hypotheses Group E) as well as whether people view their households as being net contributors to or beneficiaries from the tax and transfer system (Hypotheses Group F). These questions have either been sourced from existing studies in a series of low- and middle-income countries (Q4 and Q5) (e.g., see Hoy and Mager, 2021) or were specifically developed for this study (Q6–Q11). The questions developed for this study were based on the structure of standardized questions in the literature (e.g., they use a Likert scale), informed by expert feedback and modified based on the piloting process to ensure these new questions were adequately comprehended by respondents (see Section 3 for details of the piloting process).

### 3. Data

#### Selection of countries

The eight countries focused on in this study were selected for the following reasons. Firstly, there is limited data available about the distribution of taxes and government transfers in low- and middle-income countries. By far the largest effort that has been made to collect and disseminate this information has been through the Commitment to Equity Institute at Tulane University, which is headed by Nora Lustig (CEQ, 2021). Estimates have been produced of the difference between the gross and net GINI index in over 60 low- and middle-income countries through this work program (see Figure 1). These estimates are based on standardized household income and expenditure surveys and in 2020 a cross-country database that provided disaggregated information in a standardized way for many countries was publicly released through a joint initiative between universities, civil society and international organizations (Lustig, Mariotti and Sánchez-Páramo, 2020). However due to a range of factors, including governments' reluctance to make certain information publicly available, information about the distribution of direct and indirect taxes as well as direct and indirect government transfers (including subsidies) is limited to a far smaller subset of these countries. This subset of countries was the starting point for selecting which countries to include in this study.

Secondly, the time and costs involved in collecting data online in low- and middle-income countries are considerably lower when there is a high internet population in absolute terms. As such, countries with high populations and/or high internet penetration rates were focused on as part of this study. As can be seen in Table 2, some countries like Tanzania have higher total populations, but lower internet penetration rates, whereas other countries like Jordan have lower total populations but higher internet penetration rates.

**Table 2 – Population with access to the internet in each of the countries in this study**

Country	Total Population (mil)	Internet coverage (%)	Internet Population (mil)
Indonesia	274	54	148
Sri Lanka	22	35	8
Tanzania	60	20	12
South Africa	53	68	36
Ghana	31	53	16
Jordan	10	67	7
Colombia	51	65	33
Mexico	129	72	93

*This table shows the total population (millions of people), internet coverage (% of population) and internet population (millions of people) in each of the countries in this study (World Bank, 2021).*

Thirdly, due to funding reasons it was necessary to collect a diverse set of countries in each of the major regions with low- and middle-income countries (i.e., Latin America, West Africa, East Africa, the Middle East, South Asia and East Asia) as well as across various income levels (i.e., low, lower-middle and upper-middle income countries). This restricted the choice set considerably in some regions; for example, Indonesia was the only country in East Asia with publicly available data about the distribution of direct and indirect taxes, as well as direct and indirect government transfers (including subsidies).

Finally, efforts were made to ensure that the information about the distribution of taxes and government transfers in the database was still likely to provide a reasonable estimate of what would exist in early 2022. There were some countries, such as Iran, where there have been significant changes to the tax and transfer system since the survey included in the database took place and as a result it would not be a realistic approximation of how taxes and government transfers were likely to be distributed in early 2022.

### **Approach to data collection**

Ideally, face-to-face surveys collecting a representative sample of the general population using a sampling frame, such as a recent census, would be conducted in each of the countries in this study. Not only are the costs involved in doing this prohibitive, but there are also issues with face-to-face surveys during a pandemic. While phone surveys present a popular alternative, this is not an appropriate format for a survey along these lines. The treatments are designed to be visual in nature and it is not possible to communicate these messages fully via a phone call. This left an online survey as the most promising option for data collection, even though there are challenges with representativeness that need to be recognized and can be overcome to some extent.

A major challenge in conducting an online randomized survey experiment in low- and middle-income countries is collecting a representative sample of the total population. Unlike high-income countries where internet access is near universal, the share of the total population with internet access in the countries in this study varies from 25 to 72 percent (see Table 2). Furthermore, there is limited existing online survey “infrastructure”, such as what exists in many high-income countries where market research firms run online opinion polls daily from a large pool of pre-registered respondents who regularly complete surveys. This is far less common in low- and middle-income countries and there are reasons to be concerned about just how narrow a subset of the population would participate in an engagement like this. Similar concerns exist regarding the use of online labor platforms, such as MTurk, in a low- and middle-income country context.

Alternative approaches to online data collection in low- and middle-income countries can crudely be categorized as providing “opt-in” or “opt-out” options. An example of the former would be to advertise to social media users that they could participate in an online survey and receive compensation for doing so. While this “opt-in” approach may be attractive for soliciting responses, we identified at least two shortcomings that we felt meant this approach was not ideal for this study. Firstly, there is a clear concern regarding selection bias as people who would “opt in” to a survey based on a social media advertisement potentially have some unobservable characteristics that make them distinct from the rest of the population. It is challenging to estimate the extent to which these unobservable characteristics exist without gaining access to administrative data from social media providers. Secondly, as we are asking about a sensitive topic, tax compliance, it is possible that we would experience experimenter demand effects if participants could easily be tracked by opting into an online survey through a platform on which they were identifiable. As such, on balance we felt that an “opt-in” approach along these lines would not be suitable for our study.

We decided to collect data online using an “opt-out” approach offered by the survey firm, RIWI. They capture a sample of respondents that is broadly representative of the internet population in each country by using Random Domain Intercept Technology. This involves sampling internet users who incidentally access inactive domains (i.e., which would otherwise result in a “404 error”). As domain names regularly change and they often do not automatically redirect internet users it is commonplace for the internet using population to incidentally access inactive domains. Research suggests the likelihood of accessing an inactive domain is approximately proportional to having access to the internet (IRIS, 2021). RIWI exploits this by redirecting users from inactive domains to a website inviting them to take part in a survey. At this point people can decide whether to continue to participate in the survey or “opt out”. RIWI tracks information about the device used and operating system used by people who are redirected towards to the survey platform. In addition, the first question people are asked is about their age and sex. As a result, we can observe how “opt-out” rates from a representative sample of the internet population varies based on the characteristics of respondents (for example, we will be able to measure whether people using smartphones disproportionately opt-out of the survey, even if they do not answer a single question). A shortcoming of this “opt-out” approach is that high rates of attrition are expected early in the survey. However, given we can track how attrition varies by the characteristics of respondents and the survey experiment is at the back end of the survey, this does not undermine the integrity of the study. We discuss in Section 4 how we will factor this attrition into our analysis. Table 3 presents the expected characteristics of respondents and the survey participation rates based on RIWI’s prior experience in these countries.

**Table 3 – Expected characteristics of respondents and survey participation rates**

<b>Countries</b>	<b>Share of respondents that are male</b>	<b>Share of respondents that are aged between 18–35 years old</b>	<b>Share of respondents that are exposed to the survey platform that at least complete the first question</b>
Indonesia	76%	71%	27.0%
Sri Lanka	68%	70%	19.4%
Tanzania	68%	75%	32.2%
South Africa	64%	66%	10.6%
Ghana	71%	79%	25.6%
Jordan	61%	70%	8.6%
Mexico	63%	61%	24.2%
Colombia	61%	62%	27.5%

*This table shows the expected characteristics of respondents and participation rates based on RIWI’s experience in these countries.*

## **Details about data collection**

This study solely relies on data being collected online by the survey firm RIWI. The survey instrument in English is provided in full in Appendix B. This survey instrument has been translated into the following languages: Spanish (for use in Mexico and Colombia), Arabic (for use in Jordan), Bahasa (for use in Indonesia), Swahili (for use in Tanzania) as well as Sinhala and Tamil (for use in Sri Lanka).



In each country, we will collect completed surveys for 3600 respondents as well as all responses provided by partial respondents (i.e., those that drop out part way through the survey). In addition, we will collect information about the device used and operating system of people who are exposed to the survey but did not participate. It is expected that respondents will take around 10 minutes to complete all 20 questions and the surveys will be in the field for 6 to 8 weeks in each country (until the agreed sample size is reached).

## Sample and statistical power

The unit of analysis for this study is the individual level. The sample size in each of the eight countries is 3600 respondents who fully complete the survey. In total, this equates to 28,800 respondents. The sample size of 3600 respondents in each country (900 in each treatment and the control group) was determined by examining the sample size of other cross-country randomized survey experiments in this field as well as conducting statistical power calculations. In the case of the seminal cross-country randomized survey experiments by Alesina et al. (2018) and Alesina et al. (2021) they also included around 900 to 1000 respondents in each treatment group in each country. Other studies in this field have tended to have smaller sample sizes (e.g., Cruces et al., 2013; Karadja et al., 2017) and have still detected significant heterogeneous treatment effects based on prior beliefs. Furthermore, a *Journal of Economic Literature* article summarizing best practices in online randomized survey experiments providing information interventions suggests that having in the order of 700 to 800 respondents per treatment/control group should provide adequate power to detect an effect that is commonly observed in the literature (Haaland et al., forthcoming).

In terms of the specific statistical power calculations, we considered two broad scenarios (STATA output presented in Appendix C). The first is to consider the minimum effect we would be able to detect (with the standard probability of type one and type two errors of 0.05 and 0.2) from a single treatment on a single binary outcome (using the example of collapsing a Likert scale into a dummy variable that takes the value of 1 for strongly agree or agree and 0 otherwise). This corresponds with testing the hypotheses in Group A in one country. We could detect an effect in the order of six and a half percentage points based on the conservative assumption of 50 percent of the control group agreeing with the statement in the question.

The second is to consider the minimum effect we would be able to detect on a subset of respondents from a single treatment on a single binary outcome. This corresponds with testing the hypotheses in Groups B–F in one country. In this case, the underlying sample size in each of the treatment and control groups will shrink based on the prior beliefs and preferences of respondents (e.g., whether they believe their household is poor). We consider three examples, the first covering 75 percent of all respondents, the second covering 50 percent of all respondents and the third covering 25 percent of all respondents. We present the effect size we could detect with these smaller sample sizes in Table 4, once again based on the conservative assumption of 50 percent of the control group agreeing with the statement in the question.

**Table 4 – How the minimum detectable effect in a single country varies based on sample size**

Number of respondents per treatment/control group with a particular prior belief/preference	900 (100%)	675 (75%)	450 (50%)	225 (25%)
Minimum detectable effect (Percentage points)	6.7	7.7	9.5	13.5

*This table shows how the minimum detectable effect in a single country varies based on the sample size.*

There are multiple reasons to believe that we will potentially have much higher statistical power to detect effects than what is presented in Table 4. Firstly, in the instance whereby the mean in the control group for a binary outcome is higher or lower than 0.5, the minimum detectable effect will be smaller. Secondly, by pooling data across countries (with country fixed effects as discussed in Section 4) we will have substantially more power to detect an impact from the treatments, however it is hard to quantify this ex-ante as it will depend on variation between countries. Thirdly, we will create an index which is an unweighted Z-score for the outcomes of interest (captured in Q14–Q18) and consequently it will be easier to detect an effect on this index than on a single outcome. However, the extent to which this is the case will vary depending on how consistent the treatment effects are across each of the outcomes. Fourthly, we anticipate we will have slightly higher sample sizes for at least some of the outcome variables because we will include partial respondents in the analysis. Finally, we may pool treatments together, which would increase the sample size and consequently the statistical power. However, we do not anticipate this will be a core part of our analysis (see Section 4).

## **Pilot data**

The proposed study outlined in this registered report went through an extensive review process within the World Bank prior to being piloted in December 2021. The internal review process identified ways in which the survey instrument could reflect best practice in the literature (e.g., avoiding ceiling effects on the outcome variables by phrasing questions to ensure greater variation of responses across a Likert scale). Reviewers also emphasized that during the piloting process it will be crucial to examine whether respondents adequately comprehend the treatments and the questions. As such the primary focus of the piloting that took place was to ensure that the responses that were gathered indicated the respondents understood the survey instrument. In addition, piloting provided an opportunity to verify the assumptions made about the size of the treatment effects in the statistical power calculations and to identify ways in which the experiment could be designed in a manner to minimize attrition. These three issues are discussed one by one below following a description of the piloting process.

### Implementation of the piloting process

The survey instrument and experiment were piloted with 1061 respondents (who completed the survey) that made up a representative sample of the internet population in India in December 2021. India was selected as an appropriate location to pilot the survey as this is where the survey firm typically conducts pilots (due to the diverse, but very large, population where English is commonly used on the internet); it has a similar level of development to many of the countries in the full study and as we were not including India in the full study<sup>3</sup>, we did not need to be concerned about contaminating the pool of respondents.

There were two phases to the pilot. The first phase involved using visual stimuli for some of the questions capturing people's prior beliefs and preferences about the distribution of taxes and transfers as well as levels of inequality in their country (see Appendix D). In this version of the survey instrument that had been approved through the internal review process at the World Bank, respondents were required to select the distribution of taxes and transfers that exists in their country based on actual examples. Specifically, the options provided for respondents to select from were based on the actual progressivity of taxes in Tanzania

---

<sup>3</sup> As recent data is not available about the distribution of taxes and transfers in India, the treatments that were provided were based on what is actually the case in Mexico.

in 2011, Colombia in 2014 and Jordan in 2017. In addition, respondents were randomly allocated to receive questions from a pool of seven potential questions about their willingness to pay tax. This process helped to inform which five questions should be included in the full study. In total, 511 respondents completed this phase of the pilot.

The main change in the second phase of the pilot was replacing the questions from the first phase that involved visual stimuli with basic questions that aimed to capture people's prior beliefs and preferences about the distribution of taxes and transfers as well as levels of inequality in their country on a Likert scale (see Appendix E). This approach brought the format of these questions into line with the rest of the survey. A shortcoming of this approach was that it was no longer possible to identify whether people's beliefs and preferences matched examples of the actual level of progressivity of taxes in some low- and middle-income countries. In the second phase of the trial, respondents continued to be randomly allocated to receive a subset of questions about their willingness to pay tax. In total, 550 respondents completed this phase of the pilot.

#### Lessons learnt through the piloting process

There were three key lessons that emerged from the two phases of the pilot that informed the final survey instrument. Firstly, there was a clear need to keep the survey instrument as simple as possible. Answers to the questions that included visual stimuli in the first phase of the trial suggested respondents did not adequately comprehend the options they were presented with. Responses were very evenly distributed across the options in each of the four questions about people's beliefs and preferences in regard to the distribution of taxes and transfers in their country. To test whether this was primarily due to measurement error, in the second phase of the pilot respondents were randomly allocated to either the question format from phase one or basic questions about their views on how taxes and transfers are distributed using a Likert scale. The results were substantially different between these approaches with the basic question format returning results far more consistent with previous literature. Specifically, the results showed most people tend to prefer richer households to pay more taxes than poorer households and poorer households to receive more government transfers than richer households (i.e., most people tend to prefer progressivity in the tax and transfer system). As such we decided that the final survey instrument should rely on these basic questions to capture people's prior beliefs and preferences, even though this means that the options provided are not based on actual progressivity of taxes and transfers in countries. We believe that capturing higher quality, reliable responses is of greater importance.

Secondly, the results of the piloting process provide us with confidence that the statistical power calculations we based our sample sizes on will be more than adequate. The point estimates of the treatment effects are quite promising as they indicate variation between respondents in the treatment and control groups of an order of magnitude that we will be powered to detect at standard levels (i.e., an alpha of 0.05 and beta of 0.2) when the full sample of respondents is reached (i.e., 3600 as opposed to 1061). The main results tables for each of the hypotheses based on the pilot data are shown in Appendix F. It is important to note that due to differences in the inclusion of certain questions across the two phases of the trial there is considerable variation in sample sizes between the tables. The direction of the treatment effects is also often in line with the primary hypotheses (discussed in Section 2). This is reassuring but given we are only testing treatments that indicate the tax and/or transfer system is progressive it is important not to overemphasize these preliminary indications.

Thirdly, the piloting process highlighted ways to minimize attrition during the survey experiment and the most straightforward way was by removing list experiments from the study. Specifically, there was low attrition for the outcome variable questions included in the final survey instrument (in the order of four percent per question), whereas around one quarter of respondents dropped out during the two list experiments. Removing the list experiments from the randomized survey experiment was not a major issue for our study as there is debate in the literature about the value of this approach in general (e.g., see Chuang et al., 2021; Agerberg and Tannenberg, 2021) and we would have potentially faced considerable issues with inadequate statistical power. In the second phase of the pilot we also randomized alternative data quality check questions between respondents immediately prior to the survey experiment and found that our original question from the first phase of the pilot outperformed an alternative question that was used by Alesina et al. (2018). As such, we feel confident that including a question that asks respondents to drop out prior to the treatment if they are unwilling to complete the survey experiment will serve as an effective way to minimize attrition post treatment. We were also reassured by the lack of differential attrition observed throughout the piloting process.

## 4. Analysis

Two main types of analysis will be conducted for this study. The first type of analysis is an examination of the descriptive statistics from the survey, which will be presented in the data section of the paper. The second type of analysis is a presentation of the findings of the randomized survey experiment, which will be presented in the results section of the paper. The focus of this section of the registered report is on the second type of analysis, as are the main results tables based on the pilot data in Appendix F and the STATA do file that was used to produce these tables in Appendix G.

### Coding of variables

**Table 5 – Coding of responses to the survey**

Question	Coding of question
Q0 – Age and Gender	age1834= 1 if respondent aged 18–34 years (respondents under 18 automatically discarded), 0 if respondent aged 35 years or older male = 1 if respondent male, 0 otherwise
Q1 – Education	edu_secorless = 1 if respondent selected primary or secondary education, 0 otherwise
Q2 – Location	large_city = 1 if respondent selected large city or suburb, 0 otherwise
Q3 – Employment type	working= 1 if respondent selected employee or self-employed/small business owner, 0 otherwise
Q4 – Prefer lower inequality	lower_ineq = 1 if respondent selects strongly agree or agree, 0 otherwise

Q5 – Perceived position in national income distribution	pB40 = 1 if respondent selected poorest or second poorest quintile, 0 otherwise
Q6 – Household paid large share of income in tax	largetax = 1 if respondent selects strongly agree or agree, 0 otherwise
Q7 – Household paid more in tax than received in transfers	net_cont = 1 if respondent selects strongly agree or agree, 0 otherwise
Q8 – Perceived taxes as currently progressive	cur_prog_tax = 1 if respondent selects strongly agree or agree, 0 otherwise
Q9 – Prefer taxes to be progressive	prog_tax = 1 if respondent selects strongly agree or agree, 0 otherwise
Q10 – Perceived transfers as currently progressive	cur_prog_trans = 1 if respondent selects strongly agree or agree, 0 otherwise
Q11 – Prefer transfers to be progressive	prog_trans = 1 if respondent selects strongly agree or agree, 0 otherwise
Q12 – Data quality check	willcomplete = 1 if respondent selected yes, 0 otherwise
<i>TREATMENT PROVIDED (Will be labeled as Q13 in the data)</i>	
Q14 – Will not pay without enforcement	willpaytax = 0 if respondent selects strongly agree or agree, 1 otherwise
Q15 – Not paying tax is wrong and punishable	wrong_punish = 1 if respondent selected wrong and punishable, 0 otherwise
Q16 – Paying taxes is important	importanttopay = 1 if respondent selects strongly agree or agree, 0 otherwise
Q17 – Government has right to pay tax	righttotax = 1 if respondent selects strongly agree or agree, 0 otherwise
Q18 – Refuse to pay tax	donotrefuse_paytax = 1 if respondent selects strongly disagree or disagree, 0 otherwise

*This table provides the exact details regarding how we intend to code answers to each of the questions in the survey instrument.*

The survey experiment has only five outcomes, which means that the risk of multiple hypothesis testing being an issue is very low. However, to address potential concerns we will aggregate all five outcomes into an index. Specifically, we will follow a similar approach as to what was used in related randomized survey experiments by Alesina et al. (2018) and Karadja, Mollerstrom and Seim (2017), by creating a “Willingness to Pay Tax” Index. This index will be the unweighted average of the Z-scores of all five outcome variables

defined in Table 5 (willpaytax, wrong\_punish, importanttopay, righttotax, donotrefuse\_paytax), oriented so that a higher index means more willingness to pay tax. We will present the answers to each outcome question and the “Willingness to Pay Tax” Index in the tables of results.

## **Descriptive statistics**

This study provides a unique opportunity to identify stylized facts regarding perceptions about and preferences for the progressivity of tax and transfer systems across countries as well as how these prior beliefs and preferences relate to people’s willingness to pay tax. These descriptive statistics will be presented in the data section of the paper in the following manner:

1. Perceptions of the level of progressivity of the tax and transfer system in each country (based on Q8 and Q10 in the survey instrument in Appendix B) – Comparisons will be made with what is actually the case (i.e., this will illustrate in a qualitative manner how accurate people’s prior beliefs are). In addition, the factors (based on Q0–Q7) associated with believing the tax and transfer system is progressive or regressive will be examined using multivariate OLS regressions. The purpose of this analysis is to examine if stylized facts exist across countries, as it is not possible to establish any causal link.
2. Preferences for the level of progressivity of the tax and transfer system in each country (based on Q9 and Q11 in the survey instrument in Appendix B) – Comparisons will be made with what is actually the case (i.e., this will illustrate in a qualitative manner how well the existing distribution of taxes and transfers is in line with people’s preferences). In addition, the factors (based on Q0–Q7) associated with desiring the tax and transfer system to be more or less progressive/regressive will be examined using multivariate OLS regressions. The purpose of this analysis is to examine if stylized facts exist across countries, as it is not possible to establish any causal link.
3. Preferences for greater progressivity of the tax and transfer system in each country (based on differences between Q8 and Q9 as well as differences between Q10 and Q11 in the survey instrument in Appendix B) – The share of respondents that desire more or less progressivity in the tax and transfer system than what they perceive to exist will be determined. In addition, the factors (based on Q0–Q7) associated with desiring the tax and transfer system to be more progressive or regressive than what respondents perceive it to be, will be examined using multivariate OLS regressions. The purpose of this analysis is to examine if stylized facts exist across countries, as it is not possible to establish any causal link.
4. People’s willingness to pay taxes in each country (based on the control groups’ answers to Q14–Q18 in the survey instrument in Appendix B) – Comparisons will be made with what is actually the case in terms of tax revenue as a share of GDP and other measures of tax system performance, such as those by TADAT. In addition, the factors (based on Q0–Q7) associated with willingness to pay tax will be examined using multivariate OLS regressions. The purpose of this analysis is to examine if stylized facts exist across countries, as it is not possible to establish any causal link.
5. Relationship between people’s willingness to pay and prior beliefs and preferences for greater progressivity of the tax and transfer system (based on the control groups’ answers to Q8–Q18 in the survey instrument in Appendix B). The purpose of this descriptive analysis is to identify stylized facts in terms of the relationship between what is covered under point 4 and points 1, 2 and 3 above. In line with the conceptual framework and resulting groups of hypotheses it is expected that the clearest relationship will hold in terms of people being more willing to pay tax if they believe the tax and transfer system is progressive.

## Main statistical analysis

As discussed above, we conduct a randomized survey experiment to test the impact of accurate information about the distribution of taxes, government transfers or both on willingness to pay tax. Randomization allows for the impact of the treatments to be determined by comparing differences in mean outcomes between the control group and treatment groups. We stratify the randomization process by the age and sex of respondents. We will present balance tables for each country based on all answers provided prior to the treatment (see Q0–12 in Table 4), including measures of both individual and joint significance (i.e., both t-statistics for every variable and an f-statistic across all variables within a given country). As outlined in our discussion regarding the sample size, we have designed the survey so that we will have adequate statistical power (with the conventional levels of type one and two errors of 0.05 and 0.2) to detect a meaningful effect from the treatments if they exist. A threat to the internal validity of the experiment is differential attrition and we discuss ways to address this if it were to arise in the robustness checks subsection.

The first and most straightforward statistical analysis we will conduct in each country involves using an Ordinary Least Squares (OLS) regression in the form of a linear probability model. This will involve creating a dummy variable for each outcome of interest discussed above ( $Y_j$ ) and a dummy variable for each treatment group that takes on the value one if the respondent belongs to the specific treatment group and the value zero if the respondent belongs to the control group. This can be expressed formally as follows:

$$Y_j = \beta_0 + \beta_1 T_1 + \beta_2 T_2 + \beta_3 T_3 + X\gamma + \varepsilon, \quad (7)$$

where  $\beta_1$  captures the average difference between respondents in the taxes treatment group and the control group in regards to the outcome of interest ( $Y_j$ ),  $\beta_2$  captures the average difference between respondents in the transfers treatment group and the control group in regards to the outcome of interest ( $Y_j$ ) and  $\beta_3$  captures the average difference between respondents in the taxes and transfers treatment group and the control group in regards to the outcome of interest ( $Y_j$ ). Further,  $X$  is a vector of variables that controls for potential imbalance in background characteristics (see Q0–Q4 in Table 5) between treatment and control groups,  $\varepsilon_i$  is the model error term and  $\beta_0$  is the intercept.

Three additional types of analysis will be taken to determine our main results. Firstly, as per our hypotheses, we also explore heterogenous effects by conducting the regression analysis outlined in Equation (7) on subsets of respondents based on their responses provided prior to the treatments (see Table 6 for details about how the subsets of respondents are defined). For example, we will examine Group B of hypotheses by conducting the regression analysis outlined in Equation (7) on respondents who perceive they are in the poorest or second poorest quintile and then separating reproducing this analysis on respondents who did not perceive they are in the poorest or second poorest quintile (i.e., they perceived they were in the richest three quintiles). Secondly, we also present the main results of the survey experiment using pooled OLS regressions with country fixed effects across all countries. This approach was used by Hoy and Mager (2021). Thirdly, to maximize statistical power, we intend to combine these first two analyses so we will examine heterogenous treatment effects with pooled OLS regressions with country fixed effects. This third type of statistical analysis is in line with what was undertaken by Alesina et al. (2018) and Alesina et al. (2021).

**Table 6 – Variables corresponding with each of group of hypotheses**

<b>Group of Hypotheses</b>	<b>Variable defining subsets of respondents</b>
Group B – People’s willingness to pay tax varies based on their perceived place in the income distribution	pB40 = 1 if respondent selected poorest or second poorest quintile in Q5, 0 otherwise
Group C - People’s willingness to pay tax varies by their prior beliefs about the progressivity of tax and transfer system	cur_prog_tax = 1 if respondent selects strongly agree or agree to Q8, 0 otherwise cur_prog_trans = 1 if respondent selects strongly agree or agree to Q10, 0 otherwise
Group D – People’s willingness to pay tax varies by their preferences for the progressivity of tax and transfer system	prog_tax = 1 if respondent selects strongly agree or agree to Q9, 0 otherwise prog_trans = 1 if respondent selects strongly agree or agree to Q11, 0 otherwise
Group E – People’s willingness to pay tax varies by their preferences for the level of inequality	lower_ineq = 1 if respondent selects strongly agree or agree to Q4, 0 otherwise
Group F – People’s willingness to pay tax varies by their perceptions about whether their household is a net contributor to the tax and transfer system	net_cont = 1 if respondent selects strongly agree or agree to Q7, 0 otherwise

*This table provides details about how the subsets of respondents are defined that correspond with each group of hypotheses.*

## **Robustness Checks**

We have listed below the series of robustness checks we will conduct to examine the strength of our results.

### Alternative statistical models

As discussed above our main analysis will be based on OLS regressions, but we will also examine whether using alternative statistical models would have a material impact on our findings. Specifically, we intend to rerun our main analysis with a logit regression. In addition, because our questions are primarily ordered and categorical we will reconstruct the main outcome variable so that an ordered logit regression can be run. Based on previous studies we do not anticipate either of these alternative statistical models will substantively change our results. As such we will include them in the Appendix of the Stage 2 Report.

### Data quality checks

It is possible that some respondents may not devote their full attention to the survey and to detect whether this impacts our results we will conduct three main robustness checks. Firstly, we will rerun our analysis after excluding respondents that took too short or long a time to complete the survey. Specifically, we will winsorize at the 1, 2.5 and 5 percent levels (among respondents that complete the survey) based on time to completion.



Secondly, we follow Alesina et al. (2018) and include an explicit data quality check (question 12). We ask respondents if they are willing to fully participate in the final five questions of the survey. We exclude all respondents that do not confirm this. This question serves as a way to prompt respondents to regain their full attention prior to beginning the experiment. In addition, by stating there are only five questions remaining this reduces the likelihood of attrition once the survey has begun.

Thirdly, we will use specific questions in the survey to conduct an implicit data quality check. We will rerun our main analysis and exclude respondents that select the identical answer to questions Q6–Q11 (about their perceived and preferred levels of progressivity in the tax and transfer system as well as whether they believe their household is a net contributor or beneficiary). While it is possible that some respondents answers are accurate, for many respondents it is likely they will not have fully appreciated the differences between these questions.

### Weighting

The sample of respondents to this survey experiment are broadly representative of the internet population in each country, however this is not representative of the entire population. We reweight the results to investigate the extent to which the results may differ if the sample was based on the entire population. At a minimum we will reweight on age and sex of the adult population as these measures are more straightforward to compare across our survey and the underlying population in each country (according to the most recent census). We will also conduct a more advanced weighting exercise whereby we factor in the education of the respondent and whether they are in a rural/urban area alongside age and sex. One final reweighting exercise we will conduct is based on the characteristics of respondents who are exposed to the survey (based on device type and operating system) and/or begin the survey (age and sex), but do not complete the survey. By doing so we will check whether our main results are partly driven by the characteristics of respondents who are more likely to complete the survey.

### Attrition

The sampling strategy used in this study (see data section for details) will generate high rates of attrition early in the survey because respondents are more likely to drop out because they did not opt in to participate. We believe this sampling strategy is still preferable compared to alternative online survey options as our initial sample is more representative. In anticipation of this higher level of attrition, we ask 13 questions prior to the experiment. As such as much as possible we have tried to “frontload” the level of attrition so that it will not influence the internal validity of the survey experiment.

To minimize attrition in the survey experiment, we ask respondents if they are willing to fully participate and exclude all respondents that do not confirm this (Q12). By stating there are only five questions remaining it is expected this will reduce the likelihood of attrition once the treatments are provided. Our pilot data confirms this.

To ensure even a small amount of attrition does not drive differences between outcome variables, the order of the five questions in the survey experiment is randomized. As such, on average the number of observations for each of the outcomes should be quite similar in each country.

In the event that differential attrition occurs during the survey experiment, we will produce Lee bounds estimates of our main results (Lee, 2009). By doing so we will be able to provide guidance as to whether differential attrition is an issue of concern. This is the same approach as what has been taken by other studies in this literature (e.g., Kuziemko et al., 2015; Hoy and Mager, 2021).

High levels of attrition at the start of the survey also influence how we treat missing values. The contract with the survey firm specifies that at least 3600 surveys are completed in each country, however we will also be provided with all incomplete responses. As such for some parts of our analysis, particularly the descriptive statistics, we may have far more responses to some of the questions early in the survey. Our default approach is to use all data available to us (i.e., both the fully complete and partially complete data). However, as a robustness check we will rerun our main analysis using only answers from respondents who complete the entire survey.

## 5. Interpreting Results

The results of this study could fall into one of the following three categories.

### No meaningful impacts from the treatments

It is possible that there will be no impact from any of the treatments on people's willingness to pay tax. This could be due to either a lack of empirical support for the existing theory or issues with the study design.

#### 1. Lack of empirical support for the existing theory

Equity and reciprocity may not have a meaningful impact on people's willingness to pay tax, even though theoretically these issues could matter. To relate this directly to our model in Equation 5 this would imply that  $r_i$  is indistinguishable from 1 (i.e., reciprocity does not influence whether people pay tax) and  $\delta_i$  is indistinguishable from 0 (i.e., the degree of progressivity in the tax and transfer system does not influence whether people pay tax). If this is the case, the implications for policymakers from this study are rather straightforward; they can change the degree of equity and reciprocity in the tax and transfer system without impacting people's willingness to pay tax.

#### 2. Issues with the study design

A failure to detect meaningful impacts from the treatments could be because of issues with the study design, however numerous efforts were made to ensure this is unlikely to occur. We have an adequately powered randomized survey experiment with treatments that have been piloted extensively. We adopt best practices from the existing literature and have a series of robustness checks allowing us to detect if design issues are driving the results. We capture people's priors so we can detect the extent to which the treatments are consistent or counter to these priors. Further, the information contained in the treatments has not been widely circulated to the general public in these countries, which means it is reasonable to expect the treatments to be new to people, even if they are consistent with their priors. We present the source of the information in a similar manner to seminal randomized survey experiments (e.g., Alesina et al., 2018) so that is more likely to be trusted by respondents. While we cannot rule out the possibility of implementation issues, we believe the study design is strong enough to minimize concerns ex ante about this being an issue.

### Treatment effects are broadly consistent with the hypotheses

In the instance whereby the results are broadly consistent with the hypotheses, we believe this study will have generated unparalleled, rigorous evidence about how equity and reciprocity shape people's willingness to pay tax. As discussed throughout this registered report, there is limited empirical evidence about how equity and reciprocity shape people's willingness to pay tax, particularly in low- and middle-income countries. By developing a conceptual framework that illustrates the channels through which equity

and reciprocity are likely to influence people's willingness to pay tax and providing evidence supporting it across a diverse set of countries we will make a major contribution to the literature.

The implications for policy makers largely depend on the order of magnitude of the treatment effects. As discussed in the introduction, reforms to taxes and/or transfers that intend to improve a country's fiscal position are likely to change the equity and reciprocity in the tax and transfer system. If we show that equity and reciprocity impact people's willingness to pay tax then it is necessary for policy makers to better understand people's responses to reforms, otherwise they could undermine their intended purpose of improving the fiscal position. If we detect large effects from the treatments on willingness to pay tax this would suggest that it is possible that reforms to taxes and/or transfers that intended to improve the fiscal position of a country could actually undermine tax compliance to a point whereby the net impact is negative.

#### Treatment effects exist, but they are largely inconsistent with the hypotheses

Ex ante it is challenging to identify all the reasons why some treatment effects may exist that are largely inconsistent with the hypotheses, however we identify three potential explanations.

1. Differences between the taxes and transfers treatments

It is conceivable that survey respondents may be more likely to respond to the taxes treatment than the government transfers treatment for a number of reasons. Firstly, on average, the share of household income collected in taxes is much higher than what is provided in transfers, which means people may be more concerned about how taxes are distributed compared to transfers. Secondly, there is reason to believe that "loss aversion" could exist where people's utility is more likely to be influenced by "losing" from paying tax than by "gaining" from receiving a transfer. Thirdly, people's awareness of when they pay tax may be higher than their awareness about when they receive a transfer. For example, people are likely to be more conscious of paying income tax compared to receiving a subsidy for their fuel consumption, and consequently this could make them more responsive to information about who pays taxes as opposed to who receives transfers.

2. Differences between the measures of willingness to pay tax

The impact of the treatments may vary considerably across the five outcome variables that are used to measure people willingness to pay tax. Forthcoming work by Prichard categorizes different measures of willingness to pay tax as being conditional or unconditional, depending on whether they are dependent on some attribute of the government (Prichard, forthcoming). In the case of Q17, this question is clearly an unconditional measure of people's willingness to pay tax. In contrast, Q18 includes caveats about when people should pay tax, which means this measure is conditional. In Prichard's work he shows these measures on conditional and unconditional willingness to pay tax can yield quite different results. Consequently, this could mean aggregating results through an index may mask important differences between specific outcome measures.

3. Differences between countries

This survey experiment will be conducted across a diverse set of countries with considerable variation in the tax and transfer systems, which in turn means the specific treatments vary across countries. It is challenging to ex ante identify how the treatment effects will differ across countries as this will potentially depend on a range of factors, such as the political economy within countries, the nature of their tax and transfer systems, the share of the population with internet access, etc. Consequently, it is foreseeable that efforts to pool the data across countries may need to be more focused. For example, it may be reasonable to only pool data across upper-middle income countries in the study or across countries with higher levels of tax revenue as a share of GDP.

## 6. Bibliography

- Afrobarometer, 2012, 2013, 2015. Available online at <https://afrobarometer.org/surveys-and-methods/questionnaires>
- Agerberg, M., and M. Tannenbergh, 2021. "Dealing with measurement error in list experiments: Choosing the right control list design". *Research & Politics*, April–June, 1–8.
- Alesina, A., and P. Giuliano, 2011. "Preferences for Redistribution." In *Handbook of Social Economics*, edited by J. Benhabib, A. Bisin, and M. Jackson, 93–131. Amsterdam: North-Holland.
- Alesina, A., S., Stantcheva, and E. Teso, 2018. "Intergenerational Mobility and Preferences for Redistribution." *American Economic Review* 108(2): 521–54.
- Alesina, A., A., Miano, and S. Stantcheva, 2021. "Immigration and Redistribution". NBER Working Paper 24733, June 2018, revised August 2021.
- Ali, M., O., Fjeldstad, and I. HoemSjursen, 2014. "To Pay or Not to Pay? Citizens' Attitudes Toward Taxation in Kenya, Tanzania, Uganda, and South Africa." *World Development*, 64, 828–842.
- Allingham, M., and A. Sandmo, 1972. "Income Tax Evasion: A Theoretical Analysis," *Journal of Public Economics*, 1(3-4), 323–338.
- Antinyan, A., and Z. Asatryan, 2019. "Nudging for Tax Compliance: A Meta-Analysis," ZEW Discussion Paper No. 19-055.
- Becker, G. 1968. "Crime and Punishment: An Economic Approach," *Journal of Political Economy*, 76(2), 169–217.
- Bodea, C., and A. LeBas. 2016. "The Origins of Voluntary Compliance: Attitudes toward Taxation in Urban Nigeria." *British Journal of Political Science*, 46(1), 215–238.
- Commitment to Equity (CEQ) Institute, 2021. *Commitment to Equity Handbook: Estimating the Impact of Fiscal Policy on Inequality and Poverty*. Available online at <https://commitmenttoequity.org/>
- Chuang, E., P. Dupas, E. Huillery and J. Seban, 2021. "Sex, lies, and measurement: Consistency tests for indirect response survey methods" *Journal of Development Economics*, Volume 148.
- Cruces, G., R., Perez-Truglia, and M. Tetaz, 2013. "Biased Perceptions of Income Distribution and Preferences for Redistribution: Evidence from a Survey Experiment." *Journal of Public Economics*, 98, 100–112.
- 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," Saïd Business School Working Paper 2019-07.
- Gimpelson, V., and D. Treisman, 2018. "Misperceiving Inequality." *Economics and Politics*, 30(1): 27–54.
- Hauser, O., and M. Norton. 2017. "(Mis)perceptions of Inequality." *Current Opinion in Psychology*, 18, 21–25.

- Haaland, I., C., Roth and J. Wohlfart. Forthcoming. "Designing Information Provision Experiments." *Journal of Economic Literature*.
- Hallsworth, M., 2014. "The use of Field Experiments to Increase Tax Compliance," *Oxford Review of Economic Policy*, 30(4), 658–679.
- Hoy, C., and F. Mager, 2021. "Why Are Relatively Poor People Not More Supportive of Redistribution? Evidence from a Randomized Survey Experiment across Ten Countries." *American Economic Journal: Economic Policy*, 13(4), 299–328.
- Hoy, C., L., McKenzie, and M. Sinning. 2021. "Improving Tax Compliance without Increasing Revenue : Evidence from Population-Wide Randomized Controlled Trials in Papua New Guinea." Policy Research Working Paper No. 9539. World Bank.
- IRIS, 2021. IRIS Network. Available online at <http://www.irisnetwork.org/>
- Karadja, M., J., Mollerstrom, and D. Seim, 2017. "Richer (and Holier) than Thou? The Effect of Relative Income Improvements on Demand for Redistribution." *Review of Economics and Statistics*, 99(2), 201–12.
- Khwaja, A., O., Haq, A., Khan, B., Olken, and M. Shaukat, 2020. "Rebuilding the Social Compact: Urban services delivery and property taxes in Pakistan." International Initiative for Impact Evaluation, Available online at: <https://www.3ieimpact.org/sites/default/files/2020-04/IE117-DPW1.1005-Property-taxes-Pakistan.pdf>
- Kuziemko, I., M., Norton, E., Saez, and S. Stantcheva, 2015. "How Elastic Are Preferences for Redistribution: Evidence from Randomized Survey Experiments." *American Economic Review*, 105(4), 1478–1508.
- Lee, D. 2009. "Training, Wages, and Sample Selection: Estimating Sharp Bounds on Treatment Effects." *Review of Economic Studies* 76 (3): 1071–1102.
- Lustig, N. (Ed.). 2018. "Commitment to Equity Handbook. Estimating The Impact of Fiscal Policy on Inequality and Poverty". Brookings Institution Press.
- Lustig, N., C., Mariotti and C., Sánchez-Páramo, 2020. "The redistributive impact of fiscal policy indicator: A new global standard for assessing government effectiveness in tackling inequality within the SDG framework." *World Bank Blogs*, June 11, 2020, Available online at <https://blogs.worldbank.org/opendata/redistributive-impact-fiscal-policy-indicator-new-global-standard-assessing-government>
- 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," International Centre for Tax and Development Working Paper 58.
- Norton, M., and D. Ariely, 2011. "Building a Better America – One Wealth Quintile at a Time." *Perspectives on Psychological Science* 6(9), 9–12. Prichard, W., S., Jibao and N., Orgeira. Forthcoming. "Sub-National Tax Reform and Tax Bargaining: A Quasi-Randomized Evaluation from Sierra Leone." ICTD Working Paper. International Centre for Tax and Development.

- Prichard, W. Forthcoming. "Unpacking 'Tax Morale': Distinguishing Between Conditional and Unconditional Views of Tax Compliance." ICTD Working Paper. International Centre for Tax and Development.
- Prichard, W., A., Custers, R., Dom, S. Davenport, and M. Roscitt, 2019. "Innovations in Tax Compliance: Conceptual Framework." Policy Research Working Paper No. 9032. World Bank.
- Shimeles, A., D., Gurara, and F. Woldeyes, 2017. "Taxman's Dilemma: Coercion or Persuasion? Evidence from a Randomized Field Experiment in Ethiopia," *American Economic Association: Papers and Proceedings*, 107(5), 420–424.
- Slemrod, J. 2019. "Tax Compliance and Enforcement," *Journal of Economic Literature*, 57(4), 904–954.
- Stantcheva, S. 2021. "Understanding Tax Policy: How do People Reason?" *The Quarterly Journal of Economics*, 136(4), 2309–2369.
- World Bank. 2021. "World Development Indicators." Washington, DC: World Bank. Available online at <https://databank.worldbank.org/source/world-development-indicators>
- World Values Survey (WVS). 2020. "Wave 7." Available online at <https://www.worldvaluessurvey.org/WVSDocumentationWV7.jsp>

## 7. Administrative information

**Funding:** This work was financed as part of the Innovations for Tax Compliance (ITC) program at the World Bank. The ITC program receives funding from the Bill & Melinda Gates Foundation and the World Bank Global Tax Program.

**Institutional Review Board (ethics approval):** Ethics approval for this work was processed through the Australian National University Human Research Ethics Committee (Protocol 2021/792).

**Declaration of interest:** None.

**Acknowledgments:** I am incredibly grateful for the support, input and guidance provided by Stephen Davenport throughout this study. I am also greatly appreciative of the comments provided on an earlier version of this paper by Matthew Collins, Roel Dom, Anna Custers, Chiara Bronchi, Matthew Wai Poi, Pierre Bachas, Mahvish Ifrah Shaukat, Mathieu Cloutier, Katy Bergstrom, William Dodds and Jonathan Karver.