Working Paper 91

Teach to Comply? Evidence from a Taxpayer Education Programme in Rwanda

Giulia Mascagni, Fabrizio Santoro and Denis Mukama

February 2019

BILL & MELINDA GATES foundation

IDS Institute of Development Studies

UKaid from the British people

ICTD Working Paper 91

Teach to Comply? Evidence from a Taxpayer Education Programme in Rwanda

Giulia Mascagni, Fabrizio Santoro and Denis Mukama

February 2019

Teach to Comply? Evidence from a Taxpayer Education Programme in Rwanda
Giulia Mascagni, Fabrizio Santoro and Denis Mukama
ICTD Working Paper 91
First published by the Institute of Development Studies in February 2019

ISBN: 978-1-78118-516-2

Available from:
The International Centre for Tax and Development at the Institute of Development Studies, Brighton BN1 9RE, UK
Tel: +44 (0) 1273 606261
Email: info@ictd.ac.uk
Web: www.ictd.ac/publication

IDS is a charitable company limited by guarantee and registered in England
Charity Registration Number 306371
Charitable Company Number 877338

Teach to Comply? Evidence from a Taxpayer Education Programme in Rwanda

Giulia Mascagni, Fabrizio Santoro and Denis Mukama

Summary

The role of taxpayer education in improving tax compliance has been largely unexplored in the literature. This paper starts to fill this gap by providing the first rigorous evaluation of the effectiveness of taxpayer education on knowledge, perceptions, and compliance, which took place in Rwanda. Our analysis is based on a unique dataset that combines administrative and survey data. We show that taxpayer education results in significant and large increases in knowledge, which starts from a very low level at baseline, and that it contributes to improving compliance behaviour. Our strongest result is that training new taxpayers helps bring them into the habit of filing tax declarations – an obligation many fail to comply with. In terms of policy, our results show that the benefits of taxpayer education go beyond increased revenue in the short term, and include building a habit of tax compliance.

Keywords: taxpayer education, knowledge, tax compliance.

Giulia Mascagni is a Research Fellow at the Institute of Development Studies (IDS) and Research Director of the International Centre for Tax and Development (ICTD).
Fabrizio Santoro is a Research Officer at the Institute of Development Studies (IDS) and PhD student in Economics at the University of Sussex.
Denis Mukama is the Head of Research and Policy Analysis at the Rwanda Revenue Authority (RRA).

Contents

Tables

Figure

Acknowledgments

This study was funded by the International Centre for Tax and Development, which is funded by the UK's Department for International Development (DFID) and the Bill & Melinda Gates Foundation. This study was carried out in close collaboration with the Rwanda Revenue Authority, to which we are very grateful for its continued support. More specifically, we would like to acknowledge the contributions of the following individuals, who have been directly involved in various aspects of the project. Mrs Drocella Mukashyaka, Deputy Commissioner for Taxpayer Service, was not only instrumental in conceptualising the research idea, but also agreed to accommodate the research into the training plan. Her inputs and guidance were a valuable addition. Mrs Angelique Dusabeyezu was the focal person for the research team within the Taxpayer Services department and coordinated the related activities. Mrs Jeanne Uwigoroye, Mr Gerald Abiyingoma and Mrs Delphine Uwimana helped in organising the trainings, carrying out taxpayer registrations, and conducting the lottery. They worked closely with the external research firm during the survey process to share details of the relevant taxpayers. Mrs Joseline Umuhire, Mrs Lilian Umuhoza, Mr John Karangwa and Mr Naphtal Hakizimana provided assistance in conducting the one-to-one phone calls to taxpayers during implementation of the coaching aspect of the research. They also participated in taxpayer registrations in some of the trainings. In addition, we are grateful to the high level of support we received from Mrs Agnes Kanyangeyo and Mr Pascal Ruganintwali.

Acronyms

CIT	corporate income tax
EITC	Earned Income Tax Credit
IV	instrumental variable
ITT	intention to treat
PIT	personal income tax
PSM	propensity score matching
RRA	Rwanda Revenue Authority
TIN	Taxpayer Identification Number
TOT	treatment on treated
TPS	Taxpayer Services [department]

1 Introduction

A recurring mantra in tax research is that information is key to tax administration (Pomeranz 2015; Kleven *et al.* 2011). This is particularly true because better information about taxpayers' activities and incomes can help enforcement, even though this potential can be severely limited in low- and middle-income countries where administrative capacity is relatively weak (Carrillo, Pomeranz and Singhal 2017; Mascagni, Mengistu and Wodleyes 2018). However, a much less studied aspect is how information affects taxpayers' ability to cope and comply with the tax system. Some studies have shown that taxpayers are often confused about complex tax systems. As a result, they respond to it in ways that are inconsistent with standard tax policy (Feldman, Katuščák and Kawano 2016), and in some cases they even end up paying more than they need to because they fail to take up provisions to their benefit (Benzarti 2015). This lack of information is consistent with a broader literature showing that people are often ill informed about tax and transfers that would affect their economic choices (Chetty, Looney and Kroft 2009; Liebman and Luttmer 2015). In African countries, low taxpayer knowledge is particularly severe, as the majority of Africans do not know what taxes they owe to the government or what tax payments are for (Isbell 2017; Aiko and Logan 2014). When we asked Rwandan taxpayers when the deadline for filing taxes is, in the context of this study, less than a quarter of our survey respondents indicated the correct month.

This paper starts to fill the evidence gap in this under-researched area by providing the first ever rigorous evaluation of the effectiveness of taxpayer education on knowledge, perceptions, and compliance in Rwanda.[1] We show that providing basic tax training significantly increases taxpayer knowledge and compliance behaviour. More specifically, it contributes to bringing them into the habit of filing tax declarations – an obligation many new taxpayers fail to comply with.

Weak tax knowledge has a number of potentially serious implications. First of all, it can certainly affect tax compliance, which is the focus of our analysis. Smaller, less educated, or lower-income taxpayers are more likely to feel confused about complex tax rules, thus potentially failing to comply (e.g. not declaring or not doing so on time) or not taking full advantage of tax laws (e.g. failing to claim deductible expenses). Thus, an increase in knowledge for these taxpayers could have ambiguous effects on tax payments – positive in the first case (i.e. failure to comply) and negative in the second one (i.e. limited take-up). Beyond compliance, a poor understanding of the tax system is also likely to have broader implications. Confused taxpayers would not know with confidence how much they should pay, thus potentially being more vulnerable to corrupt officials or to be coerced into making unofficial payments. They may also be more prone to seeing the tax system as unjust, either because of corruption or because they might misperceive the benefits of paying tax. A recent study has shown that providing beneficiaries with information about the eligibility and amount of a subsidy increases the amount of subsidy they receive by about 26 per cent, thanks to lower leakage (Banerjee *et al.* 2018). The recent evidence on informal taxation suggests that similar effects are likely to occur also in the area of tax (van den Boogaard, Prichard and Jibao 2018). Ultimately, uninformed taxpayers are less likely to engage in a meaningful debate with the government about tax issues, thus limiting the potential of taxation to act as a catalyst for improved governance and accountability.

Against this background, it seems particularly relevant, both for researchers and policymakers, to investigate what can be done to increase taxpayers' knowledge – with the associated benefits that this may have, starting from improved compliance. One obvious possibility for

[1] The only other study available on this topic, Chetty and Saez (2013), only looks at reported income.

governments is to provide information and training for taxpayers. Importantly, this possibility is fully in line with a modern approach to tax administration, where increased compliance is achieved through both traditional enforcement and more friendly measures aimed at encouraging taxpayers to comply voluntarily. A large and growing experimental literature has tested the impact of letters sent by revenue authorities to nudge taxpayers to comply (for a review, see Mascagni (2018)). Such messages have generally been found to be effective in increasing compliance, including in two cases where they were tested in low-income contexts (Shimeles, Gurara and Woldeyes 2017; Mascagni, Nell and Monkam 2017). In the case of Rwanda, reminders of deadlines proved to be particularly effective, compared to message contents focused on deterrence or fiscal exchange (Mascagni *et al.* 2017). A related study collected anecdotal evidence that taxpayers particularly appreciated being reminded of declaration deadlines (Mukama, Karangwa and Hakizimana 2017), which was consistent with low levels of knowledge of basic tax parameters.

However, these messages are mostly meant as nudges, not as attempts to increase knowledge in a more substantial way. A paper looking at the longer term effects of these messages confirms that they act as simple nudges, with very little or no learning at all in the longer term (Manoli and Turner 2014). In fact, revenue authorities do a lot more than sending letters to increase taxpayers' knowledge and sensitise them on the importance to pay tax. A recent review of tax education initiatives in Africa shows that many revenue authorities run trainings and seminars for taxpayers, often focusing on specific sectors or specific groups (e.g. new taxpayers), tax education programmes in schools, and radio or TV programmes, amongst others (Mascagni and Santoro 2018). Similar initiatives on taxpayer education are likely to occur in other countries as well, but they remain almost entirely undocumented. More importantly, there is virtually no evaluation of the effect of taxpayer education on knowledge and compliance. The only rigorous study we could find on this subject looked at the effect of a two-minute training for recipients of the Earned Income Tax Credit (EITC) in the United States, showing it had no effect on average (Chetty and Saez 2013).

This paper addresses this gap in the literature by evaluating the effectiveness of a taxpayer education programme on knowledge, perceptions, and tax compliance. To the best of our knowledge, this is the first rigorous and comprehensive study of taxpayer education to consider all these relevant outcomes in any country – as well as the first evaluation of taxpayer education in any low- or middle-income country. Our study departs from the literature on information nudges (e.g. letters) mostly because of the deeper and more comprehensive nature of our intervention, which explicitly aims to educate taxpayers on all basic aspects of the tax system. More specifically, we focus on a tax education programme run every year by the Rwanda Revenue Authority (RRA) for a relatively homogeneous group of taxpayers: those who recently registered to obtain a Taxpayer Identification Number (TIN, see section 2.1). We collected survey data on these new taxpayers before and after the training, and we also observe their compliance behaviour thanks to administrative data from tax returns (see section 2.2). Although attendance to the training could not be randomised, attendees and non-attendees are fully comparable before the training, across a rich set of observable dimensions including taxpayer characteristics (e.g. size, age, gender), baseline knowledge (measured with 19 indicators) and perceptions (e.g. on the tax system, state authority, and public services). We show that those who attend the training have much higher knowledge, better perceptions about the complexity of the tax system, and better compliance outcomes after taking the training – compared to those who did not attend. These effects hold once we control for a range of potentially relevant variables. We show that our results on compliance, particularly the positive effect on the probability to declare, are robust to concerns about identification and causality. We do so by using random assignment to our survey as an instrumental variable (IV) for attendance, as well as propensity score matching (PSM) on a wide range of observable characteristics (see section 3.4). We also explore whether the main

channel for the observed compliance effect is knowledge, against other possible explanations, and compare the RRA's education programme to a potential one-to-one coaching alternative (see sections 3.3.1 and 3.3.2, respectively).

Taken together, our results suggest that the education programme of interest is highly effective to increase taxpayer knowledge and compliance behaviour. On the latter, and in line with the only other study in the field (Chetty and Saez 2013), the programme does not seem to increase reported income – or at least this result does not stand up well to further scrutiny with the available data (see section 3.4). However, it does significantly increase the probability that new taxpayers file a tax declaration in the first year since registration. This is particularly relevant to the Rwandan context, as over half of all newly registered taxpayers completely fail to file a declaration in the first year.[2] The fact that the programme increases declaration rates is important because it builds a habit of declaring for those who have never been in the tax net before. A recent paper has shown that habit can be a powerful driver of compliance – a finding that is consistent to the separate, but related, literature on get-out-the-vote (Dunning *et al.* 2017). Just the act of regularly paying taxes, or not doing so, can significantly affect compliance behaviour in the future. If taxpayer education can push new taxpayers into that habit, as we show here, there may well be positive compliance gains down the line – in addition to the immediate benefits in terms of expanding the tax base and building a compliance culture. Therefore, even if policymakers should not necessarily expect immediate revenue gains from training new taxpayers, our results still represent a clear recommendation to invest more in taxpayer education programmes.

2 Research design and background

2.1 Taxpayer education in Rwanda

When it comes to tax, Rwanda is generally seen as a success story. The tax to GDP ratio stands at just over 15 per cent in 2015 (ATAF 2017), in line with other African countries. More importantly, the RRA is widely perceived as a performing revenue authority that has fully embraced a modern approach to tax administration, where encouraging voluntary compliance plays a key role in revenue mobilisation, along with traditional enforcement. Rwanda was one of the first countries to run a yearly National Taxpayer Appreciation Day, which normally entails events stretching well beyond a single day and is aimed to appreciate taxpayers' contributions to national development through their tax payments. Partly reflecting the recent history of the country's relations with foreign donors, the government has adopted a rhetoric of self-reliance that has domestic revenue mobilisation at its core. As we show in section 3.1, all this is reflected in exceptionally strong perceptions about the importance of taxpaying in contributing to national development.

Against this background, the RRA's Taxpayer Services (TPS) department organises a wide range of initiatives on taxpayer education including the National Taxpayer Appreciation Day, tax clubs in schools, and seminars for specific sectors or types of taxpayers.[3] Amongst these is the training programme for newly registered taxpayers, which is the focus of this study. This programme involves a half-day class led by TPS officials and covers the basics of taxpaying;

[2] This figure is 57 per cent for taxpayers registered in 2017 and 53 per cent for those registered in 2015, prior to our study.

[3] For more details on these initiatives, in the context of taxpayer education in Africa more broadly, see Mascagni and Santoro (2018).

for example, explanations of the various taxes and duties, including their deadlines, what a TIN is and why it is needed, procedures for tax declaration and payment, services available for simplifying declarations and payments, such as online services, amongst others. As such, the content is more focused on providing practical information on taxpaying, rather than taking a broader approach based on accountability and citizen engagement. We would therefore expect it to have some impact on compliance behaviour, not least because it is specifically targeted at increasing taxpayer knowledge in this respect. Programme sessions are organised throughout the country, endeavouring to reach all tax centres where there are substantial numbers of new taxpayers who might need the training. Typically, the first session of the year is organised in July/August, to capture taxpayers who registered from January onwards, and other sessions continue to be held until the end of March, which is the deadline for filing income tax declarations for the year ending in December. All sessions are the same in terms of content, duration, and trainers, who are all RRA officials. The programme schedule for 2017/18 is reported in Table A1 in the Appendix.

All participants are new taxpayers: they have typically registered for income taxes – personal income tax (PIT) or corporate income tax (CIT) – and obtained their TIN for the first time.[4] They are all from new businesses, either individual non-incorporated ones (PIT) or corporations (CIT), and not taxpayers who only have employment incomes. Regardless of the business type, this is a relatively homogeneous group of generally small firms (see section 3.1). Once they have registered, the RRA invites them to the first available session in their district. Invitations are normally sent via SMS or official letters posted at the office of the relevant tax centre or local branch of the Private Sector Federation. In addition, phone calls from RRA officials (from the TPS, typically) are made to remind taxpayers to attend the session. Although this invitation process is comprehensive in principle, practical difficulties and administrative constraints mean not all intended beneficiaries are reached.[5] Nonetheless, the explicit intention of the RRA is to invite all new taxpayers and, potentially, keep the training accessible to any other taxpayer even if they were not specifically invited. In practice, the vast majority of attendees are new taxpayers from the relevant district. However, this policy intention means that we could not randomise invitations or attendance.

This programme is the key initiative that the RRA currently takes to educate new taxpayers. In this evaluation, we also piloted a potential alternative and compared it with the existing programme. We randomly selected 275 taxpayers to take part in a one-to-one coaching programme (see section 3.3.2). This programme involved two phone calls from tax officials, in the three months prior to the filing deadline, offering to answer any question participant taxpayers may have on their declaration. Similar interventions have been tried in other countries, although they are often undocumented and not evaluated, but not in Rwanda. However, the feedback we collected from taxpayers on the existing training programme suggests that there might be appetite for such an option (see section 3.3.2). It therefore seemed relevant to pilot and evaluate it, for potential consideration in the future – more likely as a complementary measure to the existing programme, rather than an alternative.

2.2 Data

When we started working on this evaluation, we immediately faced a fundamental problem. There were no usable data on attendance to the trainings, so we would not know which

[4] A taxpayer may subsequently register also for VAT, with the same TIN number for the same business, and thus would not be included in the definition of 'new taxpayer' used here.

[5] Mukama *et al.* (2017) provide examples of these practical difficulties.

taxpayers attended and, as a result, had no way of connecting attendance to declaration outcomes. However, this problem was relatively easy to solve: in collaboration with the RRA, we simply developed a new registration procedure that would allow the collection of attendance data at the time of the training, in digitalised format. This attendance data was therefore easily made available for all trainings held in 2017/18 (see list in Table A1). Nonetheless, this problem is indicative of the basic obstacles that often prevent any evaluation of programmes run by governments in low-income countries – in addition to highlighting the importance of ensuring data is available for policy evaluation and monitoring.

In addition to this new source of attendance data, we use two other sources of information: a survey and administrative data from tax returns. We can connect these three sources of information (survey data, administrative data, and attendance) thanks to unique identifiers, so we have all the relevant information for all taxpayers in the survey. Most of our analysis focuses on this survey sample (see below). For taxpayers who were not surveyed, we can still rely on attendance information and administrative data, so in section 3.5 we test some of our results on the full population of new taxpayers.

Our survey includes two rounds: one before (i.e. baseline) and one after the relevant session that taxpayers were invited to (i.e. post intervention).[6] Attrition between the two rounds is about 18 per cent, but it is not correlated to any of our variables, as shown in Table A2 in the Appendix. Nonetheless, to make sure our results are not affected, we repeat all our analysis with inverse probability weights.[7] Our results remain largely consistent, confirming that attrition is not a critical threat in our case. Importantly, the first survey round is the only baseline information we have, since administrative data is only available after taxpayers have registered – and in our case they are all new. Due to budget constraints, we limited the sample to 1,000 taxpayers and conducted the survey by phone.[8] We focused on the first three trainings in the RRA's training plan (see Table A1): one in Kicukiro, which is a busy district of the capital Kigali, and the other two in the smaller towns of Musanze and Rubavu.[9] In total, over 2,500 taxpayers registered in those three districts by July 2017. To obtain our desired sample size, and aware of attrition in survey participation, we randomly selected 1,400 taxpayers that the survey company would attempt to collect data from, and discarded the remaining 1,100. The company then started to call those 1,400 taxpayers, in random order and until the desired sample size of 1,000 was reached. We use this random allocation to the survey sample as the basis of our IV strategy, which is discussed in more detail in section 3.4.

Our survey included six modules asking detailed questions on: (1) respondents' demographics (e.g. age, gender); (2) characteristics of the business (e.g. number of employees, income); (3) reasons to register for a TIN (e.g. obeying the law, access to finance); (4) attitudes and perceptions (e.g. on the complexity of the tax system, on tax as a social duty); (5) tax knowledge (19 indicators on various aspects of the tax system); and (6) information on the RRA's programme (e.g. in round one, intention to attend; in round two, feedback on the training or reasons for not attending).[10] Since the two rounds happened within a few weeks of each other, and given that the full survey took on average 42 minutes to complete, in the second round we only collected data on our main variables of interest: all those in the modules on knowledge, attitudes, and perceptions, and the feedback on training.

---

6 We obtained permission to conduct this survey from the National Institute of Statistics in Kigali, which also granted our Ethics Approval in July 2017.

7 Results available upon request.

8 Although doing the survey by phone may have limitations, the distance created with a phone call, as opposed to an interview in person, may have put the respondents in a more comfortable position to answer some sensitive questions about their attitudes towards compliance and perceptions about the tax system.

9 Both these towns are outside the Kigali area and include both rural and urban populations.

10 Our protocol also included oral consent to take part in the survey and information about confidentiality at the beginning of the phone interview.

Administrative data nicely complements the survey particularly for what concerns compliance behaviour. Obtaining honest answers about taxpayers' compliance is typically one of the major challenges in compliance research, especially in studies based on survey data. By using administrative data from tax returns, we can directly observe compliance behaviour, most importantly compliance with the obligation to declare, but we still cannot observe income misreporting. More specifically, we can see the amount declared, but we do not know if it is true or an under-reported value.[11] We use data from income tax declarations filed between January and March 2018, which is the first declaration period for the taxpayers considered here, as they are all newly registered. This also implies that we have no baseline administrative data for these new taxpayers, as they have never filed a declaration.[12] Furthermore, one of the usual drawbacks of administrative data is the limited range of variables available in tax returns, typically only sector and location (or a few others) in addition to financial variables related to income, expenses, and the tax base. By matching administrative data with our survey, we obtain a unique and rich dataset that includes firms' characteristics, perceptions and attitudes, as well as information on their real compliance behaviour.

2.3 Empirical strategy

Designing this research project was not straightforward as we had almost no literature to rely on (see section 1) and no information on attendance to the programme (see section 2.2) – let alone its effectiveness, in Rwanda or with regard to similar programmes elsewhere. It was also clear that we would not be able to randomise training sessions, or attendance, as the RRA wanted to keep the training open to all new taxpayers who might find it useful (see section 2.1). We therefore planned to start by collecting survey and attendance data, to get some initial information, and then proceed with an encouragement design in the form of a lottery. The lottery was organised for some sessions, to be compared with others with no lottery, and taxpayers who were invited to that session *and* attended would be eligible to win one of five monetary prizes.[13] This was meant to create an exogenous difference in exposure to the training that we could then use to evaluate its impact. In the course of running the research, however, we obtained two unexpected results. First, survey data revealed that those who attended the programme are not significantly different to those who did not attend, on any of the several dimensions we can observe (see section 3.1). Second, the lottery increased attendance only marginally in one district,[14] but was not effective on average. These two findings are somewhat consistent: they suggest that there is an element of randomness in the decision to attend this programme, which cannot be much influenced by, at least, monetary incentives. Indeed, when we asked non-attendees why they did not attend, the vast majority responded with rather ad hoc reasons (32 per cent were busy at work, 30 per cent were sick or had a sick family member to care for, 13 per cent forgot), while very few gave answers that would make us think attendance is related to unobserved factors (e.g. did not think it would be useful, or prefer to learn from other sources).

11 We could still infer compliance, in terms of income under-reporting, by comparing taxpayers who received a certain intervention to those who did not, for example a letter in the typical message RCT (see Mascagni *et al.* (2017)). In our case, we can compare attendees and non-attendees, building on the good baseline balance we observe and test.

12 At least, they have not with the TIN they recently obtained. We cannot observe whether they previously had another TIN, which could refer to a completely separate business.

13 We selected three districts, each with two sessions of which one served as a treatment (with lottery) and one as control (no lottery). Taxpayers in lottery sessions were informed in advance that the lottery would take place and that they would be eligible if they attended. Five monetary prizes were drawn at the end of the sessions, of the following amounts (net of applicable taxes): (1) 102,000 RWF; (2) 59,500 RWF; (3) 42,500 RWF; (4) 34,000 RWF; and (5) 25,500 RWF. Terms and Conditions of the lottery were made available to all participants at the beginning of the relevant sessions. More information is available from the authors.

14 The lottery's effect on attendance in this one district was very small in magnitude, and only borderline significant at the 10 per cent level.

Given this information, we decided to adopt a different empirical strategy. Our main results on the programme's effectiveness (sections 3.2 and 3.3) are based on the survey sample. Although the training was not randomised, in section 3.1 we present evidence to support our view that we still ended up with two groups – attendees and non-attendees – that are largely comparable. On the one hand, the advantage of focusing on the survey sample is that we can analyse the programme's effect on all three outcomes of interest: knowledge, perceptions, and actual compliance behaviour. We can also rely on a rich set of variables to check for balance and to use as controls in our regressions. In particular, in all regressions presented in sections 3.2 and 3.3 we control for: owners' characteristics (i.e. age, gender, level of education, whether the owner had a previous business, whether they previously had another tax training), characteristics of the business (i.e. size, months since registration, location, whether they use emails, whether they have a bank account, PIT/CIT), as well as knowledge level and perceptions at baseline.[15] Below is the equation we estimate, where Y is the outcome of interest after the training (knowledge, perceptions, or compliance – see sections 3.2 and 3.3), T is attendance to the training programme, and X is the set of control variables measured at baseline ($t-1$, about one week before the training).

$$Y_{it} = b_0 + b_1 T + b_2 X_{i(t-1)} + \epsilon_i \quad (1)$$

On the other hand, however, the survey was not meant to be representative of the population. And indeed it is not, particularly in two respects. First, it is over-representative of areas outside of Kigali, compared to the reference population of new taxpayers who registered in the relevant tax centres but did not take part in the survey.[16] We therefore use sampling weights throughout the analysis.[17] Second, attendance in the survey sample (with weights) is higher than in the broader population (see section 3.1). We argue that the main, or perhaps even the only, reason for this difference is that the first survey round happened a week before the training, and therefore served as a reminder. As mentioned in section 2.1, the RRA's usual invitation process is not expected to reach all intended beneficiaries of the programme, due to administrative constraints. The survey instead reached all surveyed participants, who therefore were certainly informed about the details of the relevant session. We do not view this as a major problem in our analysis, but we exploit this feature in our IV strategy (see section 3.4).

Since we are fully aware of the potential drawbacks of an evaluation that is not actually randomised, in section 3.4 we carefully address concerns on identification and causality, in two ways. First, we use random assignment to the survey as an exogenous instrument for attendance. As argued above, the survey acted as a reminder of the training, and indeed those who took part were more likely to attend (53 per cent for those who were actually surveyed and 39 per cent for those in the list of taxpayers to be surveyed, including non-responding taxpayers) than those who did not (14 per cent and 15 per cent).[18] This

15 To avoid losing observations due to missing values, we use the method suggested by Rubin (1987) that allows us to retain the maximum possible number of observations.

16 The share of taxpayers from the capital is 64 per cent for the reference population and 47 per cent for the survey sample. This is probably due to the fact that taxpayers in the capital are generally busier, therefore have less time to spend on the phone and might even be more wary to respond to phone calls from unknown numbers. The survey sample also includes relatively more PIT taxpayers, compared to CIT, but this difference is not statistically significant once we use weights to account for the population proportion of rural/urban taxpayers (using the normalised difference test from Imbens and Rubin (2015)).

17 We do not find statistically significant differences between the survey sample and the population of new registered taxpayers on the only other dimension that are observable at baseline, namely the registration month. As discussed in section 2.2, we do not have survey information for the broader population, or much administrative data as these new taxpayers have never filed a declaration.

18 This figure refers to the reference population for the survey: all taxpayers registered from January to July 2017 in the three districts that hosted the first three sessions of the programme (see section 2.2). Note that the coaching group is dropped.

difference is still present if we compare the two randomly drawn groups of taxpayers who were allocated to the survey, whether they were eventually interviewed or not (see section 2.2), and those who were left out: respectively 36 per cent and 15 per cent. Exploiting this difference in exposure, we can use random assignment to the survey to isolate the exogenous component of attendance and evaluate its impact on compliance outcomes in a more rigorous way.[19] Since we do not have survey data for those who are not in the survey sample, we are not able to adopt a similar strategy for outcomes related to knowledge and perceptions. Second, we use propensity score matching to address any concern that our comparison group may not be a suitable counterfactual. By doing this, we are able to approximate a randomised trial and therefore provide more rigorous causal estimates (see section 3.4). The results using those two strategies, shown in section 3.4, are largely consistent with our main analysis.

Finally, we present two additional elements of analysis that strengthen our understanding of taxpayer education in Rwanda. In section 3.3.1 we investigate whether increased knowledge is the main mechanism through which attendance to the training increases compliance. In section 3.5, we attempt to generalise our results to the whole population of new taxpayers in 2017/18, beyond the survey sample, looking at the relationship between programme attendance and declaration outcomes.

3 Results

3.1 Anatomy of taxpayers at baseline and programme attendance

Our survey participants are, as expected, a relatively homogeneous group of mostly small businesses. Table 1 reports baseline values of our covariates and outcome variables (knowledge and perceptions), from the survey. Over 90 per cent have less than five employees and many have no employees (besides the owner). A large proportion does not use emails (75 per cent), does not have a bank account (56 per cent), or does not keep proper books of accounts (53 per cent). Turning to the outcome variables of interest, it is striking to note that knowledge at baseline is very low. On average, taxpayers at baseline responded correctly only to about a third of the 19 basic questions we asked about Rwanda's tax system. No respondent gave the right answer to all questions. Interestingly, 37 per cent of the people we interviewed did not know what tax type they registered for. These figures clearly confirm the importance of organising taxpayer education trainings in Rwanda, as well as the presence of large margins for improvement in this area. Low knowledge of the basic parameters of the tax system is also consistent with the fact that most new taxpayers in Rwanda fail to make a declaration in the first year since registration. For example, amongst taxpayers registered in 2017, only about 43 per cent filed a declaration, while the majority failed to declare. These figures are consistent with those from previous years, showing that this is a persistent issue in Rwanda.[20] On the other hand, perceptions are generally very good at baseline. Virtually all taxpayers agreed with the tax attitude statements we provided, for example on the government's authority to make people pay tax (95 per cent), fairness of the tax system (98 per cent), and tax as a social duty (98 per cent). These figures are fully in line with the

[19] Using random survey assignment as an IV is a more conservative approach than choosing actual participation in the survey. Although the survey company proceeded to phone taxpayers randomly allocated to the survey, whether the taxpayer responded or not to the call may be related to unobservable factors.

[20] For example, amongst those registered in 2015, 53 per cent failed to declare. This figure is based on administrative data on new registrations from 2015, before this evaluation took place.

Rwandan context and the government rhetoric of self-reliance (see section 2.1). However, they suggest that we should not expect the programme to have much effect in increasing those attitudes further.

Table 1 also reports the average of all key variables, disaggregated by attendance status. The last two columns report t-tests and the normalised difference test (∆) to check whether any variable is significantly different between these two groups (Imbens and Rubin 2015). Although our evaluation was not set up as an experiment, the two groups we aim to compare (attendees and non-attendees) are fully comparable at baseline in relation to all outcome variables and most covariates.[21] Since we control for all covariates at baseline, we do not see the very few statistically significant differences reported in Table 1 as a threat to our empirical strategy. The fact that these two groups are fully comparable on a wide range of characteristics gives us some degree of confidence to use non-attendees as a counterfactual group for attendees. However, in section 3.4 we carefully test our results against possible empirical threats.

Table 1 Summary statistics and mean differences by attendance

	Baseline results			Non-attendees		Attendees			
	Mean	St. dev	Obs.	Mean	Obs.	Mean	Obs.	Difference	∆
Covariates									
Age	33.08	9.37	971	32.48	604	34.08	367	-1.61***	0.17
Female	0.32	0.47	971	0.31	604	0.34	367	-0.02	0.05
No school or primary only	0.19	0.40	971	0.21	604	0.17	367	0.03	-0.08
University degree	0.45	0.50	971	0.46	604	0.44	367	0.02	-0.04
Owner	0.78	0.42	971	0.76	604	0.80	367	-0.03	0.08
CIT	0.47	0.50	971	0.44	604	0.51	367	-0.07	0.13
>5 employees	0.09	0.29	971	0.08	604	0.11	367	-0.04**	0.13
Life in months	4.62	2.01	952	4.56	598	4.64	382	-0.08	0.04
Located in Kigali	0.49	0.50	951	0.52	569	0.43	382	0.09***	-0.17
Email use	0.25	0.44	971	0.25	604	0.26	367	-0.01	0.02
Bank account	0.44	0.50	971	0.42	604	0.48	367	-0.05	0.11
Books of account	0.47	0.50	971	0.46	604	0.50	367	-0.04	0.08
Had previous business	0.21	0.41	971	0.20	604	0.22	367	-0.02	0.04
Had previous training	0.07	0.25	971	0.07	604	0.06	367	0.01	-0.06
Knowledge and perceptions									
Baseline Knowledge Index	6.35	2.92	971	6.29	604	6.46	367	-0.17	0.06
Complexity	0.70	0.46	939	0.69	579	0.72	360	-0.03	0.06
Evasion attitude	0.33	0.47	924	0.33	572	0.34	352	-0.01	0.02
Fairness	0.98	0.14	932	0.98	578	0.98	354	0.00	-0.00
Government authority	0.95	0.21	950	0.95	591	0.96	359	-0.01	0.04
Social duty	0.98	0.14	969	0.98	602	0.99	367	-0.01	0.06
Public service satisfaction	0.00	1	948	-0.04	588	0.06	360	-0.09	0.09

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Observations are weighted by sampling weights. T–tests are computed on the *Difference* across the two groups. ∆ is the normalised difference from Imbens and Rubin (2015), equal to the difference in means scaled by the square root of the average of the two within-group variances. St. dev – standard deviation. Obs. – observations.

Table 2 shows attendance rates across various relevant samples. The first column refers to the survey sample (with weights), which is the basis of our main results. The second column shows the out-of-survey sample that, together with the survey sample, makes up the reference population for the survey (i.e. all new registrations from January to July in the relevant districts, as explained in section 2.2). Finally, the third column shows attendance rates for all sessions organised in 2017/18, for the universe of new taxpayers. All attendance figures are disaggregated by urban (i.e. Kigali, the capital) and more rural districts (although some areas outside of the capital are urban). The pattern across all samples is similar, in that rural areas tend to have higher attendance than the capital. However, as anticipated in section 2.3, the survey had a big impact in increasing attendance amongst those who took

[21] There are no normalised differences (∆) greater than 0.25, the benchmark Imbens and Rubin (2015) suggest when testing for balance.

part. The difference between the survey and out-of-survey samples is statistically significant and, we argue, due primarily to a reminder effect (see sections 2.3 and 3.4). Importantly, attendance is generally very low – with many taxpayers failing to take up the programme.

Table 2 Take-up of training programme by samples

	Survey sample[a]	Out-of-survey sample[b]	Whole sample
Attendance	45.8%	15.3%	17.6%
Urban	42.1%	13.4%	15.5%
Rural	49.1%	17.8%	23.7%
Observations	952	1,580	15,009

[a] *Survey sample* consists of 952 taxpayers for which baseline data and attendance data are available. This number is lower than the 1,000 who took part in the survey because for a small portion of surveyed taxpayers we could not connect survey data with attendance data. [b] *Out-of-survey sample* includes taxpayers who are never surveyed, either at the baseline or at the endline. *Urban* refers to tax centres in Kigali, *Rural* all the others.

3.2 Programme impacts on knowledge and perceptions

Having presented some evidence on the comparability of the two groups of attendees and non-attendees, we proceed to evaluate the effectiveness of the programme as discussed in section 2.3. Table 3 shows the programme's impact on taxpayer knowledge.[22] The first two columns show that attendance to the programme is associated with a 27 per cent probability that taxpayers will experience any increase in knowledge between the two survey rounds. This effect is highly significant and its magnitude does not change whether we include control variables (column two) or not (column one), which is consistent with the baseline balance shown in Table 1. Columns three and four show the percentage gain of the knowledge index, which gives equal weight to all of our 19 indicators. In this case, the dependent variable is the ratio of the difference in knowledge between the two rounds to baseline knowledge. Attendees experience on average a 56 per cent increase in knowledge after the training, compared to those who did not attend. Finally, we use the actual index, re-scaled from 1 to 10, as dependent variable – once again showing relatively large increases in knowledge for attendees, especially considering the very low baseline level of the re-scaled index (about 3).

Table A3 in the Appendix also provides some insights on heterogeneous effects across four dimensions of interest: gender, location in Kigali, level of education, and taxpayer size. It shows that taxpayers in Kigali tend to experience higher knowledge gains, while large taxpayers have smaller ones, consistent with the fact that they have more resources to navigate the tax system (e.g. dedicated accountants, consultants).

Turning to perceptions, Table 4 reports results on combined indexes that bring together statements from our perceptions module that refer to similar topics. Table A5 in the Appendix reports the statements used to build each one of these indexes. For example, the complexity index brings together the two following statements: 'It is difficult to file an income tax declaration' and 'It is difficult to get in touch with RRA officials to get information'. This binary index takes the value of 1 if the respondent agreed with at least one of these two statements. This is also the only one of all perception indexes that shows a statistically significant

[22] The number of observations in these regressions is lower than the 1,000 target for sample size (see section 2.2) because of attrition between the two survey sounds. Attrition is not related to any of our observed variables, so the two groups of attendees and non-attendees remain balanced.

Table 3 Training impact on tax knowledge

	(1) Increase	(2) Increase	(3) Gain	(4) Gain	(5) Index	(6) Index
Training	0.27***	0.27***	0.56***	0.56***	1.35***	1.26***
	(0.03)	(0.03)	(0.08)	(0.07)	(0.12)	(0.09)
Baseline knowledge		-0.09***		-0.41***		0.53***
		(0.01)		(0.04)		(0.03)
Covariates	No	Yes	No	Yes	No	Yes
Mean of Y	0.714	0.714	0.571	0.571	4.538	4.538
R^2	0.092	0.206	0.069	0.359	0.150	0.508
Observations	820	820	815	815	820	820

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors in parentheses. *Increase* is a dummy variable indicating whether the post-pre knowledge score difference is positive. *Gain* is the ratio of the difference over the 0-10 pre-training score. *Index* is the post training knowledge score rescaled over 0-10 range. Covariates used are those indicated in section 2.3. All regressions are weighted by sampling weights.

difference for attendees, after the training. Attendees are about 10 per cent less likely to agree with the complexity statements after having attended, which is consistent with the fact that they now have higher knowledge and better understand how to navigate the tax system. This result is confirmed when we look at individual indicators on attendance, instead of the combined index.[23] The programme does not seem to affect other perceptions, which perhaps should not come as a surprise given the very high levels already observed at baseline (see section 3.1).

Table 4 Impacts of training on perceptions indexes

	(1) Complexity	(2) Evasion	(3) Fairness	(4) Govt. authority	(5) Social duty
Training	-0.093**	-0.005	0.003	-0.008	-0.004
	(0.039)	(0.033)	(0.005)	(0.010)	(0.002)
Baseline	0.320***	0.199***	0.018**	0.067***	0.010
	(0.044)	(0.034)	(0.008)	(0.019)	(0.006)
Covariates	Yes	Yes	Yes	Yes	Yes
Mean of Y	0.693	0.316	0.940	0.933	0.979
Observations	815	811	810	817	820

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors in parentheses. Coefficients are marginal effects from probit regressions computed at the mean of the explanatory variables. The control group are taxpayers who did not attend the training. Perceptions outcomes are indexes which range from 0 (disagree) to 1 (agree). Covariates used are those indicated in section 2.3. All regressions are weighted by sampling weights.

3.3 Programme impacts on compliance

Having shown in section 3.2 that the programme had large and significant effects on knowledge and, to a lesser extent, on perceptions around complexity, we now ask whether these benefits translate into better compliance outcomes. Compliance is a broad concept that encompasses various specific aspects, including registration, filing a declaration, doing so in time, reporting the correct amount of tax base, and paying in full and on time. It is therefore important to consider multiple outcomes in the analysis. Here we focus on three aspects that are particularly relevant to new taxpayers: the probability to declare, the probability to nilfile, and the amount of tax declared.[24] The columns of Table 5 correspond to these three outcomes of interest.[25]

[23] The results are available from the authors upon request.

[24] We also checked if the programme had any effect on the probability of making a pre-payment following the declaration. Pre-payments are payments on account, based on the previous declaration's tax amount and divided in four payments of equal

Table 5 Impact of training and coaching on tax outcomes

	(1) Declare	(2) Declare	(3) Nilfile	(4) Nilfile	(5) Log tax	(6) Log tax
Training	0.140***	0.094**	-0.097*	-0.158***	1.205*	1.146***
	(0.036)	(0.041)	(0.056)	(0.060)	(0.623)	(0.408)
Coaching	0.068	0.099*	-0.068	-0.229**	0.766	1.547**
	(0.050)	(0.056)	(0.080)	(0.091)	(0.916)	(0.772)
Covariates	No	Yes	No	Yes	No	Yes
Knowledge/perceptions	No	Yes	No	Yes	No	Yes
Observations	971	969	393	393	364	364

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors in parentheses. Coefficients are marginal effects from probit (1-4) and tobit (5-6) regressions evaluated at mean. The control group comprises taxpayers who did not receive any educational input. We control for taxpayers who received a call under the coaching option (see section 3.3.2). *Declare* and *Nilfile* are dummy variables with value 1 if the taxpayers declared and nilfile, respectively. *Log tax* is the log(tax+1) transformation of the raw tax amount variable. Covariates and Knowledge and Perceptions used are those indicated in section 2.3. All regressions weighted by sampling weights.

The training seems to have a relatively large and statistically significant effect on the probability to make a declaration, which increases by about 10 percentage points once controls are included (column two). Although we do not have a baseline value for this sample (this is the first time they are required to declare, as explained in sections 2.1 and 2.2), we know from previous years' administrative data that over half of new taxpayers typically fail to make a tax declaration. For our comparison group, the declaration rate is 34 per cent.[26] Compared to this number, an increase of ten percentage points corresponds to a 30 per cent increase in the probability to declare, which we attribute to the programme (see section 3.4).

As mentioned in section 1, bringing taxpayers into the habit of paying taxes has been shown to be crucial to their future compliance (Dunning *et al.* 2017) and is therefore an important policy impact. Data from the RRA's database on taxpayers who registered in 2015 is also highly suggestive that habit plays an important role in determining their probability to declare. A new taxpayer who files a declaration in the first year of operation (2015) had 55 per cent probability to make a tax declaration in 2016 and an 86 per cent probability to file again in 2017, conditional on having filed the previous year. In comparison, taxpayers who fail to declare in the first year (2015) have a negligible probability (less than 1 per cent) of filing a declaration in any of the following two years.

Columns three and four of Table 5 look at the probability of nil-filing. This is a common phenomenon in low-income countries,[27] which essentially involves filing a declaration with nil amounts of income and, as a result, zero tax liability. While this could be the reflection of poor business performance, especially for new taxpayers, a separate study on this topic suggests that it is also partly related to non-compliance.[28] The programme seems to reduce the probability of nil-filing by about 15 per cent, once controls are included.

amount that need to be done in advance throughout the year, towards the following year's declaration. We find no effect of the programme on the probability of making a pre-payment. This is probably due to the fact that pre-payment information is only available for those who made a positive tax declaration, which is a small sample, as others would not be required to make a pre-payment. These results are available from the authors upon request.

25 Note that the sample size in Table 5 is larger, because we do not have a problem with attrition. For example, we can observe the post-training outcome of 'declare' for all 1,000 survey participants, while we can observe 'nilfile' only for those who declared and 'log tax' only for those with a positive tax amount.

26 The declaration rate in our survey sample is lower than in the whole population because the survey included proportionally more taxpayers in rural areas, even once we include weights, and they have a lower probability to declare.

27 For example, see Almunia *et al.* (2017); Mascagni and Mengistu (2016); Mascagni, Monkam and Nell (2016).

28 This study is not yet published.

Next, we check whether attendance to the programme also affects the amount of tax paid. As mentioned in section 1, in principle taxpayer education could either increase or decrease the amount of tax paid. Columns five and six of Table 5 suggest that the programme had a large positive impact on the log of tax payable, essentially doubling it for taxpayers who attended. In section 3.4 we put this finding, as well as others, to further scrutiny and test its robustness to empirical threats.

Finally, as in section 3.2, we check for heterogeneous effects across the same variables. Table A6 in the Appendix shows that women and large firms seem to experience larger increases in declaration rates, while there do not seem to be significant heterogeneous effects on nil-filing and the tax amount, possibly because of a smaller sample size.

3.3.1 Is knowledge the channel?

Having looked into programme impacts on knowledge and compliance, a question naturally arises: is knowledge the main channel for better compliance outcomes? To address this question, we use attendance to the programme as an instrumental variable for knowledge increases. The purpose here is to isolate the effect of knowledge on tax outcomes that is due to attendance. Although one could rightfully argue that attendance is not an exogenous variable, since it was not randomised, we are somewhat encouraged to proceed based on the balance tests reported in section 3.1 and the further scrutiny of section 3.4. We would also argue that attendance is more exogenous than knowledge increases, which are more likely to be related to unobservable factors. The first stage regressions show that attendance is indeed a relevant instrument as it displays a significant coefficient.[29] Although in principle attendance might also affect compliance outcomes through other channels, in our case the exclusion restriction is also satisfied. When included in a regression similar to the second stage one, our IV (attendance) is not statistically significant.[30] Somewhat supporting this result, in section 3.2 we have shown that attendance generally does not affect perceptions, except for those on the complexity of the tax system that are more closely related to knowledge. These results are suggestive that attendance affects compliance only or primarily through knowledge, rather than other channels, although of course this argument remains debatable.

With these important caveats in mind, Table 6 provides suggestive evidence that increases in knowledge may indeed be an important channel for the programme's impact on declaration rates. As far as other compliance outcomes are concerned, Table A7 in the Appendix shows that the coefficients on nil-filing or the tax amount have the right sign and a comparable magnitude to those reported in Table 5 but they are not statistically significant, possibly because of larger standard errors in the IV estimation. While these results are to be taken with caution, they are somewhat suggestive that increased knowledge may play an important role as a channel to better compliance outcomes. An alternative potential channel would be, for example, related to enforcement and deterrence perceptions that might be affected by the programme, since it is run by RRA staff. However, deterrence is unlikely to be a channel since the programme did not seem to affect perceptions on deterrence.

[29] Results from the first stage regressions are omitted for simplicity. The coefficient on our instrument is 0.12 when using knowledge increase and 0.26 when using knowledge gain, and they are both significant at the 1 per cent level.

[30] In such auxiliary regression, the p-value of the IV (attendance) is 0.160 for knowledge increase and 0.250 for knowledge gain.

Table 6 Impacts of knowledge on declaration probability – IV strategy

	(1) Increase	(2) Increase	(3) Gain	(4) Gain
Knowledge	0.50***	0.23	0.24***	0.11
	(0.17)	(0.14)	(0.08)	(0.07)
Covariates	No	Yes	No	Yes
Knowledge/perceptions	No	Yes	No	Yes
Mean of Y	0.411	0.411	0.413	0.413
Observations	700	700	696	696

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors in parentheses. The instrument is attendance to the training. *Increase* is a dummy variable indicating whether the post-pre knowledge score difference is positive. *Gain* is the ratio of the difference over the 0-10 pre-training score. Exclusion restric- tion holds: p-value of the IV in a regression with Knowledge is 16 per cent for Increase and 25 pr cent for Gain. Covariates used are those indicated in section 2.3. All regressions weighted by sampling weights.

3.3.2 A comparison with an alternative one-to-one coaching option

As mentioned briefly in section 2.1, we compare the main RRA education programme with a possible alternative: the one-to-one coaching programme. The motivation for piloting this potential alternative is twofold. First, when we collected feedback on the programme, in the second round of our survey, many taxpayers signalled that they would benefit from some form of complement to the training programme. About 40 per cent of attendees mentioned that the session was too short, and that there was not sufficient time for questions. Second, related studies have shown that a more intensive and personalised approach might be beneficial. In the related financial education literature, personalised counselling is shown to be a useful complement to more standard financial education, to increase real financial outcomes (Carpena *et al.* 2015). The very scarce literature on this aspect of tax does not seem to be fully conclusive and it only concerns high-income countries (Chetty and Saez 2013; Gangl *et al.* 2014).

Unlike the RRA's main taxpayer education programme, we could randomise the coaching option, as this was a small pilot of a new initiative. We randomly selected 275 taxpayers, using strata (gender, location in/out of Kigali, size, and baseline knowledge) from the group of non-attendees, so that they are fully comparable with the relevant control group (i.e. non-attendees who were not selected for the coaching option) – as confirmed by balance test.[31] RRA staff called the selected taxpayers during the declaration period, after the second round of the survey was carried out, so we only have post-treatment outcomes from administrative data, while we still have all the baseline survey data. Capacity constraints played a crucial role here, as this pilot was quite burdensome for the tax administration, even if it only involved a relatively small number of taxpayers. As a result, RRA officials could reach only 160 taxpayers in both rounds. The low compliance with the treatment and lacking records are mostly due to capacity constraints on the RRA's side, as the staff allocated to this task were few and spread thinly across a number of other duties. To take this issue into account, Table 7 reports results on the coaching effect on all compliance outcomes considered in section 3.3, using both ITT (intention to treat) and TOT (treatment on treated). The latter uses the random allocation to the treatment group as an IV for treatment status (i.e. the 160 that were effectively called in both rounds). Our results show that the coaching option does not seem to be very effective in improving compliance outcomes, especially compared to the main

[31] Balance tests are shown in Table A4 in the Appendix. In addition, based on the evidence of section 3.1, we would argue that the coaching treatment and control groups are also comparable with attendees, although we do not directly compare these groups.

programme's effects (see Table 5).[32] This is potentially due to low compliance with the treatment, both because of the limited number of taxpayers who were actually reached (though this is taken into account with TOT) and the fact that some of those reached did not have any questions or only spent a very short time on the call. For example, 43 per cent of taxpayers contacted in the first round did not have any tax-related questions, while as many as 68 per cent did not have any in the second round.

Although we received anecdotal reports that taxpayers appreciated the coaching phone call, our results do not suggest that this pilot was successful. Besides our estimation results, administrative constraints are an important consideration in terms of the feasibility of a potential scale-up of this option. With current staffing, reaching 160 taxpayers represented a relatively large burden on the departments involved so it makes any further scale-up particularly unadvisable. The RRA could still adopt this more personalised approach in specific and selected cases as a complement to the main education programme, which seems to be more effective and allows reaching out to a much larger pool of beneficiaries.

Table 7 Impacts of coaching on tax outcomes

	(1) ITT declare	(2) TOT declare	(3) ITT nilfile	(4) TOT nilfile	(5) ITT log tax	(6) TOT log tax
Coaching	0.02	0.03	-0.13*	-0.13	0.36	1.09
	(0.05)	(0.06)	(0.07)	(0.08)	(0.70)	(0.93)
Covariates	Yes	Yes	Yes	Yes	Yes	Yes
R^2		0.312		0.352		0.283
Observations	618	618	236	236	216	216

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors are in parentheses. Coaching group is compared to control group, i.e. taxpayers not attending. Training group is dropped. Covariates used are those indicated in section 2.3. All regressions weighted by sampling weights. ITT – intention to treat; TOT – treatment on treated.

3.4 Addressing endogeneity

Although we have argued that our groups are comparable and balanced (see section 3.1), we are fully aware of the limitations of a design that is not randomised. While we are confident that there are no significant differences between attendees and non-attendees in the wide range of variables we can observe, including perceptions and knowledge, one could still argue that there are other variables that are unobserved and might drive differences between those groups, or that they are not fully comparable. To address these concerns, we adopt two strategies: an instrumental variable (IV) for attendance and propensity score matching (PSM).

As far as the IV strategy is concerned, we exploit the fact that allocation to the survey sample was randomised (see section 2.2) and that the survey itself resulted in an increase in attendance (see sections 2.3 and 3.1). We argue the survey resulted in higher attendance simply because of a reminder effect, as there was no other incentive inherent in our questionnaire. Importantly, the survey was administered by an independent company, and not by RRA employees. In addition, the fact that our monetary incentive, through the lottery described in section 2.3, was not effective in increasing attendance shows that participation to the programme is not easily manipulated, beyond a simple reminder effect. Since here we rely on a larger number of observations, including taxpayers who were not surveyed, we can only

[32] Note that in Table 7 we have dropped the group of attendees, which is why the the number of observations is smaller than in Table 5. However, the results are still not statistically significant if we included them in the regression, while controlling for their attendance status.

use the IV strategy to explore the programme's effect on compliance outcomes.[33] In line with the exclusion restriction, we show that our instrumental variable (random assignment to the survey sample) does not directly affect our outcomes of interest (see Table A8 in the Appendix). However, the first stage regression, reported in column one of Table 8, confirms that it significantly increases the probability to attend the programme, by a substantial 22 percentage points once we control for our full set of covariates. Columns two, three and four report the second stage regressions, where attendance is instrumented and the dependent variables are our three compliance outcomes. The results on the probability to declare remain statistically significant and increase in magnitude, suggesting that, if anything, there might have been a downward bias in the main results. Results on the other two outcomes (nil-filing and log tax) lose their significance. This may be due to the larger standard errors that are inherent to the two-stage IV estimation, coupled with a lower number of observations available for these outcomes (i.e. only those taxpayers who filed a declaration) – or it could be a genuine reflection of ineffectiveness of the programme on these outcomes, once a more rigorous estimation method is used.

Table 8 Impacts of random allocation to survey and invitation – IV strategy

	(1) 1st stage	(2) Declare	(3) Nilfile	(4) Log tax
IV Assign. survey	0.22***			
	(0.02)			
Training		0.18**	0.09	-0.53
		(0.09)	(0.15)	(1.68)
Covariates	Yes	Yes	Yes	Yes
R^2	0.09	0.22	0.03	0.01
F-stat		137.85	46.57	42.57
Observations	2,378	2,378	907	860

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors are in parentheses. *IV Assign. survey* is the instrument, referring to a dummy for whether the taxpayer is randomly assigned to the *attempt to survey* group. Coaching group has been dropped. Covariates used are those indicated in section 2.3.

Secondly, we use a PSM estimator, where the propensity score represents the probability of being assigned to a treatment (i.e. attendance to the programme), given a set of covariates. There are two assumptions for PSM to be valid. The first one is unconfoundedness, which requires that all confounding factors be included in the set of covariates used to calculate the propensity score. We cannot directly test this assumption, but we use a very large set of covariates that capture the most important potential sources of concern, including attitudes and knowledge at baseline, as well as taxpayer and business characteristics.[34] The second assumption is overlap (or common support), which requires that individuals with the same score can be either attendees or non-attendees, as if a randomised experiment was carried out. As shown in Figure A1 in the Appendix, this assumption holds in our case.[35] Importantly, since with PSM we are only focusing on the survey sample, we can report results on all outcomes of interest: knowledge, perceptions (both reported in Table 9), and compliance outcomes (Table 10). All tables report the programme effects using three different matching algorithms, which also serves as a further test for robustness.[36] The tables also report two

[33] We have post-treatment information on knowledge and perceptions only for survey participants.

[34] The variables used to calculate the propensity score are: the set of covariates as indicated in section 2.3, as well as the baseline values of knowledge and perceptions indexes. To avoid reducing the sample due to missing values, we use the method suggested by Rubin (1987).

[35] The average propensity score of non-attendees is is 46 per cent while for attendees it is 52 per cent, i.e. very close.

[36] These are: (1) Nearest-neighbour matching with replacement, where the taxpayer from the non-attendees group is chosen as a matching partner for a trained taxpayer that is closest in terms of the propensity score. With replacement, an untrained

measures of the quality of matching, which confirm its validity.[37] In line with the relevant literature, we use the common support option and bootstrap standard errors.[38]

Table 9 Training effect on knowledge and perceptions – PSM

	Increase	Gain	Index	Complexity
Panel A: N-N matching with replacement				
Training	0.28***	0.62***	1.21***	-0.05
	(0.05)	(0.08)	(0.18)	(0.05)
Common support	818	813	818	813
Ps-R^2 matched	0.032	0.031	0.032	0.031
LR chi^2 matched	34.98	34.54	34.98	34.06
Panel B: Radius matching with caliper 5%				
Training	0.25***	0.57***	1.25***	-0.10***
	(0.03)	(0.07)	(0.12)	(0.04)
Common support	818	813	818	813
Ps-R^2 matched	0.032	0.002	0.002	0.002
LR chi^2 matched	1.76	1.91	1.76	1.91
Panel C: Kernel matching				
Training	0.25***	0.57***	1.25***	-0.10***
	(0.03)	(0.07)	(0.12)	(0.04)
Common support	818	813	818	813
Ps-R^2 matched	0.002	0.002	0.002	0.002
LR chi^2 matched	1.74	1.87	1.74	1.87

$^{*}\ p < 0.10$, $^{**}\ p < 0.05$, $^{***}\ p < 0.01$. Standard errors are in parentheses are from 99 bootstrapped repetitions. Note that the Ps-R^2 unmatched is always 0.042 and the LR chi^2 unmatched is always 47.40***. *Increase* is a dummy variable indicating whether the post-pre knowledge score difference is positive. *Gain* is the ratio of the difference over the 0-10 pre-training score. *Index* is the post-training knowledge score rescaled over 0-10 range. *Complexity* is the average of two complexity costs indexes.

The results of the PSM estimation are largely comparable, and in some cases extremely similar, to the main ones reported in sections 3.2 and 3.3. The magnitude and significance of the programme's effects on knowledge and perceptions remain essentially unchanged. The same is true for the declaration probability and nil-filing, which are still statistically significant albeit with a slightly smaller coefficient. The effect on the log tax amount instead loses its significance, although the magnitude and sign of the coefficient are still in line with our main result. The fact that the results from the PSM estimation are very similar to our main analysis supports its robustness. Importantly, it also confirms that the groups we were comparing before matching are very similar, as we argued in section 3.1.

Taken together, the results from the IV and PSM estimation show that our strongest result is the programme's positive impact on declaration rates, which stands up particularly well to further scrutiny. The PSM results also support our results on knowledge, perceptions on

taxpayer is used more than once as a match; (2) Radius caliper matching with a radius of 0.5, where applying caliper matching means that a taxpayer from the untrained group is chosen as a matching partner for a treated taxpayer that lies within a given distance of the propensity score (i.e. caliper) and thus is closest in terms of propensity score; and (3) Epanechnikov kernel matching with bandwidth 0.06. Kernel matching (and local linear matching) are non-parametric matching estimators that use weighted averages of (nearly) all individuals in the control group to construct the counterfactual outcome. How many individuals are chosen from the control group depends on the kernel function. Weights depend on the distance between each individual from the control group and the participant observation for which the counterfactual is estimated.

37 As explained in Sianesi (2004), after matching there should be no additional systematic differences in the distribution of covariates between the two groups and therefore the pseudo-R^2 should be fairly low. The likelihood ratio test should be rejected in all cases when matching is applied, as is the case in all panels.

38 The common support option requires that we only include the subset of observations in the comparison group that is comparable to the treatment group (Dehejia and Wahba 1999). Heckman, Ichimura and Todd (1997) stress that a violation of the common support condition is a major source of evaluation bias.

Table 10 Impact of training on tax outcomes – PSM

	Declare	Nilfile	Log tax
Panel A: N-N matching with replacement			
Training	0.13***	-0.15*	1.09
	(0.05)	(0.07)	(0.92)
Common support	807	324	303
Ps-R^2 matched	0.022	0.099	0.105
LR chi^2 matched	26.07	54.84***	55.19***
Panel B: Radius matching with caliper 5%			
Training	0.08**	-0.12**	0.96
	(0.03)	(0.05)	(0.70)
Common support	807	324	303
Ps-R^2 matched	0.002	0.047	0.056
LR chi^2 matched	2.36	26.02	29.37
Panel C: Kernel matching			
Training	0.08**	-0.12**	0.94
	(0.03)	(0.05)	(0.71)
Common support	807	324	303
Ps-R^2 matched	0.002	0.046	0.056
LR chi^2 matched	2.14	25.78	29.18

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors in parentheses are from 99 bootstrapped repetitions. Note that the Ps-R^2 unmatched is always 0.11 and the LR chi^2 unmatched is always 44.45**. Coaching group is dropped from the analysis. The control group consists of taxpayers who did not receive any educational input. *Declare* and *Nilfile* are dummy variables with value 1 if the taxpayers declared and nilfile, respectively. *Log tax* is the log(tax+1) transformation of the raw tax amount variable.

complexity, and nil-filing. The result on log tax instead does not hold in the IV or PSM estimation, which might be due to bias in our main analysis or to the smaller sample size we can count on for this outcome.

3.5 Population-wide effects of the programme

Our results so far have shown large increases in knowledge and in declaration rates that are robust to potential empirical threats – while the results on nil-filing and tax amount do not stand up as well to further scrutiny (see section 3.4). Therefore in this section, we focus on our strongest compliance result, on the declaration probability, and address the issue of external validity, at least for what concerns the population of taxpayers in Rwanda beyond our survey sample and its reference population.[39] In other words, do our results hold when we look at the whole population of new taxpayers in Rwanda? For those registered in 2017, we can observe both attendance status and whether they declared or not, from administrative data (see section 2.2). We can therefore link the two and analyse the programmes' relation with the probability to declare, for the universe of new taxpayers. However, we have a very limited set of covariates for the whole population – essentially, only those available in the registrations database.[40] These 'administrative covariates' are: whether the taxpayer registered as a corporation (CIT) or as an unincorporated business (PIT), the months since registration, the months since the training, and fixed effects for the tax centre where they registered and where they were invited to attend the programme.

Based on these data, Table 11 reports the programme's effect as estimated once we control for the few covariates available. The first column of Table 11 shows that the coefficient of the training in the survey sample, controlling for the smaller set of administrative covariates, is very similar to the one reported in Table 5 (column two), where we control for a much larger set of covariates from the survey. This gives us some confidence that omitting those covariates might not result in a serious bias in our estimates. Comparing the programme's coefficient in the survey sample (column one) and in the population (column two), clearly shows that the effect size is much larger in the latter. This difference might be due to a different composition in terms of taxpayers based in Kigali as opposed to those based in other regions, outside of the capital (we call the latter 'rural', even though some of these areas have urban centres). In particular, the survey, as well as its reference population, is more representative of 'rural' areas than the whole population, where Kigali represents a much larger share of new taxpayers.[41] To improve comparability, we separate both the survey sample and the population into rural (columns three and four) and urban areas (columns five and six). Even using those more comparable sub-groups, the effect size in the survey sample are smaller than in the broader population, although the direction of the effect is the same. The programme still increases the probability to declare, but this effect seems to be larger in the broader population than in the survey sample – both in rural and in urban areas. Based on this comparison, we might consider the effects identified in our main analysis (section 3.3) as a lower bound for the more general effect of the programme. Table 11 suggests that the programme increases declaration rates by 30 per cent in the broader population, as opposed to about 11 per cent in the survey sample.

---

39 As mentioned in section 2.2, the reference population for our survey is composed of taxpayers who registered from January to July in the three tax centres included in the survey. In contrast, in this section we look at the total population of all taxpayers registered in any tax centre at any time in 2017.

40 Note that, since many new taxpayers fail to file, we do not have variables from the tax declaration for all relevant taxpayers here, although we can observe the outcome of interest for everyone – whether they file ornot.

41 The survey's reference population is new taxpayers registered from January to July in the three relevant tax centres (see section 2.2). The sampling weights that we describe in section 2.3 simply bring the survey sample in line with the reference population, which however is still more representative of 'rural' areas than the whole population.

Table 11 Impact of training on probability to declare

	(1) Survey	(2) All	(3) Survey rural	(4) All rural	(5) Survey urban	(6) All urban
Training	0.110***	0.308***	0.077	0.332***	0.134**	0.277***
	(0.040)	(0.012)	(0.052)	(0.020)	(0.054)	(0.015)
Admin. covariates	Yes	Yes	Yes	Yes	Yes	Yes
Tax centre FEs	Yes	Yes	Yes	Yes	Yes	Yes
Observations	971	14,994	508	4,051	461	10,942

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors are in parentheses.
Coefficients are marginal effects from probit regression computed at the mean of covariates.
All regressions are weighted by sampling weights. FEs – fixed effects.

Although the results on nil-filing and tax amount are weaker, we also check the programme's relation with those other compliance outcomes in the broader population – since for those too we have all the necessary data, although with the same smaller set of administrative covariates mentioned above. Tables A9 and A10 in the Appendix report, respectively, comparable results to Table 11 but for nil-filing and tax amount. As for the declaration probability, the results on the other two compliance outcomes are very similar once we include administrative covariates instead of the survey covariates. Perhaps the most interesting result here is on the tax amount, which seems to significantly increase for attendees in the broader population. Although the results on tax amount were not significant in the IV and PSM estimation (section 3.4), they also suffered the most from a small sample size. A larger sample size might allow for more precise and statistically significant estimates, in the presence of actual effects. The fact that, once we use a larger sample, attendance to the programme is associated with higher tax amounts, combined with the apparent absence of self-selection (see section 3.1), suggests that there might be an effect on the level of tax payable, in addition to declaration rates. However, these results need to be taken with caution as we cannot test their robustness further.

4 Conclusions

This paper aims to shed light on the effectiveness of taxpayer education, a topic that has so far been largely ignored in the literature. More specifically, we evaluate a taxpayer education programme run by the Rwanda Revenue Authority and offered to all new taxpayers. We are particularly interested to explore the programme's effects on knowledge, perceptions, and compliance outcomes. To investigate this question, we rely on a unique dataset that matches information from a survey and from administrative tax records. By doing this, we can overcome some of the common problems both of surveys (e.g. difficulty to measure compliance) and of administrative data (e.g. lack of comprehensive information on taxpayers' characteristics). Although we could not randomise attendance to the main programme of interest, we perform several tests that give us confidence that our results are robust and unbiased. In addition to showing balance across a wide range of observable variables (section 3.1), we also put our results to closer scrutiny through IV and PSM strategies (section 3.4).

Considering both our main analysis (sections 3.2 and 3.3) and our further tests (section 3.4), our results show that taxpayer education has a large impact on knowledge and on the probability to declare in the first year since registration. The programme's effects on those two

outcomes stand up particularly well to further scrutiny, therefore representing our strongest results. Still, we provide some evidence that the programme likely also affects perceptions about complexity of the tax system and other compliance outcomes, namely nil-filing and the tax amount. The fact that the latter effect is particularly weak is in line with the only article on this topic (Chetty and Saez 2013), which shows no effect on reported income for an education programme in the United States. However, as discussed in section 3.5, it might also be due to the smaller number of observations we have for this outcome. These results highlight the importance of looking at multiple aspects of compliance, especially in contexts like Rwanda where failure to declare and nil-filing are highly prevalent phenomena. Our result on declaration rates is particularly important in light of recent evidence on the role of habit in compliance (Dunning *et al.* 2017). When taxpayers register with the RRA but fail to make a tax declaration, they are less likely to comply in the future and to engage in any meaningful way in debates around tax issues – with potential implications for the relation between taxation and accountability. We plan to explore the role of habit in further research, by looking at taxpayer behaviour in subsequent years after registration.

Furthermore, we document dramatically low levels of tax knowledge in Rwanda. The majority of new taxpayers have a very low understanding of the basic parameters of the tax system, as we show in section 3.1. While this is in line with previous evidence on taxpayer knowledge (see section 1), we use a more accurate measure of knowledge, which is an original contribution to this literature. One immediate recommendation for the RRA would be to provide some basic information to taxpayers at the point of registration, for example stating clearly what taxes they registered for and what are the related obligations. In section 3.3.1, we also report some suggestive evidence that knowledge could be an important channel for the programme's impact on compliance. In addition, we would argue, even though somewhat more speculatively, that increased knowledge is likely to be beneficial more broadly to increase taxpayers' confidence to resist corrupt practices by tax officials, to engage in a constructive dialogue around tax, and ultimately to have better views on the transparency and integrity of the tax system.

In terms of policy, our results provide three important insights. First, the main objective of taxpayer education programmes should not necessarily be only immediate revenue gains, but also a change in taxpayer habits. In the case of new taxpayers, this would mean bringing them into the habit of filing tax declarations – in addition to discouraging nil-filing and reporting income in full. In section 3.3 we show that taxpayers who file a declaration in the first year of registration are much more likely to continue doing so, compared to those who fail to file in the first year. Along similar lines, the RRA also runs other programmes, such as tax clubs in schools (see Mascagni and Santoro (2018)), that aim to build a tax culture from a young age. Although these initiatives are unlikely to yield immediate revenue gains, we believe they can play a crucial role to increase and consolidate compliance in the longer term. As such, investing in taxpayer education seems to be a sensible strategy – as part of a broader range of measures to improve compliance.

Second, more specifically to the RRA's main taxpayer education programme, our results suggest that it is highly effective, but more could be done to increase its reach, which is currently quite limited (section 3.1). This might imply increasing capacity in the department in charge of taxpayer education initiatives (in the RRA's case, the Taxpayer Services department), as we also show that they are typically under-resourced (see section 3.3.2 and Mascagni and Santoro (2018)). Relatedly, our pilot of the coaching programme (section 3.3.2) suggests that a more personalised option is currently ineffective and hardly scalable. It should only be adopted for a selected group of taxpayers, or ad hoc based on specific needs, rather than be considered as an alternative to the main education programme.

Last but not least, we recommend continuing to collect information on attendance beyond the life of this evaluation study. Carrying out such a study with the data previously available would have been impossible, although a relatively simple fix allowed us to obtain reliable and digitised attendance data (see section 2.2). This practical example illustrates the broader issue that good data is an essential foundation to evidence-based policymaking. We would therefore recommend to embed proper data collection in all initiatives where that is possible at a contained cost and effort, as it would be in this case.

References

Aiko, R. and Logan, C. (2014) *Africa's Willing Taxpayers Thwarted by Opaque Tax Systems, Corruption*, Afrobarometer Policy Paper 7, http://afrobarometer.org/sites/default/files/publications/Briefing%20paper/ab_r5_policypaperno7.pdf (accessed 28 November 2018)

Almunia, M.; Gerard, F.; Hjort, J.; Knebelmann, J.; Nakyambadde, D.; Raisaro, C. and Tian, L. (2017) *An Analysis of Discrepancies in Tax Declarations Submitted under Value-Added Tax in Uganda*, Final report S-43312-UGA-1, London: International Growth Centre

ATAF (2017) *African Tax Outlook*, Pretoria: African Tax Administration Forum

Banerjee, A.; Hanna, R.; Kyle, J.; Olken, B.A. and Sumarto, S. (2018) 'Tangible Information and Citizen Empowerment: Identification Cards and Food Subsidy Programs in Indonesia', *Journal of Political Economy* 126.2: 451–91

Benzarti, Y. (2015) 'How Taxing is Tax Filing? Leaving Money on the Table because of Hassle Costs', mimeo, University of California, Berkeley

Carpena, F.; Cole, S.A.; Shapiro, J. and Zia, B.H. (2015) *The ABCs of Financial Education: Experimental Evidence on Attitudes, Behavior, and Cognitive Biases*, Policy Research Working Paper 7413, Washington DC: World Bank Group

Carrillo, P.; Pomeranz, D. and Singhal, M. (2017) 'Dodging the Taxman: Firm Misreporting and Limits to Tax Enforcement', *American Economic Journal: Applied Economics* 9.2: 144–64

Chetty, R. and Saez, E. (2013) 'Teaching the Tax Code: Earnings Responses to an Experiment with EITC Recipients', *American Economic Journal: Applied Economics* 5.1: 1–31

Chetty, R.; Looney, A. and Kroft, K. (2009) 'Salience and Taxation: Theory and Evidence', *American Economic Review* 99.4: 1145–77

Dehejia, R.H. and Wahba, S. (1999) 'Causal Effects in Nonexperimental Studies: Reevaluating the Evaluation of Training Programs', *Journal of the American Statistical Association* 94.448: 1053–62

Dunning,T.; Monestier, F.; Pineiro, R.; Rosenblatt, F. and Tunón, G. (2017) 'Is Paying Taxes Habit Forming? Theory and Evidence from Uruguay', Technical Report, mimeo, www.thaddunning.com/wp-content/uploads/2017/09/Dunning-et-al_Habit_2017.pdf (accessed 28 November 2018)

Feldman, N.; Katuščák, P. and Kawano, L. (2016) 'Taxpayer Confusion: Evidence from the Child Tax Credit', *American Economic Review* 106.3: 807–83

Gangl, K.; Torgler, B.; Kirchler, E. and Hofmann, E. (2014) 'Effects of Supervision on Tax Compliance: Evidence from a Field Experiment in Austria', *Economics Letters* 123.3: 378–82

Heckman, J.; Ichimura, H. and Todd, P.E. (1997) 'Matching as an Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Programme', *The Review of Economic Studies* 64.4: 605–54

Imbens, G. and Rubin, D.B. (2015) *Causal Inference for Statistics, Social, and Biomedical Sciences*, New York NY: Cambridge University Press

Isbell, T. (2017) *Tax Compliance: Africans Affirm Civic Duty but Lack Trust in Tax Department*, Afrobarometer Policy Paper 43, http://afrobarometer.org/sites/default/files/publications/Policy%20papers/ab_r6_policypaperno43_tax_compliance_in_africa-afrobarometer.pdf (accessed 28 November 2018)

Kleven, H.J.; Knudsen, M.B.; Kreiner, C.T.; Pedersen, S. and Saez, E. (2011) 'Unwilling or Unable to Cheat? Evidence from a Tax Audit Experiment in Denmark', *Econometrica* 79.3: 651–92

Liebman, J.B. and Luttmer, E.F.P. (2015) 'Would People Behave Differently if they Better Understood Social Security? Evidence from a Field Experiment', *American Economic Journal: Economic Policy* 7.1: 275–99

Manoli, D.S. and Turner, N. (2014) *Nudges and Learning: Evidence from Informational Interventions for Low-Income Taxpayers*, NBER Working Paper 20718, Cambridge MA: National Bureau of Economic Research

Mascagni, G. (2018) 'From the Lab to the Field: A Review of Tax Experiments', *Journal of Economic Surveys* 32.2: 273–301

Mascagni, G. and Mengistu, A. (2016) *The Corporate Tax Burden in Ethiopia: Evidence from Anonymised Tax Returns*, ICTD Working Paper 48, Brighton: International Centre for Taxation and Development

Mascagni, G.; Mengistu, A.T. and Wodleyes, F.B. (2018) *Can ICTS Increase Tax? Experimental Evidence from Ethiopia*, ICTD Working Paper 82, Brighton: International Centre for Tax and Development

Mascagni, G.; Monkam, N. and Nell, C. (2016) *Unlocking the Potential of Administrative Data in Africa: Tax Compliance and Progressivity in Rwanda*, ICTD Working Paper 56, Brighton: International Centre for Taxation and Development

Mascagni, G.; Nell, C. and Monkam, N. (2017) *One Size Does Not Fit All: A Field Experiment on Drivers of Compliance and Delivery Methods in Rwanda*, ICTD Working Paper 58, Brighton: International Centre for Taxation and Development

Mascagni, G. and Santoro, F. (2018) *What is the Role of Taxpayer Education in Africa?*, African Tax Administration Paper 1, Brighton: International Centre for Tax and Development

Mukama, D.; Karangwa, J. and Hakizimana, N. (2017) *Communicating to Improve Compliance: Taxpayers' Feedback on Message and Mode of Delivery in Rwanda*, ICTD Working Paper 59, Brighton: International Centre for Taxation and Development

Pomeranz, D. (2015) 'No Taxation Without Information: Deterrence and Self-Enforcement in the Value Added Tax', *American Economic Review* 105.8: 2539–69

Rubin, D. (1987) *Multiple Imputation for Nonresponse in Surveys*, New York NY: John Wiley

Shimeles, A.; Gurara, D.Z. and Woldeyes, F. (2017) 'Taxman's Dilemma: Coercion or Persuasion? Evidence from a Randomized Field Experiment in Ethiopia', *American Economic Review* 107.5: 420–24

Sianesi, B. (2004) ‘An Evaluation of the Swedish System of Active Labour Market Programmes in the 1990s’, *Review of Economics and Statistics* 86.1: 133–55

van den Boogaard, V.; Prichard, W. and Jibao, S. (2018) *Norms, Networks, Power, and Control: Understanding Informal Payments and Brokerage in Cross-Border Trade in Sierra Leone*, ICTD Working Paper 74, Brighton: International Centre for Tax and Development

Appendix

Table A1 The TPS training plan for 2017/18

	District	Date
1	Kicukiro	21 August 2017
2	Musanze	29 August 2017
3	Rubavu	30 August 2017
4	Muhanga	22 November 2017
5	Rusizi	29 November 2017
6	Huye	6 December 2017
7	Nyagatare	20 December 2017
8	Kicukiro	23 January 2018
9	Gasabo	6 February 2018
10	Nyarugenge	9 February 2018
11	Kicukiro	20 February 2018
12	Gasabo	22 February 2018
13	Nyarugenge	7 March 2018

Table A2 Balance tests by attrition status

	Non-attrited		Attrited			
	Mean	Obs.	Mean	Obs.	Difference	Δ
Covariates						
Age	33.24	821	32.82	186	0.42	-0.04
Female	0.30	821	0.41	186	-0.11***	0.22
No school or primary only	0.20	821	0.14	186	0.06*	-0.16
University degree	0.43	821	0.54	186	-0.10**	0.20
Owner	0.79	821	0.62	186	0.17***	-0.38*
Individual business	0.52	821	0.45	186	0.07*	-0.13
>5 employees	0.09	821	0.11	186	-0.02	0.07
Email use	0.26	821	0.27	186	-0.01	0.02
Bank account	0.45	821	0.44	186	0.01	-0.03
Books of account	0.48	821	0.51	186	-0.03	0.06
Tax time use days	2.00	605	1.32	134	0.68	-0.15
Had previous business	0.21	821	0.23	186	-0.03	0.06
Had previous training	0.06	821	0.10	186	-0.03	0.12
# RRA visits	1.39	716	1.32	158	0.07	-0.03
Knowledge and perceptions						
Baseline Knowledge Index	6.37	821	6.42	186	-0.06	0.02
Complexity	0.71	796	0.62	177	0.10**	-0.21
Evasion attitude	0.33	782	0.34	173	-0.01	0.02
Fairness	0.98	788	0.98	176	0.00	-0.02
Government authority	0.95	804	0.96	181	-0.00	0.01
Social duty	0.98	819	0.97	185	0.01	-0.06
Public service satisfaction	0.007	802	-0.03	182	0.04	-0.04

Observations weighted by sampling weights. T-test are computed on the *Difference* across the two groups. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Δ is the normalised difference from Imbens and Rubin (2015), equal to the difference in means scaled by the square root of the average of the two within-group variances. Obs. – observations.

Table A3 Training effect on knowledge – interactions

	(1) F Increase	(2) F Gain	(3) K Increase	(4) K Gain	(5) P Increase	(6) P Gain	(7) L Increase	(8) L Gain
Training	0.245***	0.529***	0.190***	0.602***	0.270***	0.574***	0.280***	0.587***
	(0.036)	(0.079)	(0.042)	(0.086)	(0.033)	(0.079)	(0.030)	(0.074)
Group	-0.066	-0.182**	-0.083*	0.02	-0.080	-0.256***	-0.075	0.062
	(0.054)	(0.083)	(0.047)	(0.064)	(0.061)	(0.085)	(0.118)	(0.153)
Group*Training	0.080	0.092	0.126**	-0.070	-0.004	-0.099	-0.122	-0.346**
	(0.063)	(0.144)	(0.058)	(0.122)	(0.075)	(0.124)	(0.118)	(0.153)
Covariates	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Mean of Y	0.714	0.571	0.714	0.571	0.714	0.571	0.714	0.571
R^2	0.208	0.360	0.211	0.360	0.206	0.360	0.208	0.361
Observations	820	815	820	815	820	815	820	815

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors in parentheses. *Increase* is a dummy variable indicating whether the post-pre knowledge scores difference is positive. *Gain* is the ratio of post-pre scores difference over the pre-training score. Interactions used are with dummies for female, Kigali, primary education and large size. Covariates used are those indicated in section 2.3. All regressions are weighted by sampling weights.

Table A4 Balance tests by assignment to coaching

	Control		Coaching		
	Obs.	Mean	Obs.	Mean	Difference
Covariates					
Age	279	32.40	293	32.56	-0.15
Female	279	0.30	293	0.31	-0.01
No school or primary only	279	0.25	293	0.18	0.07**
University degree	279	0.45	293	0.44	0.01
Owner	279	0.76	293	0.82	-0.06
CIT	270	0.46	283	0.43	0.03
>5 employees	279	0.06	293	0.08	-0.02
Life in months	270	4.44	283	4.63	-0.18
Located in Kigali	272	0.51	281	0.52	-0.00
Email use	279	0.24	293	0.25	-0.01
Bank account	279	0.45	293	0.40	0.05
Books of account	279	0.46	293	0.44	0.03
Tax time use days	214	1.73	200	2.24	-0.52
Had previous business	279	0.18	293	0.22	-0.04
Had previous training	279	0.06	293	0.08	-0.02
Knowledge and perceptions					
Baseline Knowledge Index	279	6.19	293	6.25	-0.05
Complexity	267	0.71	281	0.70	0.01
Evasion attitude	264	0.32	278	0.32	0.00
Fairness	265	0.97	281	0.98	-0.01
Government authority	272	0.95	287	0.95	-0.00
Social duty	279	0.96	291	0.99	-0.03**
Public service satisfaction	273	0.01	283	-0.01	0.02
Observations	572				

Observations weighted by sampling weights. T-test are computed on the *Difference* across the two groups. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.
Obs. – observations.

Table A5 Perception indexes description

Index	Components	Baseline
Complexity	Difficult to file tax declaration – Difficult to get in touch with RRA	0.693
Evasion	Right for some people not to pay taxes – Fair not to pay tax if neighbours don't pay – Most businesses evade at least part of their taxes	0.316
Fairness	Tax system is fair – Businesses sometimes bribe tax officials – Trust in RRA – Enforcement probability is high	0.940
Government authority	Government can make people pay to increase public services – Government always has a right to make people pay	0.933
Social duty	Must pay tax for Rwanda to develop – Tax is a social duty	0.979

Indexes are 1-0 dummies with the following meaning: 1 – agree or neutral; 0 – disagree.

Table A6 Training effect on probability to declare – interactions

	(1) No interact.	(2) Female	(3) Kigali	(4) Primary	(5) Large	(6) Top know
Training	0.12***	0.07	0.10**	0.10**	0.11**	0.12***
	(0.04)	(0.05)	(0.05)	(0.04)	(0.04)	(0.05)
Group		-0.02	0.00	-0.17**	-0.11	0.19**
		(0.06)	(0.05)	(0.06)	(0.09)	(0.09)
Group*Attend		0.15*	0.02	0.17*	0.11	-0.02
		(0.08)	(0.07)	(0.10)	(0.12)	(0.09)
Covariates	Yes	Yes	Yes	Yes	Yes	Yes
Knowledge/perceptions	Yes	Yes	Yes	Yes	Yes	Yes
Observations	980	980	980	980	980	980

$p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors in parentheses. Coefficients are marginal effects from probit regressions. Covariates used are those indicated in section 2.3. All regressions are weighted by sampling weights.

Table A7 Impacts of knowledge increase on tax outcomes – IV strategy

	(1) Inc. nilfile	(2) Inc. nilfile	(3) Gain nilfile	(4) Gain nilfile	(5) Inc. tax	(6) Inc. tax	(7) Gain tax	(8) Gain tax
Knowledge	-0.26	-0.26	-0.10	-0.12	3.56	3.34	1.21	1.35
	(0.31)	(0.23)	(0.11)	(0.10)	(3.51)	(2.74)	(1.13)	(1.06)
Covariates	No	Yes	No	Yes	No	Yes	No	Yes
Knowledge/perceptions	No	Yes	No	Yes	No	Yes	No	Yes
Mean of Y	0.605	0.605	0.605	0.605	3.480	3.480	3.480	3.480
Observations	280	280	280	280	262	262	262	262

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors in parentheses. The instrument is attendance to the training. *Increase* is a dummy variable indicating whether the post-pre knowledge score difference is positive. *Gain* is the ratio of the difference over the 0-10 pre-training score. Exclusion restriction holds: p-value of the IV in a regression with Knowledge is 11% for both Inc. nilfile and Inc. tax and 26% for Gain nilfile and 17% for Gain tax. Covariates used are those indicated in section 2.3. All regressions weighted by sampling weights.

Table A8 Tests for exclusion restriction

	(1) Declare	(2) Nilfile	(3) Log tax
Training	0.17***	-0.11***	1.00***
	(0.03)	(0.03)	(0.39)
Assignment survey	0.01	0.05	-0.34
	(0.02)	(0.03)	(0.38)
Covariates	Yes	Yes	Yes
R^2	0.217	0.070	0.025
Observations	2,378	907	860

Standard errors are in parentheses.
Coaching group has been dropped.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A9 Impact of training on probability to nilfile

	(1) Survey	(2) All	(3) Survey rural	(4) All rural	(5) Survey urban	(6) All urban
Training	-0.123**	-0.073***	-0.133	-0.195***	-0.102	0.000
	(0.060)	(0.016)	(0.087)	(0.040)	(0.078)	(0.017)
Admin. covariates	Yes	Yes	Yes	Yes	Yes	Yes
Tax centre FEs	Yes	Yes	Yes	Yes	Yes	Yes
Observations	393	6,410	188	1,169	203	5,240

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors in parentheses.
Coefficients are marginal effects from probit regression computed at the mean of covariates.
All regressions are weighted by sampling weights. FEs – fixed effects.

Table A10 Impact of training on the log tax declared

	(1) Survey	(2) All	(3) Survey rural	(4) All rural	(5) Survey urban	(6) All urban
Training	1.065*	0.543***	0.910	1.716***	0.871	-0.167
	(0.643)	(0.159)	(0.892)	(0.368)	(0.862)	(0.194)
Admin. covariates	Yes	Yes	Yes	Yes	Yes	Yes
Tax centre FEs	Yes	Yes	Yes	Yes	Yes	No
Observations	361	6,078	174	1,034	187	5,044

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors in parentheses.
Coefficients are marginal effects from tobit regression computed at the mean of covariates.
All regressions are weighted by sampling weights. FEs – fixed effects.

Figure A1 Propensity score distribution by training groups