Revealing Tax Evasion: Experimental Evidence from a Representative Survey of Indonesian Firms*

Christopher Hoy† Filip Jolevski‡ Anthony Obeyesekere§

November 6, 2024

Abstract

This paper examines the pervasiveness of tax evasion in Indonesia and the characteristics associated with non-compliance through a randomized, double-list experiment embedded in a nationally representative survey of 2,955 registered firms. This revealed that around a quarter of firms indirectly admit to evading tax, which is equivalent to a revenue loss of two percent of GDP. Firms that do not export, face informal sector competition, and believe tax administration is a major obstacle are the most likely to evade taxes. These findings help inform tax authorities, which face substantial challenges in estimating levels of tax evasion and identifying non-compliant taxpayers.

Keywords: Tax Evasion; Experiment; Firms; Public Finance, Compliance

JEL Codes: H25, H26, E26, D22, O17, O53, C83, C93

*The findings, interpretations, and conclusions expressed in this paper are entirely those of the authors. They do not necessarily represent the views of the World Bank and its affiliated organizations, or those of the Executive Directors of the World Bank or the governments they represent. The authors are incredibly grateful for the support received throughout this study from Jorge Luis Rodriguez Meza, Habib Rab, Alexandre Laure, Rong Qian and Stephen Davenport. We also want to thank Revocatus Paul for his outstanding support with data analysis. The authors are also very appreciative of comments received from Lars Christian Moller, Emilia Skrok, Anne Brockmeyer, Jonathan David Timmis, Mahvish Shaukat, Khudadad Chattha, David C. Francis, Norman Loayza, Joshua Wimpey, Davide Mare, Attique Ur Rehman, and Nazım Tamkoç. In addition, we are very thankful for comments and suggestions provided by staff at the Indonesia Directorate General of Taxation, participants in the World Bank DECIG Seminar. The work was made possible by funding support from European Union, SECO and Canada under Indonesia PFM MDTF Phase III Program, together with funding from the World Bank Global Tax Program and the Gates Foundation.

†Christopher Hoy (corresponding author, choy@worldbank.org) is an Economist with the Poverty and Equity Global Practice at the World Bank.

‡Filip Jolevski (fjolevski@worldbank.org) is an Economist with the Enterprise Analysis Unit, Development Economics at the World Bank and with the Department of Economics at George Mason University.

§Anthony Obeyesekere (aobeyesekere@worldbank.org) is a Country Economist in the Indonesia team of the Macroeconomics, Trade and Investment Global Practice.

1 Introduction

Tax evasion by firms in middle-income countries is often thought to be one of the key reasons why revenue levels as a share of GDP remain well below those in high-income countries (Slemrod, 2019; World Bank, 2023). However, it can be challenging to measure how pervasive tax evasion is by firms by drawing on tax administrative and survey data (Slemrod and Weber, 2012; Timofte (Coca) et al., 2019). Estimates based on administrative data typically rely on tax declarations being checked for accuracy with the limited third-party information that is available in these settings, and it can be unclear whether lower-than-expected tax declarations are due to changes in economic activity, misreporting, or genuine evasion (Pomeranz, 2015; Carrillo et al., 2017). On the other hand, relying on surveys of taxpayers also has its challenges as respondents may be hesitant to directly admit to evading tax, which leaves many studies using questions about "tax morale" as a proxy for compliance (e.g., see discussion in Luttmer and Singhal, 2014; Ali et al., 2014; and Hoy, 2022). The absence of reliable information about tax evasion levels can undermine revenue authorities' ability to improve firms' compliance and identify the types of firms they should focus their enforcement efforts on (Slemrod, 2008; Mascagni et al., 2023).

To overcome these challenges, we embedded an experiment that indirectly captures levels of tax evasion as part of a nationally representative survey of 2,955 firms in Indonesia. Specifically, we use a list experiment (also known as the "item count technique") whereby respondents only state the number of statements on a list that apply to them but do not reveal directly which of the statements are true (see detailed explanation of this approach in Ahlquist, 2018, Blair and Imai, 2012, Imai, 2011 and Rosenfeld et al., 2016). We draw on best practices in measuring sensitive topics, such as drug use and risky sexual behavior, by fielding a "double" list experiment that overcomes some of the concerns with traditional "single" list experiments[1] (e.g., see Chuang et al., 2021). Namely, all respondents are randomly

[1]For example, one of the most common criticisms of list experiments is that respondents' willingness to answer honestly may be driven by the non-sensitive statements on the list. A double list experiment allows us to check whether, across the two lists with different non-sensitive statements, there are similar levels of

allocated to the treatment group in the first or second of two sequential list experiments. This approach means that estimates across the two list experiments can be directly compared, serving as an internal robustness check to demonstrate the accuracy of the results. We implement this double list experiment as part of one of the largest World Bank Enterprise Surveys (WBES) to date, covering a nationally representative sample of 2,955 registered firms with at least five employees. WBES are among the most comprehensive and reliable surveys about the activities of small, medium, and large firms, containing detailed information about firm characteristics and beliefs, which allows us to examine the heterogeneity of tax evasion extensively.

The results show that around one-quarter of firms report evading taxes and evasion rates vary substantially on only a relatively small number of dimensions. The "double" list experiment produced internally consistent estimates of around 26 percent (varying from 25 to 27 percent) of firms admitting to not paying all the taxes they are required to pay (which is an extensive margin estimate of tax evasion). This result is almost identical across both experiments, which provides considerable reassurance of its accuracy, and it represents a lower bound estimate as it is based on self-reported tax evasion, which is likely to be lower than actual tax evasion. A back-of-the-envelope calculation based on this result suggests that revenue lost from tax evasion by registered firms is around two percent of GDP. We use machine learning to identify, among an extensive set of firm characteristics, the dimensions in which substantial and consistent variation in tax evasion exists. This revealed that firms that do not export, face substantial competition from the informal sector, and believe tax administration is a major obstacle to their business activities are the most likely to evade taxes. There was limited variation across both list experiments in reported tax evasion based on other dimensions, highlighting just how widespread this behavior is.

These results provide important insights into how common tax evasion is among firms and the specific firm characteristics and beliefs most associated with evading tax. Many

tax evasion reported.

challenges associated with firms evading taxes in Indonesia are similar to those in other middle-income countries where tax revenue to GDP has remained stubbornly low (Alm, 2019; Basri et al., 2021). This study is particularly valuable for revenue authorities in these settings that face sizable challenges in accurately estimating levels of tax evasion and identifying non-compliant taxpayers. The results show that substantially more revenue could be collected if the government increased compliance by registered firms.[2] This study may also assist revenue authorities with their compliance activities by identifying the dimensions where the most significant heterogeneity exists in tax evasion between firms. For example, the higher rates of tax evasion among non-exporting firms suggest that many firms may perceive that the revenue authority lacks third-party information about economic activities that do not cross international borders. Furthermore, the notably higher rates of tax evasion by firms that see tax administration as a major obstacle to their business activities provide suggestive evidence that efforts to simplify the tax system may lead to increases in compliance.

This study contributes to two strands of the existing literature about tax evasion by firms and methods for surveying sensitive topics. Regarding research about tax evasion by firms, we make three main contributions. First, we focus on measuring levels of tax evasion in a middle-income country in a novel way compared to previous studies that have typically relied on tax administrative data for a subset of firms (e.g., see Pomeranz, 2015; Carrillo et al., 2017; Basri et al., 2021; Waseem, 2022; Waseem, 2023; Artavanis et al., 2016)[3] or survey questions that either directly ask about compliance or focus on tax morale (both of which are likely to underestimate tax evasion) (e.g., see Ali et al., 2014; Luttmer and Singhal, 2014; Hoy, 2022). Secondly, our focus is on the tax compliance behavior of almost exclu-

[2]However efforts by the authorities to increase compliance could raise the incentives for firms to remain informal.

[3]Some of these studies draw solely on tax administrative data to estimate evasion among taxpayers close to thresholds or between different sectors. While this work provides incredibly valuable insights into tax evasion among specific subsets of firms, it is rarely generalizable to all registered firms.

sively domestically owned firms,[4] as opposed to multi-national firms, which have received far more attention in the recent literature (e.g., see Tørsløv et al., 2023) and can evade taxes in unique ways (such as transfer pricing across countries). Thirdly, by implementing a double list experiment as part of the WBES, we have far more comprehensive data about firm characteristics to examine heterogeneity than prior studies. It is worth noting that while there has been rapid growth of work examining ways to increase tax compliance (primarily through using behavioral science), these studies have not focused directly on measuring tax evasion (instead, they aim to improve compliance through multiple channels, such as increasing the salience of deadlines) (e.g., see Suharnoko et al., 2020; Mascagni, 2017 and Hoy et al., 2024).

Concerning survey methodology, this work contributes to the literature on how indirect elicitation methods can be used to provide a more accurate depiction of sensitive topics in surveys (see discussion about these approaches in Ahlquist, 2018, Blair and Imai, 2012, Imai, 2011 and Rosenfeld et al., 2016). While traditional single list experiments have been used in an array of studies on different topics in low- and middle-income countries (e.g., see Chuang et al., 2021; Blair et al., 2014; McKenzie and Siegel, 2013; Karlan and Zinman, 2012), the use of double list experiments is still in a nascent stage. We add considerable value to this line of inquiry by implementing one of the most extensive double list experiments to date (see Yang and Moerenhout, 2024 for an example of a large-scale single list experiment in a middle-income country) and collecting data from a nationally representative sample so that inferences can be made about the general population. In addition, this is the first study to apply the double list experiment approach to tax evasion of firms (single list experiments examining tax evasion include Genest-Grégoire et al., 2022, Heide-Jørgensen, 2022 and Iraman et al., 2022). Furthermore, to the best of our knowledge, this is the first study to combine the use of a double list experiment and machine learning algorithms to extensively examine variation in levels of the sensitive item among subgroups of respondents.

This paper is structured as follows. Section 2 provides details about the study's set-

[4]Only 97 of the 2,955 firms were foreign owned. As such we refer to the survey sample as effectively being solely domestically owned firms.

ting, Section 3 describes the data, and Section 4 explains the methodology used. Section 5 presents the findings, and Section 6 discusses the implications of the results. The Appendix provides more details about the study's design and the analysis that was conducted.

2 Setting

Domestic revenue mobilization efforts in Indonesia lag many comparable countries. Between 2009 and 2019, Indonesia's revenue collection as a share of GDP was less than two-thirds of the average for countries in Emerging East Asia and less than half of the average for Emerging Market economies more generally. Even though reasonably similar tax policies are in place in some neighboring countries (e.g., the Philippines, Cambodia, and Malaysia), Indonesia collects several percentage points of GDP less in tax revenue. Moreover, Indonesia's revenue as a share of GDP exhibited a downward trend from a peak of around 20 percent in 2009 and falling to 15 percent of GDP by 2019 (World Bank 2024a; IMF, 2024). Weak tax compliance is often blamed for the country's relatively low and worsening revenue collection.

Taxes paid by firms constitute the primary source of government revenue and thus are essential for examining why relatively low tax revenue exists. Over the past three decades, taxes collected from firms (e.g., Corporate Income Tax (CIT) and Value Added Tax (VAT), among others) have gained increased significance in Indonesia's tax revenue composition. These taxes increased from 60 percent of total tax revenue in the 1990s to around 67 percent in the 2010s.[5] This has occurred concomitantly with a decline (around six percentage points) in the share of tax revenue collected in personal income taxes. Indonesia has a statutory CIT rate of 22 percent on business profits. Firms with gross annual revenue less than around US 3.2 million receive a 50 percent tax rate discount on the profits attributable to their first US 310,000 in gross revenue. When a firm is newly registered, it may be temporarily eligible for a simplified alternative scheme with a tax rate of 0.5 percent on gross revenues (rather than the CIT on profits). Firms are eligible so long as gross annual revenue is less than around US

[5]IMF data and World Bank staff calculations.

310,000, and eligibility expires after 3 years for incorporated companies or 4 years for other registered businesses. Firms in the construction, real estate, shipping, and airline services sectors are also subject to special final tax rates on their gross turnover, instead of the CIT. There is a statutory VAT rate of 11 percent. There are various exempted or zero-tax-rated categories, including essential food, health, education, financial services, and exports. Participation in the VAT regime is only compulsory for firms with gross annual revenues of at least US 310,000, though smaller firms may voluntarily opt-in.

There is considerable evidence that incomplete tax compliance by firms is pervasive in Indonesia. Previous studies have found that nine out of ten firms were not registered with the revenue authority (Rothenberg et al., 2016), and that most registered firms reported zero net income and paid no income tax (e.g., Ikhsan et al., 2005). Audit data suggests that in some instances there is substantial misreporting between taxpayer self-declared liabilities and auditor-assessed liabilities among firms that do pay tax.[6] Moreover, such non-compliance seems exceptionally high among professional services corporations (Breuer et al., 2018). Recent World Bank (2024b) research has estimated Indonesia's combined CIT and VAT compliance gap at around 3.8 percent of GDP in lost revenues per year from 2016 to 2021. This includes losses due to deliberate or unintentional underpayment of businesses' tax obligations, such as the type of tax evasion focused on in this paper. More generally, a few studies have also estimated Indonesia's tax losses based on estimates of the informal economy. Tax losses are calculated by applying an average tax rate to this economic activity which is assumed to have remained outside the tax net. These studies have arrived at a wide range of tax-loss estimates, from 0.7 to 3.8 percent of GDP per annum (see Tatariyanto, 2014; Mulyawan, 2017; Nizar and Purnomo, 2011; Ramadhan, 2019; Indupurnahayu and Walujadi, 2019).

The revenue authority in Indonesia, the Directorate General of Taxation (DGT), has

[6]Based on a joint World Bank and Directorate General of Taxation exercise conducted in 2017. It analyzed a sample of business tax returns, comparing income tax estimated by taxpayers with that estimated by tax auditors.

implemented many strategies to address low compliance by firms. Identification of unregistered firms includes drawing on data from business registries, licensing regimes, and banking systems. To identify tax evasion among registered firms, DGT takes steps to determine the accuracy of tax returns, including implementing a compliance risk management model that predicts the likelihood of non-compliance. These predictions are partly based on tax audit results covering around 2 percent of registered firms each year. However, there is no public reporting on the scale of misreporting identified by these audits, nor any estimates of total tax losses if rates of identified misreporting were to be extrapolated.

3 Data

The data used in this analysis comes from the 2023 WBES in Indonesia, which interviewed the top managers or owners of 2,955 firms. This sample was nationally representative of firms possessing a Company Registration Certificate (TDP) or Business Identification Number (NIB) with five or more employees with at least 1 percent of private ownership and who do not have legal status as cooperatives. Firms are selected through stratified random sampling from the 2016 Economic Census conducted by the Central Agency of Statistics of Indonesia. Stratification is based on sector, firm size (employment),[7] and location. The sectoral distribution covers firms from the manufacturing and major services sectors. The survey covers all 38 provinces of Indonesia, though some are aggregated, resulting in a total of 22 regional strata. The sample design ensures no more than 7.5 percent margins of error and 90 percent confidence intervals at each of the stratification levels: size, sector, and region. All interviews were conducted face-to-face with top managers or business owners in Bahasa Indonesian through Computer-Assisted Personal Interviews (CAPI) using tablets and Survey Solutions as the software for data collection. Data collection started in December 2022 and concluded in September 2023, achieving a response rate of 41.2 percent, similar

[7]The stratification by firm size is defined using the number of employees in the firm. The definitions are consistent across all countries where the WBES has been implemented and are defined as small (5-19 employees), medium (20 to 99 employees), and large (100 and more).

to response rates for WBESs in other countries. The replacement of businesses refusing to participate in the survey was done within the same stratum, following a randomly generated preference order of which businesses should be contacted for an interview. All interviews were subject to the WBES protocols[8] of quality assurance. Prior to the launch of data collection, the team of interviewers underwent a 5-day training to understand the methodology, protocols, and each question that appears in the instrument. The training was used to verify and improve the instrument's translation, and this was fine-tuned through pilot interviews with a small sample of firms before the formal launch of data collection.

The survey primarily consisted of detailed questions about firm characteristics and their operating environment, followed by the double list experiment (see the following section). The questions about firm characteristics and their operating environment are part of standardized modules included in WBESs that have been implemented in more than 150 countries. This includes specific questions about how firms have interacted with the government (e.g., the time required to obtain permits and encounters with corruption) and firms' experience paying tax (e.g., which types of tax they must pay and if tax officials have visited them). The extensive list of questions about firm characteristics and their operating environment provides a rich dataset to examine the heterogeneity of the double list experiment.

4 Design of the Experiment

A "double" list experiment was included in the 2023 WBES in Indonesia to estimate levels of tax evasion by firms. List experiments (also known as the item count technique) provide a way to estimate levels of tax evasion without having to directly ask respondents if they engage in such activities. This indirect approach is typically seen as a more credible way of asking questions about sensitive topics as respondents provide "concealed" answers that can still be used to generate estimates of sensitive responses (originally proposed and elaborated on by Raghavarao and Federer, 1979; Miller, 1984; Droitcour et al., 2004). Specifically, list

[8]Details about the methodology and protocols can be found here: WBES Methodology

experiments involve randomly allocating respondents into two groups. They are then shown a list of statements and asked to state the number of statements (but not which specific statements) are true in their case. The list shown to the "treatment" group includes one extra statement about the sensitive issue (e.g., not paying all the taxes that they owe). While it is not known which specific respondents have agreed with the statement on the sensitive issue, differences in the average number of statements reported as true between the treatment and control groups can be used to estimate the prevalence of the sensitive response. If respondents do not engage in sensitive activities, the mean number of items should be the same between the control and treatment groups. The double list experiment approach is an extension of the traditional single list experiment: respondents complete two list experiments sequentially and are randomly allocated to the treatment group in one of the two.

The double list experiment included as part of the WBES is shown in Table 1. Respondents were randomly assigned to either Group A or Group B. Each group was given two list questions to respond to, but with a slight variation between the groups. Group A received the treatment in List 2 while being the control for List 1 (note the statement about taxes in Group A's List 2). Group B had the opposite arrangement, receiving the treatment in List 1 while serving as control in List 2. The treatment is the inclusion of a sensitive item asking about tax evasion, specifically "This establishment does not pay all the taxes it is required to pay." The exact location of this statement in the list of statements was randomized. This text is identical to what was used in previous studies asking directly about tax evasion and has been trialed as part of single list experiments on firms in other middle-income countries (e.g., see Dom et al., 2022). The other statements in the double list experiment were carefully determined to minimize the risk that respondents could gauge the focus of the experiments. Consultations with private sector experts in Indonesia were conducted to ensure the statements achieve this aim, and they were subsequently fine-tuned through extensive piloting. In the case of List 1 in Table 1, the first statement is expected

to apply to most respondents, while the second item is expected to be very uncommon. In the case of List 2, the second item (and the third item to a lesser extent) was expected to apply to most respondents, while the first item was expected to be much less common.

Table 1

The experiment was administered during the interview with the top manager or owner of the firm, following the survey section about interactions with the government. The assignment to group A or B was randomly determined when the interview appointment was scheduled (see the balance table, Table A1 in the Appendix, showing that random assignment was successful). The interviewers were not informed whether the respondent was assigned to group A or B. Once the interview reached the point in which the experiment was administered, the interviewers informed the respondent that they were only interested in the number of items that were accurate for them, emphasizing that they were not interested in which ones were accurate and that they could not comment on or provide any clarifications of any of the items. The interviewers showed respondents the list of items to count on the tablet on which survey responses were being collected. The CAPI system randomized the order in which the items appeared. Respondents only provided the number of accurate items in each list.

The double list experiment was designed with several considerations in mind based on an extensive review of best practices in the recent literature (e.g., Imai, 2011; Blair and Imai, 2012; Glynn, 2013; Rosenfeld et al., 2016; Ahlquist, 2018; Chuang et al., 2021). Firstly, the number of items provided to respondents must be brief to minimize respondent fatigue but long enough to reassure respondents that their answers will be concealed credibly. As such, we chose three "non-sensitive" items alongside the tax evasion item, which is consistent with many recent studies. Secondly, it is essential to avoid "ceiling" and "zero" effects (where most respondents report none of or all the items on the list). Otherwise, it will be evident to

the respondent that their answer to the tax evasion item will not be concealed. Therefore, we included one non-sensitive item most respondents were expected to select and another most respondents were not likely to. Thirdly, the theme of the non-sensitive items should be broadly related to the sensitive item. Otherwise, the latter may stand out too starkly to respondents, and they may be less inclined to report honestly. To address this, we ensure all non-sensitive items in both list experiments are framed around the business activities of the "establishment" that the respondent is answering on behalf of. Finally, we field a double list experiment with many strengths relative to a traditional single list experiment, including the fact that answers can be compared between the two list experiments to see if a similar prevalence of tax evasion is reported. By doing so, we provide credible evidence to dispel one of the most common criticisms of list experiments: that the non-sensitive items on the list may drive respondents' willingness to answer honestly. A double list experiment allows us to check whether similar levels of tax evasion are reported across the two lists with different non-sensitive items.

5 Findings

5.1 Main Results

Figure 1 shows the distribution of responses to the two list questions, disaggregated by treatment (blue) and control (pink) groups. The horizontal axis represents the number of accurate statements in the list experiments selected by respondents. The treatment group was presented with a list of four statements, while the control group was presented with a list of three statements. Most respondents had selected that just one or two statements were accurate in both lists of experiments. Relatively few selected all or none of the statements (which provides strong evidence against ceiling or floor effects being a significant concern). Both list experiments demonstrated that respondents allocated to the treatment group (i.e., Group B in List 1 and Group A in List 2) were more likely to state a higher number of items

being correct. This is consistent with tax evasion being quite prevalent among respondents.

Figure 1

Table 2 shows that both list experiments suggest that around one-quarter of businesses admit to not paying all the taxes they are required to pay. In the first list experiment, the average number of statements reported as true were 1.43 for the treatment group (which were given the additional tax evasion statement) and 1.16 for the control group. The difference in means, 0.27, equates to a tax evasion prevalence of 27 percent among respondents.[9] In the second list experiment, the average number of statements reported as true was 1.71 for the treatment group and 1.46 and for the control group. The difference in means, 0.25, equates to a tax evasion prevalence of 25 percent among respondents. There was no significant difference in the size of the treatment effects between the first and second list experiments, which provides strong evidence that these effects are not simply a matter of chance. Moreover, respondent in groups A and B each selected around 2.86 items on average across both list experiments (see row 3 of Table 2), which provides considerable reassurance that what is included as non-sensitive items in each of the specific list experiments is not affecting the results. In addition, there were no differences in the background characteristics of firms across the two groups (see Table A1 in the Appendix). The only meaningful difference between groups A and B was whether they were randomly allocated to receive the tax evasion item in the first or second list experiment.

Table 2

Several robustness checks were conducted to illustrate the reliability of the main results.

[9]Put differently, when presented with one additional statement about evading taxes, the average number of statements reported as being true was 0.27 higher. This may be interpreted as 27 percent of treatment group respondents agreeing with the tax evasion statement.

Firstly, we reproduce the results by applying weights to the sample of businesses to generate the weighted average treatment effects. The weights employed were based on firm size, sector, and region. This did not have a qualitative impact on the findings (see Table A2 in the Appendix), but as expected, the level of variance increases once weights are applied. Table A2 is restricted to all respondents who participated in both listing experiments. This shrinks the sample size to 2,272 but does not meaningfully change our results. Finally, we reproduce our analysis, including all respondents who refused to answer these questions and treated them as if they did not evade taxes. This still shows that substantial evasion is likely to be present.

5.2 Heterogeneous Effects

The double list experiment was included as part of the WBES which collects extensive information on firm characteristics and the beliefs of top managers or owners, which means that there are many dimensions along which heterogeneity in the main findings could be explored. We rely on a machine learning algorithm to identify the dimensions where substantial and consistent heterogeneity exists. Specifically, we use a causal forest technique[10] to identify where the greatest heterogeneity occurs across 27 potential dimensions that are captured in the survey prior to the double list experiment questions. A list of these 27 dimensions together with a measure of their relative heterogeneity in tax evasion prevalence is presented in Table A3 in the Appendix. We use this inductive approach to identify heterogeneity as ex ante, it is not immediately clear which dimensions will exhibit the most significant variation in tax evasion. Importantly, we focus our analysis on the dimensions where substantial and consistent heterogeneity exists in both list experiments. There are instances where differences across some dimensions are meaningful in only one of the list experiments and where the direction of heterogeneity varies. As we are interested in the most robust results, we do not focus on these dimensions where heterogeneity is not substantial or consistent across

[10]The causal forest function in the R package grf.

both list experiments.

Table 3

Table 3 shows the three dimensions where substantial and consistent heterogeneity in reported tax evasion rates is most notable. Firstly, firms that do not export are more likely to report evading tax than those that export across both list experiments. The reported tax evasion rate varies from 26.0 to 30.7 percent for non-exporting firms compared to 16.9 to 21.5 percent for exporting firms. While the exact level varies, the order of magnitude of tax evasion is much higher for non-exporting firms. Secondly, firms which report that tax administration is a major obstacle to business activities are also found to have significantly higher levels of tax evasion. The tax evasion rate varies from 31.9 to 38.1 percent for firms that claim tax administration is a major obstacle compared to 20.1 to 24.3 percent for those that do not. That equates to an evasion rate of up to 90 percent higher for firms where tax administration is considered a major obstacle to business activities. Thirdly, firms facing substantial competition from the informal sector were more likely to report evading taxes. The reported tax evasion rate varies from 28.5 to 35.4 percent for firms that face substantial competition from the informal sector compared to 21.5 to 23.3 percent for those that do not. While some of these differences are not statistically significant at traditional levels, partly due to sample size constraints, the differences are still economically meaningful.

The absence of substantial and consistent heterogeneity of reported tax evasion on other dimensions is striking. The survey covered a range of aspects of firm characteristics and beliefs, such as industry sector, whether firms had been visited by tax officials, firm size, ownership structure, and where they used an auditor. However, based on these characteristics, there was no significant and consistent variation in tax evasion across the two list experiments. For example, firms with an owner-manager reported a tax evasion rate of 24.4 to 28.9 percent compared to 20.4 to 23.6 percent for those that did not (these differences

were not statistically significant).

5.3 Estimate of revenue loss due to tax evasion by registered firms

We produce a back-of-the-envelope estimate of the revenue lost due to tax evasion by registered firms in Indonesia by combining survey results about the share of firms evading tax (extensive margin) and the extent to which these firms avoid paying tax (intensive margin). The results of the double list experiment estimate the former. The latter is captured in a separate question where respondents are asked what percentage of their tax obligations non-compliant firms pay. The extensive margin results of 25-27 percent are outlined throughout this paper, while the intensive margin results are presented in the Appendix. On average, respondents estimate that non-compliant firms only pay around 23 percent of the taxes they owe. These figures are applied to the existing share of GDP currently collected by the central government in business income taxes and VAT from firms in Indonesia, which is around 8 percent. This results in an estimate of around two percentage points of GDP lost in revenue due to tax evasion by registered firms (the complete working for this estimate is presented in the Appendix).

6 Discussion and Conclusion

This study shows that a sizable share of registered firms in Indonesia admit to not paying all the tax they owe, and this is likely to substantially reduce the total amount of business income tax and VAT collected. The double list experiment provides rigorous, internally consistent results illustrating that around one-quarter of firms report evading taxes. A back-of-the-envelope calculation based on this result suggests that revenue lost from tax evasion by registered firms is around two percent of GDP. This is the most common among firms that do not export, face substantial competition from the informal sector, and believe tax administration is a major obstacle to their business activities. This is based on self-reported

tax evasion, so it likely represents a lower bound.

These findings provide policy makers with a much clearer picture of just how widespread tax evasion is by registered firms in Indonesia. To date, there have not been precise estimates of the underlying level of tax evasion as only limited audit data exists, and surveys asking directly about tax compliance and/or tax morale can have substantial issues. Further, while analyses based on tax administrative data can provide valuable insights about the tax-paying behavior of firms, they do not necessarily shed light on intentional tax evasion in the same way that a double list experiment can. Arguably, the findings we present, which are based on nationally representative data and are internally consistent, provide among the best estimates of the levels of tax evasion by firms in Indonesia. Existing studies using single-list experiments to estimate tax evasion have typically focused on individuals as opposed to firms, and in these instances, reported tax evasion has been lower than what we observe (e.g., see Genest-Grégoire et al., 2022 and Iraman et al., 2022). At a minimum, our findings suggest substantial increases in tax revenue are feasible if the government could increase the compliance of registered firms.

The double list experiment also provides unique insights for policymakers into the types of firms that are more likely to be non-compliant,[11] by extensively exploring heterogeneity and identifying three key dimensions. Firstly, there are higher rates of tax evasion among firms that do not export. A potential explanation for this heterogeneity is that exporting firms are more likely to have a paper trail of their economic activity (e.g., from custom licenses), which means that the volumes of sales of these firms are easier to monitor. In contrast, firms that do not export may perceive that the revenue authority has less third-party information about their business activities. Consequently, they are more willing to evade taxes. Secondly, firms that find tax administration to be a major obstacle to their business activities are much more likely to admit to evading tax. This suggests efforts to minimize the complexity and,

[11]The optimal targeting of enforcement activities would extend beyond just evasion rates, including the ease of detecting/proofing evasion, expected recovery rate on outstanding tax due, and deterrence value of enforcement.

consequently, the tax system's burden on firms may also increase compliance. Thirdly, firms that face substantial competition from the informal sector are more likely to report higher levels of tax evasion. This is likely because formal firms in competition with informal firms have a disincentive to comply, as this will lower their competitiveness in the market (such as by lowering their profit margins). Interestingly, this result suggests that there may be a positive externality from registering informal firms if this leads to formal firms being more likely to comply. Notably, there were limited differences in tax evasion rates across other dimensions. As such, any efforts by tax collectors to reduce the rate of tax evasion by most types of firms are likely to raise revenue.

This study has also shown the value of embedding double-list experiments in large-scale surveys. While single-list experiments have grown popular over time, there has also been increasing recognition of some of their shortcomings, particularly regarding the possibility of specific non-sensitive items in a list influencing the likelihood of respondents also counting the sensitive item. Double-list experiments verify if this issue is present and can help illustrate the internal consistency of findings. Our examination of heterogeneity across 27 dimensions with a large sample size is also an important learning exercise as it shows the value of combining machine learning algorithms with indirect solicitation techniques. However, the double-list experiment approach also raises important questions about the best way to report heterogeneity in instances that are not substantial and/or consistent across both list experiments. At a minimum, our study shows how, at a low marginal cost, a double list experiment can provide much more credible insights about the prevalence of a sensitive behavior than a single list experiment and this approach can identify the dimensions of heterogeneity in the prevalence of a sensitive behavior.

Future research on this topic could take several directions. Firstly, additional research could be conducted measuring levels of tax evasion in Indonesia to validate our findings by analyzing tax administrative data with third-party information and/or survey data that directly captures tax evasion and tax morale. In addition, there may be value in exploring

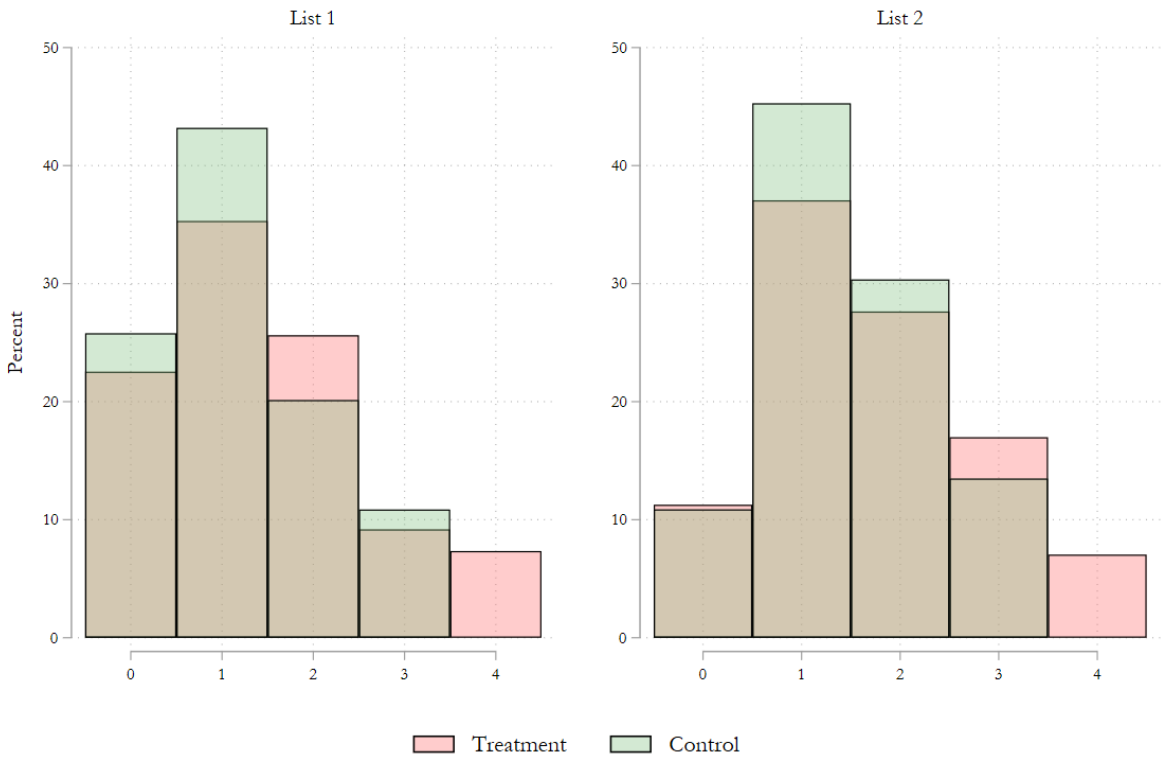
more subtle forms of tax evasion, such as when firms artificially lower their profits by exaggerating their expenditure. Secondly, this approach of using a double list experiment to measure tax evasion could be trialed in other countries to test the generalizability of this methodology. Thirdly, while this study has provided valuable insights about "what" the levels of tax evasion are, far more analysis is required to capture "how" to stop firms from evading taxes in middle-income countries with relatively low revenue levels as a share of GDP.

Table and Figures

Table 1: Design of Experiment

	Group A	Group B
List 1	a) This establishment had to let go of an employee over the last year b) This establishment's last month sales increased by 200% c) The establishment was temporarily closed during the COVID-19 pandemic	a) This establishment had to let go of an employee over the last year b) This establishment's last month sales increased by 200% c) The establishment was temporarily closed during the COVID-19 pandemic **d) This establishment does not pay all the taxes it is required to pay**
List 2	a) This establishment almost went bankrupt in the last year b) At least one of the employees of this establishment contracted COVID-19 since the start of the pandemic c) The price of the main product of this establishment changed over the past year **d) This establishment does not pay all the taxes it is required to pay**	a) This establishment almost went bankrupt in the last year b) At least one of the employees of this establishment contracted COVID-19 since the start of the pandemic c) The price of the main product of this establishment changed over the past year

Figure 1: Distribution of Responses



Note: This figure shows the number of items selected by respondents in the treatment and control groups in the first and second list experiments. Group B received the treatment in List 1, and Group A received the treatment in List 2.

Table 2: Main Results

	Control	Treatment	Difference	SE	N
List 1	1.161	1.435	0.274***	0.044	2330
List 2	1.465	1.715	0.250***	0.041	2322
List 1+2	1.316	1.571	0.255***	0.030	4652

Note: The first two rows of this table show the average number of true statements reported by respondents in groups A and B, for the first and second list experiments, as well as the differences between these two groups. The third row of the table shows pools both experiments together. Standard errors are presented in parentheses. *** corresponds with a p-value below 0.01.

Table 3: Heterogeneity of tax evasion by firm characteristics

	List 1		List 2	
	(1)	(2)	(3)	(4)
Full Sample	0.274***	0.268***	0.250***	0.248***
	(0.044)	(.043)	(.041)	(.040)
Exporters	0.174**	0.169**	0.215***	0.213***
	(0.085)	(0.084)	(0.079)	(0.079)
Non-exporters	0.307***	0.301***	0.261***	0.260***
	(0.051)	(0.050)	(0.047)	(0.047)
P-value difference	0.177	0.174	0.623	0.611
Tax administration obstacle	0.345***	0.319***	0.369***	0.381***
	(0.082)	(0.081)	(0.077)	(0.076)
Tax administration NOT an obstacle	0.243***	0.243***	0.207***	0.201***
	(0.051)	(0.050)	(0.048)	(0.047)
P-value difference	0.290	0.422	0.073	0.045
Informal Competition	0.354***	0.332***	0.285***	0.291***
	(0.068)	(0.068)	(0.064)	(0.064)
No informal competition	0.215***	0.225***	0.233***	0.219***
	(0.056)	(0.056)	(0.052)	(0.052)
P-value difference	0.117	0.224	0.529	0.382
Sample Size	2330	2330	2322	2322
Controls	No	Yes	No	Yes

Note: This table shows differences in the average number of items selected by respondents in the treatment and control groups in the first and second list experiments (with and without control variables). Specifically, the results of the list experiments are presented for subgroups of respondents based on market competition (competition from the informal sector or not), target market (exporter or not), and if respondents believe tax administration is an obstacle to their business.

References

Ahlquist, J. S. (2018). List Experiment Design, Non-Strategic Respondent Error, and Item Count Technique Estimators. *Political Analysis*, *26*(1), 34–53. https://doi.org/10.1017/pan.2017.31

Ali, M., Fjeldstad, O.-H., & Sjursen, I. H. (2014). To Pay or Not to Pay? Citizens' Attitudes Toward Taxation in Kenya, Tanzania, Uganda, and South Africa. *World Development*, *64*, 828–842. https://doi.org/10.1016/j.worlddev.2014.07.006

Alm, J. (2019). Can Indonesia reform its tax system? Problems and Options. *Tulane Economics Working Paper Series*. https://doi.org/https://EconPapers.repec.org/RePEc:tul:wpaper:1906

Artavanis, N., Morse, A., & Tsoutsoura, M. (2016). Measuring Income Tax Evasion Using Bank Credit: Evidence from Greece. *The Quarterly Journal of Economics*, *131*(2), 739–798. https://doi.org/10.1093/qje/qjw009

Basri, M. C., Felix, M., Hanna, R., & Olken, B. A. (2021). Tax Administration versus Tax Rates: Evidence from Corporate Taxation in Indonesia. *American Economic Review*, *111*(12), 3827–3871. https://doi.org/10.1257/aer.20201237

Blair, G., & Imai, K. (2012). Statistical Analysis of List Experiments. *Political Analysis*, *20*(1), 47–77. https://doi.org/https://doi.org/10.1093/pan/mpr048

Blair, G., Imai, K., & Lyall, J. (2014). Comparing and Combining List and Endorsement Experiments: Evidence from Afghanistan. *American Journal of Political Science*, *58*(4), 1043–1063. https://doi.org/https://doi.org/10.1111/ajps.12086

Breuer, M. L. E., Guajardo, M. J., & Kinda, M. T. (2018, August). *Realizing Indonesia's Economic Potential*. International Monetary Fund.

Carrillo, P., Pomeranz, D., & Singhal, M. (2017). Dodging the Taxman: Firm Misreporting and Limits to Tax Enforcement. *American Economic Journal: Applied Economics*, *9*(2), 144–164. https://doi.org/10.1257/app.20140495

Chuang, E., Dupas, P., Huillery, E., & Seban, J. (2021). Sex, lies, and measurement: Consistency tests for indirect response survey methods. *Journal of Development Economics*, *148*, 102582. https://doi.org/https://doi.org/10.1016/j.jdeveco.2020.102582

Dom, R., Custers, A., Davenport, S., & Prichard, W. (2022, January). *Innovations in Tax Compliance: Building Trust, Navigating Politics, and Tailoring Reform*. World Bank Publications.

Droitcour, J., Caspar, R. A., Hubbard, M. L., Parsley, T. L., Visscher, W., & Ezzati, T. M. (2004, August). The Item Count Technique as a Method of Indirect Questioning: A Review of Its Development and a Case Study Application. In *Measurement Errors in Surveys* (pp. 185–210). https://doi.org/10.1002/9781118150382.ch11

Genest-Grégoire, A., Godbout, L., & Guay, J.-H. (2022). Lists: A Novel Experimental Method to Measure Tax Evasion [Publisher: The University of Chicago Press]. *National Tax Journal*, *75*(3), 517–537. https://doi.org/10.1086/719469

Glynn, A. (2013). What Can We Learn with Statistical Truth Serum? Design and Analysis of the List Experiment. *Public Opinion Quarterly*, *77*(S1), 159–172. https://doi.org/https://doi.org/10.1093/poq/nfs070

Heide-Jørgensen, T. (2022). Does Social Desirability Bias Distort Survey Analyses of Ideology and Self-Interest? Evidence from a List Experiment on Progressive Taxation. *Public Opinion Quarterly*, *86*(4), 899–912. https://doi.org/10.1093/poq/nfac050

Hoy, C. (2022). How does the progressivity of taxes and government transfers impact people's willingness to pay tax? Experimental evidence across developing countries. *World Bank Policy Research Working Papers*, *10167*. https://doi.org/https://doi.org/10.1596/1813-9450-10167

Hoy, C., McKenzie, L., & Sinning, M. (2024). Improving Tax Compliance without Increasing Revenue: Evidence from Population-Wide Randomized Controlled Trials in Papua New Guinea [Publisher: The University of Chicago Press]. *Economic Development and Cultural Change*, *72*(2), 691–723. https://doi.org/10.1086/721650

Ikhsan, M., Trialdi, L., & Syahrial, S. (2005). Indonesia's new tax reform: Potential and direction. *Journal of Asian Economics*, *16*(6), 1029–1046. https://doi.org/10.1016/j.asieco.2005.10.004

Imai, K. (2011). Multivariate Regression Analysis for the Item Count Technique [Publisher: Taylor & Francis _eprint: https://doi.org/10.1198/jasa.2011.ap10415]. *Journal of the American Statistical Association*, *106*(494), 407–416. https://doi.org/10.1198/jasa.2011.ap10415

Indupurnahayu, I., & Walujadi, D. (2019). Shadow Economy and Loss of Potential Taxes in Indonesia: Empirical Study 1990-2018. *Opción*, *35*, 872–894. Retrieved May 22, 2024, from https://produccioncientificaluz.org/index.php/opcion/article/view/31628

Iraman, E., Ono, Y., & Kakinaka, M. (2022). Tax compliance and social desirability bias of taxpayers: Experimental evidence from Indonesia. *Journal of Public Policy*, *42*(1), 92–109. https://doi.org/10.1017/S0143814X21000040

Karlan, D. S., & Zinman, J. (2012). List randomization for sensitive behavior: An application for measuring use of loan proceeds. *Journal of Development Economics*, *98*(1), 71–75. https://doi.org/10.1016/j.jdeveco.2011.08.006

Luttmer, E. F. P., & Singhal, M. (2014). Tax Morale. *Journal of Economic Perspectives*, *28*(4), 149–168. https://doi.org/10.1257/jep.28.4.149

Mascagni, G. (2017). From The Lab To The Field: A Review Of Tax Experiments. *Journal of Economic Surveys*, *32*(2), 273–301. https://doi.org/https://doi.org/10.1111/joes.12201

Mascagni, G., Dom, R., Santoro, F., & Mukama, D. (2023). The VAT in practice: Equity, enforcement, and complexity. *International Tax and Public Finance*, *30*(2), 525–563. https://doi.org/10.1007/s10797-022-09743-z

McKenzie, D., & Siegel, M. (2013). Eliciting illegal migration rates through list randomization. *Migration Studies*, *1*(3), 276–291. https://doi.org/10.1093/migration/mnt018

Miller, J. D. (1984). *A New Survey Technique for Studying Deviant Behavior* [Doctoral dissertation, George Washington University]. Retrieved May 22, 2024, from https://www.proquest.com/openview/2254c7ccdade4a691e417bd5734aa874/1?cbl=18750&diss=y&pq-origsite=gscholar&parentSessionId=fxQe4GJ6QOmCLsfOJsoaL7QlsKkT7Yqud81VlxS808A%3D

Mulyawan, A. W. (2017). Analisis underground economy Indonesia dan potensi penerimaan pajak [Number: 1]. *Info Artha*, *1*(1), 60–78. https://doi.org/10.31092/jia.v1i1.18

Nizar, M. A., & Purnomo, K. (2011). Potensi penerimaan pajak. *Kajian Ekonomi dan Keuangan*, *15*(2). https://doi.org/https://doi.org/10.31685/kek.v15i2.93

Pomeranz, D. (2015). No Taxation without Information: Deterrence and Self-Enforcement in the Value Added Tax. *American Economic Review*, *105*(8), 2539–2569. https://doi.org/10.1257/aer.20130393

Raghavarao, D., & Federer, W. T. (1979). Block Total Response as an Alternative to the Randomized Response Method in Surveys. *Journal of the Royal Statistical Society: Series B (Methodological)*, *41*(1), 40–45. https://doi.org/10.1111/j.2517-6161.1979.tb01055.x

Ramadhan, F. M. (2019). Estimation The Size Of Underground Economy And Tax Potential Loss In Indonesia From 2000 – 2017 [Number: 2]. *Jurnal Ilmiah Mahasiswa FEB*, *7*(2). Retrieved May 22, 2024, from https://jimfeb.ub.ac.id/index.php/jimfeb/article/view/5855

Rosenfeld, B., Imai, K., & Shapiro, J. N. (2016). An Empirical Validation Study of Popular Survey Methodologies for Sensitive Questions. *American Journal of Political Science*, *60*(3), 783–802. https://doi.org/10.1111/ajps.12205

Rothenberg, A. D., Gaduh, A., Burger, N. E., Chazali, C., Tjandraningsih, I., Radikun, R., Sutera, C., & Weilant, S. (2016). Rethinking Indonesia's Informal Sector. *World Development*, *80*, 96–113. https://doi.org/10.1016/j.worlddev.2015.11.005

Slemrod, J. (2008). Does It Matter Who Writes the Check to the Government? The Economics of Tax Remittance [Publisher: The University of Chicago Press]. *National Tax Journal*, *61*(2), 251–275. https://doi.org/10.17310/ntj.2008.2.05

Slemrod, J., & Weber, C. (2012). Evidence of the invisible: Toward a credibility revolution in the empirical analysis of tax evasion and the informal economy. *International Tax and Public Finance*, *19*(1), 25–53. https://doi.org/10.1007/s10797-011-9181-0

Suharnoko, B., Castaneda Nunez, J. L., Chairunissa, T. F., Farid, M. N., Poerwanto, J., & Wirapati, B. A. (2020). *Applying Behavioral Insights to Promote SME Tax Compliance in Indonesia : Activity Report* (Text/HTML No. 157456). The World Bank. Retrieved May 22, 2024, from https://documents.worldbank.org/en/publication/documents-reports/documentdetail/448871615957692506/Applying-Behavioral-Insights-to-Promote-SME-Tax-Compliance-in-Indonesia-Activity-Report

Tatariyanto, F. (2014). Taxing the Underground Economy: The Case of Indonesia. *Journal of Economics and Sustainable Development*, *5*(27), 236–250. https://doi.org/https://api.semanticscholar.org/CorpusID:55248921

Timofte (Coca), C., Socoliuc, M., & Grosu, V. (2019). Measuring tax evasion. Conceptual approaches. *The USV Annals of Economics and Public Administration*, *19*(1(29)), 116–121. Retrieved May 22, 2024, from http://annals.feaa.usv.ro/index.php/annals/article/view/1149

Tørsløv, T., Wier, L., & Zucman, G. (2023). The Missing Profits of Nations. *The Review of Economic Studies*, *90*(3), 1499–1534. https://doi.org/10.1093/restud/rdac049

Waseem, M. (2022). The Role of Withholding in the Self-Enforcement of a Value-Added Tax: Evidence from Pakistan. *The Review of Economics and Statistics*, *104*(2), 336–354. https://doi.org/https://doi.org/10.1162/rest_a_00959

Waseem, M. (2023). Overclaimed refunds, undeclared sales, and invoice mills: Nature and extent of noncompliance in a value-added tax. *Journal of Public Economics*, *218*, 104783. https://doi.org/10.1016/j.jpubeco.2022.104783

Yang, J., & Moerenhout, T. (2024). Information campaigns and public perceptions of structural reforms: Evidence from a survey experiment on gasoline subsidy reform in Nigeria. *Journal of Policy Analysis and Management*, *42*(2), 509–529. https://doi.org/https://doi.org/10.1002/pam.22561

Appendix

Table A1: Balancing Table

	Group A		Group B		(A)-(B)	
Variable	N	Mean	N	Mean	N	Mean difference
Sole Proprietorship	1405	0.601	1550	0.554	2955	0.047***
		(0.013)		(0.013)		
Less than 10 years of age	1405	0.273	1550	0.270	2955	0.002
		(0.012)		(0.011)		
Manufacturing	1405	0.470	1550	0.468	2955	0.002
		(0.013)		(0.013)		
Facing informal competition	1405	0.363	1550	0.378	2955	-0.015
		(0.013)		(0.012)		
Retail/Wholesale	1405	0.220	1550	0.205	2955	0.015
		(0.011)		(0.010)		
Services	1405	0.323	1550	0.308	2955	0.015
		(0.012)		(0.012)		
Local Market (Main)	1405	0.683	1550	0.657	2955	0.026
		(0.012)		(0.012)		
Informality is an obstacle	1405	0.373	1550	0.358	2955	0.015
		(0.013)		(0.012)		
Has a bank account	1405	0.723	1550	0.709	2955	0.014
		(0.012)		(0.012)		
External audit	1405	0.180	1550	0.187	2955	-0.007
		(0.010)		(0.010)		
Inspected by tax authority	1405	0.130	1550	0.144	2955	-0.014
		(0.009)		(0.009)		
Subject to CIT	1405	0.532	1550	0.535	2955	-0.003
		(0.013)		(0.013)		
Subject to VAT	1405	0.496	1550	0.517	2955	-0.021
		(0.013)		(0.013)		
Government Contract	1405	0.094	1550	0.105	2955	-0.011
		(0.008)		(0.008)		
Large firm	1405	0.063	1550	0.056	2955	0.007
		(0.007)		(0.006)		
F-test of joint significance (F-stat)						1.382
F-test, number of observations						2955

Note: This table shows the average share of firms reported to have these baseline characteristics in Groups A and B and the differences between the groups. Standard errors are presented in square parenthesis. *** corresponds with a p-value ¡ 0.01.

Table A2: Main Results (weighted using a paired test)

	Control	Treatment	Difference	SE	N
List 1	0.994	1.342	0.348***	0.091	2272
List 2	1.559	1.772	0.213**	0.095	2272
List 1+2	1.268	1.564	0.296***	0.065	2272

Note: The first two rows of this table show the average number of items that were selected by respondents in Groups A and B in the first and second list experiments and the differences between the groups. The third row of the table shows the average number of items selected by respondents in Groups A and B across both list experiments. Standard errors are presented in parenthesis. *** corresponds with a p-value below 0.01. The weights employed are those that ensure the WBES's representativeness in Indonesia across firm size, sector, and region.

Table A3: Degree of Variation in Heterogeneous Treatment Effects

Dimension	List 1	List 2
Sole proprietor	0.028	0.030
Local sales	0.023	0.025
Small	0.092	0.022
Medium	0.035	0.038
Large	0.016	0.006
Less than 10yrs	0.031	0.067
Manager less than 10yrs	0.030	0.028
Manufacturing	0.025	0.033
Retail/Wholesale trade	0.030	0.023
Sales under100mil	0.035	0.034
Sales 100-500mil	0.034	0.025
Sales over500mil	0.020	0.027
Services	0.053	0.047
Informal competition	0.035	0.067
Uses auditor	0.027	0.025
Bank account	0.044	0.028
Visited by DGT	0.077	0.027
Government contract	0.067	0.016
Tax rates obstacle	0.041	0.071
Tax admin obstacle	0.042	0.097
Located in Java	0.028	0.078
Owner manager	0.023	0.018
Female manager	0.032	0.055
Foreign-owned	0.004	0.001
Solely Domestic Sales	0.059	0.031
Believes DGT is fair	0.024	0.045
Makes informal payments	0.045	0.034

Note: The output in this table is based on the causal forest function within the Generalized Random Forest R package. The degree of heterogeneity in reported tax evasion is calculated for each dimension, and the relative degree of variation in heterogeneity for each dimension is placed on a scale from 0 to 1 (i.e. so that the total across all dimensions is equal to 1). This exercise is completed independently for the two list experiments.

Details about estimate of revenue loss due to tax evasion by registered firms

Following the double list experiment, respondents were asked "It is said that establishments sometimes are not fully compliant with their tax liabilities. On average, what percentage of each of the following taxes it owes, do establishments like this one pay?" Respondents were asked to provide an answer on a scale from 0 to 100% for VAT and CIT separately. This question has been used extensively across WBES in other countries. On average, respondents estimate that non-compliant firms only pay around 23 percent of the VAT and CIT that they owe. We combine survey results about the share of firms evading tax (extensive margin) and the extent to which these firms avoid paying tax (intensive margin) to make the following calculation:

Revenue loss=% of firms evading tax $\times$ Mean amount of tax not paid by evading firms

The results of the double list experiment provide an estimate of the share of firms evading tax. This is around 26 percent. The results of the intensive margin question provide a way to estimate the average amount of tax not paid by evading firms. Specifically, respondents estimate that on average non-compliant firms only pay around 23 percent of the taxes they owe. As such, the average amount of tax not paid by evading firms is 77 percent. This would suggest that another 20 percent ($0.26 \times 0.77=0.2$) of tax would be collected if tax evasion by registered firms with five or more employees was eliminated.

This estimate is, at best, only indicative and should be interpreted cautiously for many reasons. Firstly, while the extensive margin estimate of tax evasion is based on the double list experiment, the intensive margin estimate is potentially less reliable as it is based on respondents' educated guess about the behavior of other firms. Secondly, the overall estimate is based on assuming a similar amount of tax would be paid by firms that are currently non-compliant as those that are compliant, which may not be the case. Thirdly, the sample of firms in the survey is representative of registered firms, with at least five employees, in most manufacturing and services sub-sectors. However, they do not represent the exact distribution of all business income tax and VAT payers in Indonesia. Fourthly, there is a risk that firms will shift from formal to informal operations due to enforcement activities, undermining the total amount of tax that could be collected from formal firms. Finally, we do not consider how tax evasion may vary by firm size, as we do not observe any differences in the double list experiments based on this dimension.