*The World Bank Economic Review*, 38(3), 2024, 625–646
https://doi.org10.1093/wber/lhad030
Article

OXFORD

# Addressing Social Desirability Bias When Measuring Child Labor Use: An Application to Cocoa Farms in Côte d'Ivoire

**Marine Jouvin**

## Abstract

This paper proposes new estimates of child labor use in Côte d'Ivoire's cocoa farms that are certified free of child labor. The study relies on list experiments (LE) to overcome social desirability bias associated with measuring sensitive issues, implemented on a sample of 4,458 Ivorian cocoa farmers. Findings show that 24 percent of them were helped by at least one child under 16 for harvesting and breaking the cocoa pods during the past 12 months, 21 percent for preparing their farm, and 25 percent employed and paid at least one child to perform any task on their cocoa farm. These results are twice as high as those declared by farmers when directly questioning them about their child labor use. This study provides evidence that the LE method, while more cognitively demanding than the direct questioning method, can be successfully understood by populations with low levels of education. Findings further show that, in this setting, the LE estimates are robust to specific LE design changes.

**JEL classification:** C83, J23, J43, J81

**Keywords:** list experiment, social desirability bias, child labor, certification schemes

## 1. Introduction

Recent evidence based on household surveys suggests that the Ivorian cocoa sector is still struggling with child labor, with an estimated 790,000 children between 5 and 17 years laboring on cocoa farms in 2019 (Sadhu et al. 2020). The cocoa industry thus remains under pressure from national and international regulators, as children working on cocoa farms are widely exposed to hazardous work, one of the worst forms of child labor according to the International Labour Organization (ILO). In Côte d'Ivoire, 97 percent of children working on cocoa farms are thus estimated to be involved in such hazardous child labor (Sadhu et al. 2020). The ILO defines the latter as "work which, by its nature or the circumstances in which it is carried out, is likely to harm the health, safety or morals of children".[1] Pressure from various

Marine Jouvin is a PhD candidate at the Bordeaux School of Economics, University of Bordeaux, Bordeaux, France; her email address is marine.jouvin@u-bordeaux.fr. The author thanks Touton S.A. for the valuable partnership that made this project possible, and gives special thanks to the enumerators in Côte d'Ivoire for their outstanding data collection work for this study. The author is grateful to Tanguy Bernard and Delphine Boutin for their guidance and support, as well as to the editor, Eric Edmonds, and anonymous referees for valuable comments that have greatly improved the manuscript. The author also appreciates the helpful comments of conference participants at the 2021 International Conference in Development Economics. A supplementary online appendix is available with this article at *The World Bank Economic Review* website.

1 Article 3 of ILO's Convention No. 182.

civil society stakeholders to reduce child labor in cocoa has in part been fueled by a growing body of research related to the adverse effects of child labor on children's physical and mental health (Guarcello et al. 2004; Fekadu, Alem, and Hägglöf 2006; Roggero et al. 2007; Trinh 2020) or education (Udry 2006; Beegle, Dehejia, and Gatti 2009; Putnick and Bornstein 2015) since the early 2000s.[2]

In the late 2000s, traders and manufacturers began to use independent certification labels to meet growing consumer demand for ethical and sustainable cocoa.[3] These labels reward cocoa farmers with the farmgate price and a certification premium, the latter rewarding a more sustainable cocoa production, without deforestation and free of child labor.[4] Farmers that are members of certified cooperatives benefit from various trainings and are sensitized to the negative consequences of child labor, especially hazardous tasks, on children's well-being. Third-party auditors annually audit certified cooperatives to ensure that their affiliated farmers comply with certification programs' social and environmental requirements. A farmer who does not respect these requirements risks being excluded from the certified cocoa value chains, which implies losing all the benefits of certification, including the certification premium. Certification schemes and the services they provide to farmers have gradually attracted the attention of researchers in the past few years, mainly focusing on assessing the impact of certification on farmers' incomes. As far as we know, no studies have investigated whether child labor use remains prevalent in the specific context of certified cocoa farms.

Investigating farmers' use of child labor on certified cocoa farms implies addressing methodological challenges to correctly measure this complex and sensitive phenomenon. Guarcello et al. (2012) highlight that child labor prevalence fluctuates significantly from one year to another in various countries, calling into question the way in which it is measured. The subsequent literature on the measurement of adult and child labor has identified key factors that may cause labor measurement errors, although evidence remains scarce for Côte d'Ivoire.

The choice of the respondent to be asked about child labor is an important consideration. Proxy respondents, such as household heads, generally under-report when asked about child labor issues compared to self-report from children (Dillon 2010; Bardasi et al. 2011; Galdo, Dammert, and Abebaw 2018; Janzen 2018). Findings from Galdo, Dammert, and Abebaw (2020) also suggest that discrepancies between child and proxy respondent responses might depend on the child's gender. However, whether children can provide accurate answers will vary depending on their age and cognitive development (Borgers et al. 2000; Fuchs 2009). While children aged 16 and over can be considered adults in their survey responses, younger children are subject to cognitive difficulties that differ according to their age group (Borgers et al.). The recall period over which respondents are asked about labor issues is also likely to generate erroneous answers, as it can be difficult for them to recall accurately and over long periods the amount of labor used, or even their own working time. Several studies have confirmed the existence of such recall bias in labor surveys in the African context, suggesting that the choice of recall period sometimes leads to underestimating or overestimating labor use or working time, depending on the context (Beegle, Carletto, and Himelein 2012; Arthi et al. 2018; Gaddis et al. 2021). Bardasi et al. (2011) further highlight that labor-related estimates are sensitive to the length of the module covering labor issues, thus suggesting that survey design is also a factor that may influence labor-related estimates.

It is also likely that, depending on the type of child labor concerned, national legislation, and social standards, the measurement of child-labor-related behaviors may be prone to social desirability bias.

2 See Boutin and Jouvin (2022) for a comprehensive review.

3 UTZ and Rainforest Alliance (RA) for sustainable development labels, Fairtrade International (Max Havelaar) for fair trade label, or Organic for cocoa that follows organic production methods. All sustainable development programs set up by chocolate producers themselves (e.g., Cocoa Life for Mondelez) are excluded here.

4 In Côte d'Ivoire, the cocoa farmgate price is set at the beginning of each season by the Conseil Café Cacao (CCC). During the survey period (2019–2020 cocoa season), the farmgate price per kilo of cocoa was set at 825 CFA (1.48 dollars).

This means that respondents may not honestly answer child labor questions to provide socially desirable answers, and avoid negative legal or social consequences (Tourangeau and Yan 2007). This behavior is common when respondents are questioned about sensitive topics, both from a personal or societal perspective (Tourangeau and Yan 2007), such as politics (Çarkoğlu and Aytaç 2015), health (De Cao and Lutz 2018; Moseson et al. 2015), or intimate partner violence (Cullen 2020).

Attempting to detect cocoa households relying on child labor on their farms, the cocoa industry is increasingly implementing child labor monitoring systems among farmers. These systems, which aim at identifying and addressing cases of child labor, usually rely on surveying cocoa households specifically on their use of child labor, in a dedicated child labor questionnaire that focuses on both light and hazardous work, and uses direct questioning methods.[5] We argue that a social desirability bias is likely at play when cocoa farmers from certified cooperatives are asked about their use of child labor, given that it is strictly prohibited and that they could suffer negative economic repercussions, by being excluded from certified cocoa value chains. This bias may be important since the questionnaires used so far (a) are usually conducted as part of a CLMRS, i.e., a system specifically designed to identify and address child labor, (b) ask questions exclusively or almost exclusively about child labor, (c) use direct questioning methods which explicitly signals to the respondent that the focus is on child labor issues.

This social desirability bias has not received much attention in the measurement of child labor issues, and has rarely been considered in reports assessing the prevalence of child labor use in the mainstream or certified cocoa industry. Yet estimates of child labor use or prevalence obtained through a direct questioning method could significantly underestimate the phenomenon in the case of a proven social desirability bias.

This paper relies on the list experiment (LE) approach, an indirect questioning method, to measure the proportion of certified cocoa farmers in Côte d'Ivoire who use child labor to perform hazardous work on their farms. By hiding the respondent's answer to the sensitive question among other answers to innocuous questions, this method is expected to reduce social desirability bias (Imai 2011). This approach is used to measure farmers' use of children under the age of 16 on their cocoa farms. Specifically, respondents are asked whether a child helped them with harvesting and pod breaking, preparing the farm, or if they had employed and paid a child for any work on their cocoa farm. Results from the LEs are compared to those obtained from directly questioning farmers about their use of child labor for hazardous tasks, as done in existing surveys aiming at measuring child labor use. This study covers 4,458 cocoa farmers certified by the UTZ and RA labels, surveyed between December 2019 and March 2020. Two additional survey waves conducted in 2022 are used to test the robustness of the LE estimates to design changes, and farmers' levels of understanding of the LE method.

Findings show that the prevalence of child labor use estimated using the indirect measure is twice as large as that obtained from direct questioning, indicating that cocoa farmers likely exhibit social desirability bias when asked about their use of child labor. Using LEs, this paper provides evidence that 24 percent of certified cocoa farmers were helped by a child under 16 for harvesting and breaking cocoa pods during the past 12 months, 21 percent for preparing their farm, and 25 percent employed and paid a child to perform any task on their cocoa farm. Findings further show that the LE method is well understood by cocoa farmers, despite their low levels of education. There is no evidence of a better understanding of LE among educated farmers compared to those with no education. Last, this paper suggests that LE estimates are robust to design changes in this setting. Specifically, light modifications to the control items, as well as reduction of the recall period, have a negative average effect on the number of items to which farmers respond "yes", but in a uniform way between control and treated farmers, thus not affecting the difference in means between the two groups.

5 Usually called the Child Labour Monitoring and Remediation Systems (CLMRS) survey.

This paper contributes to the literature on child labor in two ways. First, unlike most studies that attempt to measure child-labor-related behavior, an indirect method of questioning is used to reduce the social desirability bias among respondents and obtain a more accurate measure of child labor use. Findings support the existence of a social desirability bias and question the relevance of using only the direct questioning method in this context, which can contribute to the design of poorly targeted policies. Accordingly, surveys used by the cocoa industry, such as CLMRS surveys, likely underestimate the magnitude of child labor and farmers' reliance on children to work on cocoa farms. More broadly, the use of an indirect questioning method, such as the list experiment, should be considered along with direct questioning when individuals are asked about child labor, and have incentives to hide the truth. Second, this study focuses on measuring hazardous child labor use on certified cocoa farms. These findings contribute to the literature on the use of certification schemes in agricultural supply chains, by providing evidence that certification bodies are failing to keep Ivorian cocoa farms free of child labor, despite their commitment. It raises concerns about the effectiveness of certification labels in reducing child labor in the cocoa sector, as well as the effectiveness of currently implemented monitoring systems.

This study also contributes to the list experiment literature, suggesting that this indirect questioning method can be used to measure the use of child labor in particularly sensitive contexts. Although considered more cognitively demanding than the direct questioning method, this paper shows that cocoa farmers with low levels of education can successfully understand the method. Results from the robustness test also contribute to the discussion regarding the LE estimates' sensitivity to design changes. Although the recent literature highlights a significant sensitivity of LE estimates to the choice of control items (Chuang et al. 2021), findings from this setting show that, along with a reduction in the recall period, slight modifications of the control items have no significant impact on the difference-in-means estimator.

While the list experiment has become increasingly popular for measuring sensitive behaviors, this indirect questioning method also presents some limitations. Future research should focus on assessing the robustness of the list experiment to survey design and implementation, as evidence on this matter remains limited. It may help researchers and practitioners appreciate the contexts where the use of LE would be most appropriate, and adopt better implementation practices. In addition, any attempt to validate the list experiment method against objective data would also increase confidence in using this method to better measure child-labor-related behaviors, and overcome the associated social desirability bias.

The remainder of this article is organized as follows. Section 2 provides background on child labor in the cocoa sector in Côte d'Ivoire. Section 3 presents the data, the design and implementation of the list experiment, and the empirical strategy. Results are discussed in Section 4, along with robustness checks. Section 5 concludes.

## 2. Child Labor in the Ivorian Cocoa Sector

### 2.1. Cocoa Sustainability Standards and Child Labor

Child labor has become a major issue in the cocoa sector since the early 2000s, when the cocoa industry committed to reducing the worst forms of child labor by 70 percent by 2020.[6] Despite ongoing efforts to achieve this goal by both governments and cocoa companies, the sector has failed to meet its commitment. Indeed, many children continue to work on cocoa farms in West Africa, many of whom perform dangerous tasks that may threaten their health, such as land clearing, pruning, pod-breaking with a sharp tool, harvesting with a machete, or manipulating agrochemicals (Sadhu et al. 2020). In Côte d'Ivoire, all these activities are prohibited for children under the age of 18 and recognized as hazardous under Ivorian

6 In September 2001, several companies from the cocoa industry further ratified the Harkin–Engel Protocol, an international public–private partnership aimed at eradicating the worst forms of child labor in cocoa value chains in West Africa.

legislation.[7] In response to growing international demand for sustainable cocoa, as well as increasing legal requirements in both importing and producing countries, cocoa traders are constantly trying to improve the detection and remediation of child labor in their supply chains.

Since the late 2010s, the various independent cocoa certification labels (Fairtrade, RA) represent one of the potential tools for reducing child labor in the cocoa sector, now covering nearly 60 percent of the country's cocoa farmers (The World Bank 2019). These labels allow certified farmers to obtain an additional premium and to benefit from various trainings,[8] related to good agricultural and social practices. In return, farmers must comply with social and environmental requirements, including non-reliance on child labor on their farms. This ban is associated with awareness campaigns to inform certified cocoa farmers about the harmful effects of child labor, the benefits of schooling for children, and the legal minimum age at which children can start working. Certified cocoa cooperatives must also implement systems to prevent, monitor, and address child labor at the community level. Similarly, cocoa-buying companies are required to implement a child labor monitoring system[9] among certified cocoa farmers in their supply chain. For this, they implement household surveys in which cocoa farmers are asked about their labor practices on their cocoa farms, including child labor. These requirements are assumed to contribute to reducing child labor on cocoa farms.

The impact of certification schemes on farmers has received increased attention from researchers in recent years. Mainly studied in the coffee sector, the economic effects are mixed and differ according to countries, types of certification, and context (de Janvry, McIntosh, and Sadoulet 2015; Oya, Schaefer, and Skalidou 2018; Meemken 2020). In the cocoa sector, studies assessing the impact of certification on cocoa farmers show a positive effect on cocoa productivity, income, and consumption (Astrid Fenger et al. 2017; Ingram et al. 2018; Knößlsdorfer, Sellare, and Qaim 2021). Nevertheless, these positive results seem conditioned by the quality of the services and the supervision that farmers receive from the cooperative to which they are affiliated. Recently, Sellare et al. (2020) confirmed the importance of accounting for the heterogeneity of cocoa cooperatives to improve the evaluation of certification schemes in Côte d'Ivoire. Mainly focused on the impacts of certification on farmers' incomes or productivity, the literature has paid only little attention to other outcomes likely impacted by farmers' participation in a certification program (Chen et al. 2017). Child labor, which is prohibited by all cocoa certification schemes in Côte d'Ivoire, is one of those topics on which research remains limited.

### 2.2. Social Desirability Bias among Certified Cocoa Farmers

While it is prohibited by both national legislation and certification bodies in the context of certified cocoa value chains, certified cocoa farmers may still rely on child labor for a number of tasks on their farms.

In some cases, the use of child labor may be the only available option for some agricultural households. According to Cigno (2009), the poorest may not be able to afford adult labor, and others may need the additional wage earned by their children to reach a decent level of income. Empirical evidence among cocoa households in Côte d'Ivoire supports the idea that child labor is rooted in household poverty (Nkamleu and Kielland 2006). However, previous literature also indicates that land-rich agricultural households are also likely to use child labor (Bhalotra and Heady 2003), as each additional hectare of land increases the household's need for labor. This wealth paradox has further been requalified as an inverted U-shaped relationship between land wealth and child labor (Basu, Das, and Dutta 2010). A combination of market failures, particularly in rural areas of developing countries (Dumas 2013), may explain this paradox. While several studies highlight the positive impact of certification on cocoa farmers' incomes (Astrid

7 Dangerous work can be allowed from the age of 16 provided that children have received specific instruction or vocational training.

8 As an example, Fairtrade provides training guides and materials for farmers on sustainability, biodiversity, gender equality, and child labor, available on their dedicated website: producerlibrary.org.

9 Usually based on the CLMRS developed by the International Cocoa Initiative (ICI).

Fenger et al. 2017; Ingram et al. 2018), the literature on child labor thus suggests that the relationship between income and child labor is still unclear and depends on many other factors.

In Côte d'Ivoire, cocoa-producing households face a significant need for labor as cocoa-related activities are not mechanized and thus are labor intensive. While some studies have found a positive impact of certification on cocoa productivity (Ingram et al. 2018), it may nonetheless put upward pressure on farmers' need for any type of labor, including children. Moreover, while cocoa is a year-round labor-intensive activity, the labor demand peaks during the main cocoa season, when cocoa production is most intensive (Davis Pluess 2018). Cocoa farmers are, however, usually clustered in cocoa-growing communities, which means that during peak season, farmers from the same community simultaneously seek additional labor for their farms. Considering the remoteness of cocoa communities, accessing an external labor force might be an issue for cocoa farmers. All these factors increase the likelihood that a cocoa-producing household will face difficulties finding available adult labor in a tight labor market.[10]

Constraints on the labor market could, in principle, be mitigated by efficient land or credit markets. The sale or lease of land may reduce farmers' need for labor while increasing their income. However, farmers in Côte d'Ivoire continue to hold customary land titles despite successive policy attempts to facilitate the official registration of property rights.[11] Land exchanges are therefore limited because they are fraught with uncertainty. Access to credit can facilitate income smoothing, limiting the need to take children out of school for financial reasons, or to make them work to increase household income (Guarcello, Mealli, and Rosati 2009). However, few Ivorian cocoa farmers are affiliated with a bank or microcredit institute, and most of their loans come from informal sources (Lonie et al. 2018) and are of small amounts. The lack of school infrastructure at the community level or within a reasonable distance may also increase education costs and reduce parents' incentive to send their children to school (Fafchamps and Wahba 2006).

All these market failures are present in both non-certified and certified cocoa value chains, suggesting that despite some directly observable income and productivity benefits of certification for cocoa farmers, their structural difficulties in accessing labor remain. Beyond economic factors explaining why certified cocoa farmers might be using child labor despite its interdiction by certification schemes, the existing monitoring difficulties might also play against the ban. Cocoa farms are often isolated and difficult to reach, making the monitoring process more costly and less effective.[12] Enforcement of the ban, which relies on the awareness, involvement, and cooperation of farmers and various stakeholders at the community level, is also likely to be limited. Indeed, the very definition of a child, and therefore of child labor, established by the international community is inconsistent with that of Ivorian cocoa communities. The latter do not assign tasks to children according to their age, but rather to their physical capacities, which they are able to appreciate (Buono and Babo 2013). At the community level, it is therefore unlikely that the ILO standard would prevail, as it sometimes conflicts with social traditions of Ivorian cocoa households. The efficiency of community-based systems that detect and monitor child labor is thus probably undermined by the discrepancies between standards. Yet previous literature on child labor has shown that the ability to enforce a policy is a key factor for its success. Consequently, where enforcement is low, policies to reduce child labor are shown to have little or no effect (Edmonds and Shrestha 2012; Bargain and Boutin 2021).

The literature on the determinants of child labor suggests that cocoa farmers may be using child labor on their farms, even if they belong to certified schemes. Yet, by using child labor on their certified cocoa farms, farmers face the risk of being excluded from certified cocoa value chains. This would imply losing all the associated benefits, including the extra income they receive from selling certified cocoa. Thus,

10 The difficulty in finding adult labor was also discussed in the 2020 Cocoa Barometer.

11 Ministry of Agriculture and Rural Development, Declaration of Rural Land Policy of Cote d'Ivoire, 2017.

12 Audits performed by certification bodies have recently been declared ineffective in addressing child labor (Fountain and Huetz-Adams 2020).

certified cocoa farmers are likely to exhibit a social desirability bias when directly questioned about their use of child labor, especially when it involves hazardous tasks. While stakeholders in the cocoa sector conduct household or CLMRS surveys to collect data on child labor from cocoa farmers, they only rely on direct questioning methods.

### 2.3. The List Experiment

List experiments[13] are one of the main methods used as an alternative to direct questioning about sensitive topics, seeking to reduce respondents' social desirability bias. They consist of showing participants a list of statements (usually four or five) and asking how many apply to their case. Since the answer provided by the respondent is a global response to a set of questions, the enumerator has no way of knowing precisely which statements apply to the respondent's case, which gives the latter confidentiality and should encourage them to respond more honestly. In the simplest version of the list experiment, respondents are randomly divided into two groups. The control group receives a list of $n$ non-sensitive items, and the treatment group receives a list with the exact same $n$ sensitive items as the control group, plus a sensitive one. The proportion of the sample engaging in sensitive behavior is then estimated by a simple difference in mean responses between the two groups.

Initially proposed by Raghavarao and Federer (1979) in a slightly different version, the list experiment has since been used in many areas to study various topics such as politics (Çarkoğlu and Aytaç 2015; Comşa and Postelnicu 2013; Corstange 2010; Holbrook and Krosnick 2010), health (Lépine et al. 2020; Moseson et al. 2015), homophobia (Lax, Phillips, and Stollwerk 2016; Co Man, Co Man, and Keith 2017), intimate partner violence (Cullen 2020), or food security (Tadesse, Abate, and Zewdie 2020). The use of list experiments has revealed the existence of a social desirability bias for various sensitive topics, raising concerns about previous estimates obtained from responses to direct questioning. As an example, Gonzalez-Ocantos et al. (2012) found that 24 percent of Nicaraguan voters exchanged their vote for a gift or service using the list experiment approach, compared to only 2 percent when asked directly. Similarly, Co Man, Co Man, and Keith (2017) found a 67 percent increase in the proportion of homophobic sentiment in the workplace through the list experiment in an online survey, compared to the direct question. The better performance of the list experiment compared to other methods of questioning (direct and indirect) has been highlighted by several other empirical studies (Çarkoğlu and Aytaç 2015; Holbrook and Krosnick 2010; Moseson et al. 2015; Wolter and Laier 2014).

Yet the list experiment sometimes leads to insignificant results or of the same magnitude as those obtained via the direct question (Arentoft et al. 2016; Krebs et al. 2011; Lax, Phillips, and Stollwerk 2016), which may occur when experiments are underpowered. Indeed, one limitation of the list experiment is that it requires a sufficiently large sample size to detect social desirability bias, as it produces estimates with relatively high variance due to the non-sensitive items. A thoughtful design of the LE is thus essential for its performance. In particular, it is recommended to prevent the use of exclusively high or low prevalence items (Blair and Imai 2012; Glynn 2013; Tsuchiya, Hirai, and Ono 2007), to avoid a ceiling or floor effect issue (further explained in the estimation strategy session), and to propose negatively correlated items to reduce the variance of LE estimators (Glynn 2013). Recently, Chuang et al. (2021) also showed that the choice of non-sensitive items could also influence respondent compliance and the results obtained with the list experiment. Focusing on sexual and reproductive behavior in Côte d'Ivoire, they show that choosing non-sensitive items related to the sensitive one implied a better LE performance than non-sensitive items fully innocuous and unrelated to the studied phenomenon. The lack of difference between the list experiment approach and direct questioning may also reveal the absence of social desirability bias, contrary to the researchers' initial assumptions (Blair, Coppock, and Moor 2020).

13 Also known as the unmatched count technique or the item count technique.

Although used in many research settings, the list experiment has not yet been used to measure child-labor-related behavior.

## 3. Research Design

### 3.1. The Data

This paper uses data from a socioeconomic survey, designed by the author and carried out by an international cocoa trading company, among Ivorian certified cocoa households that are part of its supply chain. Two questionnaire versions are used to implement a list experiment.[14] The survey is divided into two complementary questionnaires: (a) a household-level questionnaire, which collects household characteristics such as size, composition, and basic characteristics of its members, along with information on banking, savings and credit practices, and wealth; (b) a farm-level questionnaire which covers issues such as farm size, farm age, pesticide use, the incidence of pests and diseases, use of family and hired labor, and cocoa-related income. The survey is only addressed to the certified cocoa farmer, who is the head of the household in most cases. No other respondents, such as other household members, were surveyed. Data collection started during the main cropping season in December 2019 and stopped in mid-March 2020. The enumerators collected information in several cocoa-growing areas in the regions of Agnéby-Tiassa, Cavally, Grands Ponts, Guémon, Gôh, Lôh-Djiboua, Nawa, and Tonkpi. All enumerators were fluent in French but were also proficient in local languages[15], making it possible to survey non-French-speaking farmers. The translation of the list experiment questions and proposed items into different local languages was discussed and defined during the enumerators' training.

The final sample consisted of 4,458 UTZ[16] (88 percent) and RA (12 percent) certified cocoa farmers,[17] randomly selected among sampled cooperatives.[18] The great majority of them are Ivorian (74 percent) or Burkinabe (24.5 percent), and almost all of the farmers are men (95 percent) (see supplementary online appendix table S1.1). Cocoa farmers are on average 42 years old, generally married (89 percent), and live in a household with at least one child (90 percent). The levels of farmers' education remain relatively low: 57 percent of them never attended school, and only 17 percent of them have reached secondary or higher education. The median size of the cocoa farm is 2.09 ha, but cocoa farmers also grow other crops on their farm in 51 percent of cases.

### 3.2. Design and Implementation of the List Experiment

The implemented list experiments focus on three different categories of work performed by children under 16 on certified cocoa farms. First, this study aims to estimate the proportion of cocoa farmers being helped by a child for tasks classified as hazardous by the ILO. The focus is on hazardous work related to farm preparation (applying pesticides, clearing, tree felling, burning plots), and harvesting and pod breaking. Second, we seek to estimate the proportion of cocoa farmers who hire and pay a child to perform any task(s) on their farm. The light work children perform on cocoa farms is not considered here, as this is not harmful to the child and is commonly encouraged by their family. Those tasks, tolerated by the ILO, are generally perceived by cocoa farming households as an integral part of the child's education and

14 Randomization of the survey version among farmers was also carried out and monitored by the author.

15 Such as Baule, Dioula, or Senoufo.

16 Although the survey ended before reaching the target sample size due to the COVID-19 outbreak, post hoc power calculations show that our sample allows us to detect a minimum detectable effect of 8 percent.

17 Farmers in our sample benefited from several awareness sessions on the negative effects of child labor. Each cocoa farmer and their household are individually sensitized to child labor when they start working for a certified cooperative. A second awareness session, this time at the community level, is conducted every year.

18 The cooperatives sampled were all certified cooperatives that were part of the supply chain of the cocoa trading company with whom the author worked with.

**Table 1.** List Experiments and Direct Questions

| | Wording of the questions |
|---|---|
| Harvesting & pod-breaking | ***List experiment #1***<br>During the last campaign, for harvesting and breaking the pods:<br>(1) No woman over 16 helped me.<br>(2) At least one member of my family helped me.<br>**(3) One or more people under 16 helped me.**<br>(4) A group of communal laborers helped me.<br>In total, how many statements apply to your case?<br>***Direct question***<br>During the last campaign, for harvesting and breaking the pods, did one or more people under 16 help you? |
| Farm Preparation | ***List experiment #2***<br>During the last campaign, to prepare my plot (applying pesticides, clearing, felling of trees, burning of plots):<br>(1) No woman over 16 helped me.<br>(2) At least one member of my family helped me.<br>**(3) One or more people under 16 helped me.**<br>(4) A group of communal laborers helped me.<br>In total, how many statements apply to your case?<br>***Direct question***<br>During the last campaign, to prepare your plot (applying pesticides, clearing, felling of trees, burning of plots), did one or more people under 16 help you? |
| Employed child labor | ***List experiment #3***<br>During the last campaign, to do some work on my farm (for all types of activities):<br>(1) I have not employed any woman over 16.<br>(2) At least one member of my family helped me.<br>**(3) I have employed and paid one or more people under 16 who are not members of my family.**<br>(4) I employed a group of communal laborers.<br>In total, how many statements apply to your case?<br>***No direct question*** |

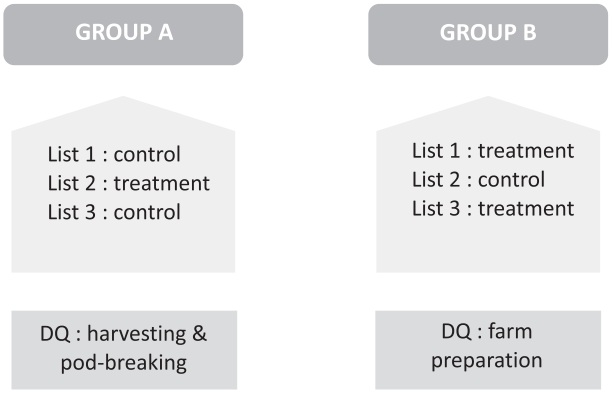
*Source:* Author.
*Note:* For each list experiment, the bolded item is the sensitive one, related to child labor. It is only read to cocoa farmers who belong to the treatment group for that list experiment.

personal development (Buono and Babo 2013). The scope of this study is also restricted to children under 16, in line with Ivorian legislation which, under certain conditions, allows children between 16 and 18 to perform hazardous work.[19] Thus, this study focuses on the worst forms of child labor.

Three list experiments were proposed to cocoa farmers, one for each of the aforementioned categories of work (table 1). Cocoa farmers were randomly assigned to two groups (fig. 1). Group A serves as a control group for list experiments 1 and 3, i.e., farmers in group A are not exposed to the sensitive item in lists 1 and 3; and as a treatment group for list 2. The inverse is true for group B, which serves as a control group for list experiment 2 and treatment group for list experiments 1 and 3. The list experiments were always administered to respondents in the same order, regardless of the respondent group. Section entitled "Potential Source of Measurement Errors" discusses whether and how this choice may have affected the reliability of the LE estimates.

19 See decree No. 2017-017 MEPS/CAB of June 2, 2017, determining the list of hazardous work prohibited for children.

**Figure 1.** Sequencing of the List Experiments and Direct Questions.




*Source:* Author.

*Note:* DQ stands for direct question. All farmers are asked the three list experiments. Farmers in group A belong to the control group for the list experiments about harvesting and pod-breaking (list 1), and child employment (list 3), and to the treatment group for farm preparation (list 3). Conversely, farmers in group B are in the control group for farm preparation (list 3), and in the treatment group for harvesting and pod-breaking (list 1) and child employment (list 3). Group A farmers are also directly questioned about their use of child labor for harvesting and pod-breaking, and farmers in Group B about their use of child labor for farm preparation.

The non-sensitive items used in the list experiments are related to the sensitive item's topic, i.e., the farmer's use of labor on their certified cocoa farm. We argue that including non-sensitive items unrelated to the sensitive one would only increase its salience to respondents, which is in line with findings from Chuang et al. (2021). In their study about LEs and sexual practices in Côte d'Ivoire, they found that the LE performed worse when the control items were non-sensitive or unrelated to the sensitive item. The LEs used in this study are hence focusing on different types of labor that cocoa farmers may use on their farms. Thus, enumerators ask cocoa farmers about their overall labor practices, rather than specifically about their use of child labor. The non-sensitive items also relate to types of labor that cocoa farmers are typically asked about in questionnaires administered by different stakeholders in the sector.[20] Although the LEs also mention child labor and may still lead to farmers' mistrust, it is expected that an LE's format and wording, as well as the proposed control items, help reduce the salience of the sensitive item (i.e., the use of child labor), as it is framed as one of several types of labor possibly used by cocoa farmers.

The non-sensitive items were furthermore designed following the recommendations provided by the list experiment literature. Two negatively correlated items (items 1 and 2) were chosen to minimize the variance of the LE estimator, as recommended by Glynn (2013).[21] This also prevents a treated farmer from answering yes or no to all items, thus revealing one's answer to the sensitive item (Tsuchiya, Hirai, and Ono 2007). The non-sensitive items used in the LEs were also slightly modified from those designed initially, based on feedback from enumerators who had the opportunity to test the LEs during a one-week test with certified farmers. The modifications resulted in more relevant and understandable non-sensitive items for cocoa farmers.

Before proceeding with the list experiment, each enumerator explained to the respondent how this particular question worked, as follows:

> Please put your hand behind your back. I am going to read three or four statements about the labor used on your cocoa farm in the last 12 months. If the statement read is true, and applies to your case, lift a finger behind your

20 Certified cocoa farmers are subjected to different types of surveys including modules on their labor use, such as CLMRS surveys conducted by cocoa traders or surveys conducted by researchers (NORC survey).

21 Items 1 and 2 are, indeed, negatively correlated. In most cases, the farmer is at least helped by his wife on his cocoa farm. Based on the information collected about household members in the survey, we can also state that the wives of the sampled farmers are also older than 16 years old.

back. If the statement is not true, do not lift a finger. After I read you all the statements, show me your hand and the number of fingers you raised. I do not want to know which statements apply to you, only the number of statements that are true for you. Please do not answer yes or no when I read statements to you. Your answers must remain confidential.

Enumerators were further encouraged to provide an example to the farmer in case he seemed confused about the question's wording. The proposed example was a simple one, using familiar terms and representing a common situation for farmers. Acknowledging that the list experiment may involve a higher cognitive load than a direct question, farmers were not asked to count the number of statements that applied to their case, but only to raise a finger behind their backs for each positive response to an item. Adding up the number of positive answers given to each item was left to the enumerators.

In addition to the list experiments, farmers from group A were asked directly about the use of child labor for harvesting and pod breaking, and farmers from group B were asked directly about their reliance on children for plot preparation. To be able to compare the answers given to the list experiment and the direct question, the wording of the direct questions was identical to that of the corresponding list experiment (table 1). No direct question was asked about employed child labor. During the author's fieldwork, this issue was found to be far too sensitive and could have caused farmers to refuse to continue the survey, as the employment of children is sometimes associated with child trafficking and forced labor.

The design and sequencing of the direct questions were used to minimize potential response bias from farmers. First, farmers in each group were asked directly about only one type of child labor, as asking several direct sensitive questions might raise their suspicions or be perceived as offensive, which could distort the answers given to the second direct question and lead to downward bias. Second, farmers assigned to the treatment group for one type of child labor with the LE were not asked directly about that same sensitive topic, thus minimizing the potential salience effect and downward bias affecting the response to the direct question.

### 3.3. Estimation Strategy

A difference-in-means estimator is used to obtain the estimated prevalence of use of the three types of child labor mentioned above. This estimator is used to obtain the average number of items to which respondents answered yes in both the control and treatment groups. Any difference in means between the treatment and control groups is attributed to the sensitive item. Following Blair and Imai (2012), the estimator is defined as follows:

$$\hat{\tau} = \frac{1}{N_1}\sum_{i=1}^{N} T_i Y_i - \frac{1}{N_0}\sum_{i=1}^{N}(1 - T_i)Y_i,$$

with $T_i$ taking the value of 0 if the respondent is in the control group, 1 otherwise. The variable $Y_i$ is the answer given by the respondent, which can take a value from 0 to $J$, $J$ being the maximum possible value, i.e., the total number of proposed items. The values $N_1$ and $N_0$ correspond to the size of the treatment and control groups, respectively.

The validity of the difference-in-means estimator relies on three key assumptions (Blair and Imai 2012): a random allocation of respondents between the control group and treatment group (1), the assumption of no design effect (2), and the assumption of no liars (3). If these three assumptions are not simultaneously satisfied, the estimator does not yield valid estimates.

Assumption (1) cannot be rejected, as farmers who received questionnaires A and B display similar characteristics, supporting the successful random allocation of treatment between the two farmer groups (table 2).

**Table 2.** Basic Characteristics of Respondents in Groups A and B

| | Control | Treatment | (C-T) *p*-value | Obs. |
|---|---|---|---|---|
| Age | 41.99 | 41.88 | 0.749 | 4,458 |
| Married | 0.89 | 0.89 | 0.951 | 4,458 |
| No education | 0.57 | 0.56 | 0.289 | 4,458 |
| Ivorian | 0.73 | 0.74 | 0.947 | 4,458 |
| Number of children | 2.62 | 2.61 | 0.726 | 4,458 |
| Number of adults | 3.29 | 3.38 | 0.117 | 4,458 |
| Cocoa income (XOF) | 995,001 | 992,207 | 0.913 | 4,458 |
| Yields (kg/ha) | 656.57 | 639.16 | 0.392 | 4,458 |
| Other crop | 0.51 | 0.51 | 0.911 | 4,458 |

*Source:* Author's calculation.
*Note:* The table displays the mean of each variable for control and treatment groups, as well as the p-value from t-test indicating whether the difference between these means is significant.

**Table 3.** No-Design Effect Test Results

| | Harvesting & pod-breaking | | | | Farm preparation | | | | Employed child labor | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *y*-value | $\hat{\pi}_{y0}$ | SE | $\hat{\pi}_{y1}$ | SE | $\hat{\pi}_{y0}$ | SE | $\hat{\pi}_{y1}$ | SE | $\hat{\pi}_{y0}$ | SE | $\hat{\pi}_{y1}$ | SE |
| 0 | 2.41 | 0.003 | 1.01 | 0.005 | 2.91 | 0.003 | 0.11 | 0.005 | 2.74 | 0.003 | 0.69 | 0.005 |
| 1 | 24.75 | 0.010 | 4.81 | 0.014 | 27.93 | 0.010 | 3.73 | 0.014 | 25.73 | 0.010 | 0.92 | 0.014 |
| 2 | 44.50 | 0.013 | 8.37 | 0.012 | 39.28 | 0.014 | 5.89 | 0.013 | 40.35 | 0.014 | 8.61 | 0.013 |
| 3 | 4.21 | 0.010 | 9.94 | 0.006 | 8.13 | 0.011 | 12.02 | 0.007 | 5.55 | 0.011 | 15.41 | 0.008 |

*Source:* Author's calculation.
*Note:* The table shows the estimated proportion of respondents characterized by the total number of affirmative answers to the non-sensitive items ($y$) and their answer for the sensitive item (1 stands for a positive response to the sensitive item, 0 for a negative one). These proportions are estimated using the R package proposed by Blair and Imai (2012). For example, the value in the fourth row and third column suggests that an estimated 9.94 percent of farmers answered no to all non-sensitive items but yes to the sensitive item. Standard errors are also provided for each estimated proportion. We use Blair and Imai (2012)'s notation.

Assumption (2) implies that the inclusion of the sensitive item does not affect the respondents' answers to control items. That is, the treatment group's answers to non-sensitive items should be approximately equivalent to those of the control group. Additionally, the average total number of items in the treatment group must be equal to or greater than that of the control group, but at most by one item. This condition corresponds to the null hypothesis of the test proposed by Blair and Imai (2012) to detect the presence of a design effect. Keeping their initial notation, we have

$$H_0 \begin{cases} \Pr(Y_i \leq y \mid T_i = 0) \geq \Pr(Y_i \leq y \mid T_i = 1) & \text{for all } y = 0, \ldots, J-1, \\ \Pr(Y_i \leq y \mid T_i = 1) \geq \Pr(Y_i \leq y-1 \mid T_i = 0) & \text{for all } y = 1, \ldots, J. \end{cases}$$

This test yields estimates of the different proportions of each type of respondent ($\hat{\pi}_{yz}$), characterized by the total number of affirmative answers to the non-sensitive items ($y$, which can take a value from 0 to 3) and their answer for the sensitive item ($x$, which takes the value of 0 for a negative answer, and 1 for a positive one). If all these proportions are positive, the null hypothesis of no design effect cannot be rejected (Blair and Imai 2012). This is the case for the three list experiments, as shown in table 3.

Finally, assumption (3) implies that respondents answer honestly to the sensitive item. While there is no statistical test to verify this hypothesis, the LE approach should, by design, enhance honest responses from respondents to the sensitive item. Nevertheless, the validity of this assumption may be questioned when the respondent would like to answer yes (ceiling effect) or no (floor effect) to all the items. In this case, the enumerator can infer the respondent's answer to the sensitive item, which removes the private nature of the answer and may encourage respondents to misreport their true answer. Such a concealment

**Table 4.** Distribution of Farmers according to Their Answer to the List Experiments

| | Harvesting & pod-breaking | | | | Farm preparation | | | | Employed child labor | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Control | | Treatment | | Control | | Treatment | | Control | | Treatment | |
| *y*-value | Freq. | (%) | Freq. | (%) | Freq. | (%) | Freq. | (%) | Freq. | (%) | Freq. | (%) |
| 0 | 79 | 3.43 | 52 | 2.41 | 65 | 3.02 | 67 | 2.91 | 79 | 3.43 | 59 | 2.74 |
| 1 | 681 | 29.56 | 555 | 25.77 | 682 | 31.66 | 646 | 28.04 | 614 | 26.65 | 569 | 26.42 |
| 2 | 1,218 | 52.86 | 1,062 | 49.30 | 973 | 45.17 | 991 | 43.01 | 1,128 | 48.96 | 889 | 41.27 |
| 3 | 326 | 14.15 | 271 | 12.58 | 434 | 20.15 | 323 | 14.02 | 483 | 20.96 | 305 | 14.16 |
| 4 | | | 214 | 9.94 | | | 277 | 12.02 | | | 332 | 15.41 |
| Total | 2,304 | | 2,154 | | 2,154 | | 2,304 | | 2,304 | | 2,154 | |

*Source:* Author's calculation.

*Note:* The table shows the distribution of farmers based on the total number of items to which they answered yes. For example, 52.86 percent of farmers in the control group answered affirmatively to two items for harvesting and pod-breaking, compared to 49.30 percent in the treatment group.

strategy, known as strategic measurement error, calls for specific care to avoid floor or ceiling effects when designing an LE. In particular, researchers should avoid proposing lists of high-prevalence or low-prevalence items (Blair and Imai 2012; Glynn 2013; Tsuchiya, Hirai, and Ono 2007), as this increases the likelihood that a respondent will answer yes or no to all the items. Glynn (2013) advises proposing negatively correlated non-sensitive items, which has the added advantage of reducing the variance of the difference-in-means estimator, which is known to be less powered than the direct method. For each list experiment, these recommendations were followed in the choice of non-sensitive items (table 1). Yet it is likely that a ceiling effect occurred, with 15 percent to 20 percent of respondents answering yes to all items in the control group (table 4). Thus, it is possible that some farmers from the treatment group who used child labor, but who had already answered yes to all the control items, did not answer truthfully to the sensitive one. This suggests that the difference-in-means estimates are to be interpreted as lower bounds of the true prevalence of child labor use.

## 4. Results

### 4.1. Differences between Direct Questioning and LE Responses

The results of the list experiment (column 3), the direct question (column 4), and the difference between the two measures (column 5) for each type of child labor are reported in table 5. The difference in means between the treatment and control groups is positive, statistically significant, and relatively large for all three types of child labor. The findings suggest that the use of child labor is not an isolated practice among certified cocoa farmers: 24.1 percent of the farmers in the study sample used child labor for harvesting and breaking the cocoa pods during the past 12 months, 21.7 percent for farm preparation, and 25.6 percent of them employed and paid at least one child under 16 to perform any task on their cocoa farm.

Column 4 of table 5 reports the prevalence of farmers who responded positively to the direct question regarding child labor for harvesting and pod-breaking, and farm preparation. When cocoa farmers are asked directly, 12.76 percent of them report having been helped by a child under 16 for harvesting and breaking the pods in the last 12 months, and 18 percent of them declared that they used child labor for farm preparation. These results are half as high as those obtained using the list experiment, supporting the existence of a social desirability bias among certified cocoa farmers when directly asked about their use of child labor to perform hazardous tasks. Although farmers were not asked directly about hiring and paying children under 16 to perform any task on their farm, it is likely that this work category is prone to similar social desirability bias as hazardous child labor. Indeed, employing and remunerating a child

**Table 5.** Proportion of Farmers Using Child Labor Estimated by List Experiment and Direct Questioning Method

| | (1) Control mean | (2) Treatment mean | (3) Difference in means (%) | (4) Direct question (%) | (5) Difference (pp) |
|---|---|---|---|---|---|
| List 1: Harvesting & pod-breaking | | | | | |
| | 1.777 | 2.018 | 24.12 | 12.76 | 11.36*** |
| *N* | 2,304 | 2,154 | 4,458 | 2,304 | |
| List 2: Preparing the plot | | | | | |
| | 1.824 | 2.042 | 21.75 | 11.28 | 10.47*** |
| *N* | 2,154 | 2,304 | 4,458 | 2,154 | |
| List 3: Employed child labor | | | | | |
| | 1.874 | 2.130 | 25.63 | | |
| *N* | 2,304 | 2,154 | 4,458 | | |

*Source:* Author's calculation.
*Note:* $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

under 16, regardless of the work performed, is legally prohibited under Ivorian law.[22] This information is also common knowledge in cocoa farming communities that benefit from child-labor awareness-raising campaigns.

The difference of over 10 percentage points between the direct method and the list experiment indicates that at least half of the farmers who do use child labor are not willing to declare it (column 5). These results are in line with the initial hypothesis about the existence of a social desirability bias. From the author's field observations, these findings are interpreted as a rational response to farmers' double pressure regarding hazardous child labor, which is prohibited both in Côte d'Ivoire and by the certification schemes.

## 4.2. Robustness Checks

Estimating the proportion of certified cocoa farmers relying on child labor is challenging as it is prone to measurement errors, as discussed in Section 4.2.1. Sections 4.2.2 and 4.2.3 test for the robustness of our LE estimates to design changes and the extent to which the LE method is understood by farmers.

### 4.2.1. Potential Source of Measurement Errors

While respondent type may impact answers given to child-labor-related questions (Bardasi et al. 2011), this possibility is ruled out in this setting. Since the survey was administered to certified cocoa farmers, the risk of getting inaccurate reporting associated with questioning proxy respondents is removed.

Memory bias might be present due to the length of our recall period, which refers to the last cocoa season, i.e., the past 12 months. Although such a long recall period may increase the likelihood of memory bias, this risk is lowered for two reasons. First, the annual cocoa season is divided into two 6-month periods: the main crop, when cocoa production peaks, and the light crop, when cocoa production is lower. These two seasons were discussed at length by the enumerators throughout the survey, facilitating the recovery of farmers' memories, including those related to their use of labor. Second, while the literature on recall bias associated with the measurement of farm labor shows that recall period length matters, whether a longer recall period overestimates or underestimates the phenomenon is still unclear, depending on the context (Gaddis et al. 2021; Arthi et al. 2018). Further, it is likely that measuring the occurrence of a binary behavior (whether or not the farmer used child labor on his farm) decreases the likelihood of recall bias compared to other more complex measures of farm labor, such as hours worked or the number of workers. Last, the fact that the same recall period was used for both the direct question and the list experiment suggests that if any memory bias occurs, it should affect both measures in the same way.

22 Article 23.2 of the Labour Code.

The use of the list experiment is expected to reduce social desirability bias, which remains an important issue associated with estimating child-labor-related measures. Yet this method itself may be prone to measurement errors, mainly because the LE is considered more cognitively demanding for respondents than a direct question. People with cognitive difficulties or low levels of education may struggle to understand this indirect questioning method, which may lead them to give a random answer to the enumerator. As a result, measurement errors would correlate with respondents' education levels. To mitigate this issue, farmers were only asked to lift a finger behind their backs when an LE statement applied to their case, thus removing the need for them to count and aggregate their answers. The "List experiment and cognitive load" section reports the results of a comprehension test of the LE from another survey also carried out with Ivorian certified cocoa farmers, which supports a good level of understanding of the LE among this population, even for the less educated farmers.

One might also be concerned about any measurement issue arising from the fact that not all of the items composing the LEs are mutually exclusive. Thus, control items 2 and 4, which are used in the three list experiments and asked to all farmers, whether in the control or treatment group, can overlap. Similarly, farmers belonging to the treatment group for LEs 1 (harvesting and pod-breaking) and 2 (farm preparation) receive items 2 and 3, which can also overlap if one was helped by a child under 16 who also belongs to one's family. It is therefore possible that a farmer might answer yes to both control item 2 and sensitive item 3 when considering the same person (a child of his family under 16). This could be an issue under the following two conditions: (a) farmers who were helped by a child under 16 belonging to their family have different response patterns to item 2, i.e., in some cases, they answer yes to item 2, while in other cases they do not; (b) these different response patterns are associated with treatment or control group assignment. As the farmers in the control and treatment groups have socioeconomic characteristics that are not statistically different from each other on average (table 2), their average response to item 2 should also be similar, thus not impacting the difference in means between the two groups. The same reasoning holds for the overlap of control items 2 and 4.

Respondents' answers to LEs may also be influenced by the order in which they are presented, as each farmer is administered the three list experiments. To avoid any order effect, cocoa farmers were first presented with the list experiment related to tasks that are a priori the most often performed by children on cocoa farms (Sadhu et al. 2020). The last list experiment asks about the most sensitive practice (i.e., hiring children outside one's family). This sequence was designed to avoid offending the respondent by asking first a question on the most sensitive topic. The statements of the list experiments were also neutral and did not contain any judgments or terms that could offend the farmers. We did not use the term "children" or describe the tasks we were interested in as "hazardous." Despite these precautions, an order effect may persist, but it is likely to be limited. It is unlikely that cocoa farmers in group A experience an order effect, as they are only exposed to one list experiment as a treatment group, implying that they are only presented with the sensitive item once. If any order effect occurred, it would likely be among farmers belonging to group B, who act as treatment groups for two LEs, thus being exposed to the sensitive item twice, possibly increasing its salience. This may lead them to give less truthful answers to the sensitive item of the second list experiment for which they serve as a treatment group. Acknowledging that an order effect may exist in that case, we consider that, if anything, LE estimates on children's employment (list 3) are probably a lower bound, as it may have been affected by such a downward effect.

#### 4.2.2. Sensitivity of LE Estimates to Design Changes

Two new survey rounds were conducted between December 2021 and May 2022 among certified cocoa farmers, similar to 2019. The farmers surveyed in these new survey rounds are either (a) farmers who were not surveyed in the 2019 survey due to the unexpected end of data collection due to the COVID-19 outbreak, or (b) new farmers who entered the cocoa trading company's supply chain after the 2019 survey. These new survey waves involved different respondents from 2019, so they cannot be considered

**Table 6.** Survey Characteristics across Waves

| | 2019 | 2022-V1 | 2022-V2 |
|---|---|---|---|
| *Survey characteristics* | | | |
| **Sampled individuals** | Certified cocoa farmers | Certified cocoa farmers | Certified cocoa farmers |
| **Sample size** | 4,458 | 7,447 | 5,593 |
| **Sampled regions** | Agnéby-Tiassa<br>Cavally<br>Gôh<br>Grands Ponts<br>Guémon<br>Lôh-Djiboua<br>Nawa | Agnéby-Tiassa<br>Cavally<br>Gôh<br>Grands Ponts<br>Guémon<br>Haut-Sassandra<br>La Mé<br>Lôh-Djiboua<br>Sud-Comoé | Agnéby-Tiassa<br>Cavally<br>Gôh<br>Guémon<br>Haut-Sassandra<br>La mé<br>Lôh-Djiboua<br>Sud-Comoé<br>San-Pédro |
| **Data collection period** | December 2019–March 2020 | December 2021–February 2022 | February 2022–May 2022 |
| *List experiment design* | | | |
| **Non-sensitive items** | Initial | Modified | Initial |
| **LE test** | No test | 2 tests before LE | 1 test after LE |
| **Recall period** | 12 months | 6 months | 12 months |

*Source:* Author.

*Note:* The table reports the survey characteristics and list experiment design for all three surveys used: 2019, 2022 version 1 (V1) and 2 (V2). LE stands for List Experiment.

panel data. They can nevertheless offer insight into the sensitivity of the LE to design modifications. The different LE designs used in each survey wave are summarized in table 6.

In the first survey wave of 2022 (referred to as 2022-V1 hereafter), changes were made to the LEs used in 2019, both to the design of the non-sensitive items, making them all mutually exclusive, and to the recall period, which was shortened from 12 to 6 months. Two LE comprehension tests preceding the three real LEs on child labor were also included. These modifications were made across the entire sample, making it impossible to determine a specific effect related to each design change, but rather an overall effect. The second wave of 2022 (2022-V2) restored the same LE characteristics as in the 2019 survey.

The two additional data sets successfully pass the balance tests for farmer characteristics between the control and treatment groups (tables S1.2 and S1.3 in the supplementary online appendix), thus implying successful randomization of the LEs across farmers.

Specifically, this analysis aims to test whether the changes in LE design are associated with different response behaviors for farmers belonging to the treatment group for LEs. If so, the difference in means between the treatment and control groups would reflect not only a social desirability bias, but also measurement errors related to LE design features.

We rely on the following pooled regression for each type of child labor:

$$Y_i = \alpha + \beta_0 T_i + \beta_1 \text{Year} + \beta_2 \text{Year} * T_i + \beta_3 \text{Design} + \beta_4 \text{Design} * T_i + \delta_i + \gamma_i + \epsilon_i,$$

where $Y_i$ is the farmer's response to the LE question, i.e., the number of items to which the individual $i$ answered yes (ranging from 0 to 3 for farmers in the control group, and 0 to 4 for farmers in the treatment group). The variable $T_i$ is a dummy indicating whether the cocoa farmer was assigned to the control or treatment group; Year is a dummy equal to 1 if the survey took place in 2022, 0 if it was conducted in 2019; Design is a dummy equal to 1 if the LE design is that of 2022-V1, 0 for the 2019 and 2022-V2 version; $\delta_i$ stands for a vector of household control variables (including farmer's gender, age, education, and cocoa income) and $\gamma_i$ for enumerators' fixed effects. Standard errors are clustered at the regional level.

**Table 7.** Regression Analysis of List Experiments across Survey Waves

| | Harvesting & pod-breaking | Preparing the farm | Employed child labor |
|---|---|---|---|
| Treatment group | 0.210*** | 0.209*** | 0.238*** |
| | (0.027) | (0.046) | (0.041) |
| 2022 | −0.099 | −0.040 | −0.023 |
| | (0.075) | (0.122) | (0.093) |
| 2022 ∗ treatment group | −0.186** | −0.232** | −0.197*** |
| | (0.045) | (0.055) | (0.024) |
| LE design 2022-V1 | −0.680*** | −0.721*** | −0.769*** |
| | (0.121) | (0.122) | (0.110) |
| LE design 2022-V1 ∗ treatment group | 0.014 | 0.057* | −0.014 |
| | (0.026) | (0.023) | (0.022) |
| Constant | 1.809*** | 1.693*** | 1.689*** |
| | (0.099) | (0.087) | (0.117) |
| *N* | 13,211 | 13,165 | 13,068 |
| *R*-squared | 0.447 | 0.453 | 0.473 |
| Controls | Yes | Yes | Yes |

*Source:* Author's calculation.

*Note:* Included controls: enumerators fixed effects and household characteristics (farmer's gender, age, education, and cocoa income). Clustered robust standard errors at the regional level are shown in parentheses. Level of significance: $*p < 0.05$, $**p < 0.01$, $***p < 0.001$. LE stands for List Experiment.

The results from these regressions are reported in table 7 for the three types of child labor, using the overall sample (2019, 2022-V1, and 2022-V2).

Findings show that being surveyed with LEs during the main cocoa season in 2022 versus 2019 has no significant average effect on the number of items farmers answer yes to, but it does for the farmers in the treatment group. Since the only difference between the control and treatment groups is the sensitive item related to child labor use, two assumptions may explain this result. Either there have been structural factors leading to a decline in child labor use between 2019 and 2022, or differences in training and implementation of the LE may have impacted how farmers answered the LE. The descriptive statistics reported in table S1.4 suggest a reduction in the use of child labor between 2019 and 2022 when looking at answers to direct questions. Although these results should be interpreted with caution, as they do not come from panel data and rely on direct questioning, they provide some indication of the overall trend. However, this decline in the use of child labor is not associated with a decrease in the use of family or hired labor.

Changes in the design of the list experiment (reduction of the recall period to 6 months, addition of LE comprehension tests, and modification of non-sensitive items to make them mutually exclusive) have a large negative average effect on the number of items to which farmers respond yes. As expected, this result is driven by the reduction of the recall period from 12 to 6 months, which mechanically led to a downward effect. This finding is also likely driven by the fact that modifying non-sensitive items to make them mutually exclusive implied adding more restrictive items (tables S1.5 and S1.6 in the supplementary online appendix). Descriptive statistics show that the distribution of the number of items to which farmers answered yes is thus much more concentrated than with the original design (2019 and 2022-V2), in both the control and treatment groups (tables S1.7, S1.8, and S1.9 in the supplementary online appendix). However, the interaction coefficient between the treatment variable and the LE design version indicates that changes in LE design have the same effect on control and treatment groups for the LEs related to harvesting and pod-breaking, and employed and paid child labor. The reason why changes in LE design

**Table 8.** Proportion of Farmers Who Passed the List Experiment Comprehension Test

| | 1st test | | 2nd test | |
|---|---|---|---|---|
| | (1) % | (2) *N* | (3) % | (4) *N* |
| Overall sample | 96.96 | 13,040 | 97.31 | 7,447 |
| **By education level** | | | | |
| No education | 96.83 | 7,727 | 97.38 | 4,737 |
| Primary education or higher | 97.16 | 5,308 | 97.19 | 2,706 |

*Source:* Author's calculation.
*Note:* Column 1 reports the success rate of the 1st list experiment comprehension test, administered in the 2022-V1 and 2022-V2 surveys. Column 3 reports the success rate of the 2nd list experiment comprehension test, which was only administered in the 2022-V1 survey.

are associated with a higher number of items to which the farmers answer yes in the treatment group for child labor related to farm preparation remains, however, unclear.

Overall, findings provide evidence that changes in LE design (length of recall period, non-sensitive items, and addition of a comprehension test before LEs) are strongly associated with a lower average number of items to which farmers respond yes. This is likely driven by the reduction of the recall period and the addition of more restrictive non-sensitive items. However, this downward trend in LE responses is similarly observed across farmers in both the control and treatment groups. That is, while changes in the LE design are associated with a reduction in the number of average items to which farmers respond yes, this reduction is uniform between the control and treatment groups, thus not affecting the difference in means between the two groups, i.e., the LE estimates.

Another insightful finding is that using the LE across years to measure the same sensitive behavior can lead to different results. In this study setting, it is unclear why being surveyed in 2022 versus 2019 is associated with a reduction in positive answers provided to the LEs only among farmers in the treatment group. This may be driven by a structural downward trend in the use of child labor, or differences in LE survey training and implementation across the years.

#### 4.2.3. List Experiment and Cognitive Load

The inclusion of two new survey waves also provided an opportunity to test the understanding of the LE method on a similar sample to 2019, i.e., certified cocoa farmers. For this purpose, the new questionnaire included either two (2022-V1) or one (2022-V2) comprehension test for the LE. Tests are conducted as follows: a list experiment is used to ask cocoa farmers about a non-sensitive topic, followed by direct questions for each item used in the LE test. The number of items the farmer answered yes to when asked the LE was compared to the number of direct questions to which the farmer answered yes. If both numbers are equal, then it is considered that the cocoa farmer has understood the LE method correctly. If answers differ, it is assumed that the farmer misunderstood the LE or made a mistake while answering one of the two questioning methods. The wording and proposed items for each of the LE comprehension tests are presented in the supplementary online appendix (table S1.10).

The test results indicate a high rate of understanding of the LE method by the cocoa farmers, around 97 percent (table 8). Although these tests were not conducted during the 2019 survey, they provide evidence of farmers' ability to understand the LE method, even though the cognitive load is greater than that of direct questions. Descriptive statistics also show no particular difference in LE comprehension between cocoa farmers who have never been to school and those with primary education or higher.

## 5. Conclusions

Eradicating child labor remains a major challenge for both public and private sectors. To provide effective and well-targeted policies, as well as to monitor progress and evaluate the effectiveness of interventions aimed at reducing child labor use, it is essential to measure this phenomenon accurately. Yet measuring child labor use is challenging given its sensitive nature: misreporting the use of child labor is likely common, given the social, economic, and legal sanctions that one who relies on child labor can be exposed to. This paper relies on the list experiment approach, implemented among 4,458 Ivorian cocoa households to estimate the proportion of farmers using child labor on cocoa farms certified as child-labor-free. This indirect survey method is used to overcome the social desirability bias associated with direct questions related to hazardous child labor, which is prohibited both by national legislation and cocoa certification schemes. Two additional survey rounds are used to test the robustness of the estimates to LE design changes, and assess farmers' understanding of this indirect questioning method.

LE estimates show that 24 percent of certified cocoa farmers were helped by at least one child under 16 for harvesting and breaking the cocoa pods during the 2019 cocoa season, 21 percent for preparing their farm, and 25 percent employed and paid a child to perform any task on their cocoa farm. These results are twice as high as those declared by farmers when directly questioning them about their child labor use, indicating a social desirability bias among cocoa farmers when asked about such a sensitive topic. Through the implementation of a comprehension test of the LE method, it is shown that despite its higher cognitive load than direct questioning, the LE is well understood by farmers. Findings further provide evidence that, in this setting, if changes in LE design are associated with a lower average number of LE items to which farmers respond yes, this reduction is uniform between the control and treatment groups, thus not affecting the LE estimates.

This paper's results are of interest for three reasons. First, this study highlights an additional source of error associated with measuring child labor use that has so far received little attention in the existing literature. The findings support the existence of a social desirability bias associated with directly questioning certified cocoa farmers about their use of child labor on their farms, in a context where hazardous child labor is prohibited by both the government and the certifying schemes. The results suggest that direct questioning leads to underestimating the prevalence of child labor use by at least 50 percent.

Second, the findings of this study highlight that in Côte d'Ivoire, certified cocoa farmers are using child labor despite the ban imposed by the government and certification schemes, even to perform dangerous tasks on their farms. This raises questions about the effectiveness of certification schemes to reduce child labor in the cocoa sector, and the monitoring processes implemented by certification bodies.

Last, results from the robustness tests suggest that the LE method, which is cognitively more demanding than direct questioning, can be understood even by populations with low levels of education, if implemented in a way that lowers issues of misunderstanding. In this study setting, respondents were not required to remember or add up their positive responses to the LE items. Instead, they only had to raise a finger behind their back when an item applied to their case, and show their hand to the enumerator at the end of the LE question. Adding up the respondent's positive answers was the sole responsibility of the enumerator.

Several caveats are in order. First, while direct questioning provides a lower bound on the use of child labor at an individual level, indirect questioning methods can provide a broader identification of farmers using child labor at an aggregate level. Nevertheless, this method is probably more appropriate for measuring a simple binary phenomenon, i.e., whether a farmer relies on child labor, than for more complex measures of child labor use, such as the number of hours worked by children, due to a higher cognitive load than the direct questioning method. Second, we were not able to further validate the measures obtained through indirect questioning methods with objective measures of child labor. The inability to do so represents the main limitation of this study. Such an exercise would make it possible to assess

the remaining gap between the true value of child labor use and the estimates obtained using the LE approach.

## Data Availability

Data are available from the author upon reasonable request and with permission of Touton S.A.

## References


Arentoft, A., K. Van Dyk, A. D. Thames, P. Sayegh, N. Thaler, D. Schonfeld, J Labrie *et al.* 2016. "Comparing the Unmatched Count Technique and Direct Self-Report for Sensitive Health-Risk Behaviors in HIV+ Adults." *AIDS Care - Psychological and Socio-Medical Aspects of AIDS/HIV* 28(3): 370–75.

Arthi, V., K. Beegle, J. De Weerdt, and A. Palacios-López 2018. "Not Your Average Job: Measuring Farm Labor in Tanzania." *Journal of Development Economics* 130: 160–172.

Astrid Fenger, A., N. Skovmand Bosselmann, R. Asare, and A. de Neergaard 2017. "The Impact of Certification on the Natural and Financial Capitals of Ghanaian Cocoa Farmers." *Agroecology and Sustainable Food Systems* 41(2): 143–66.

Bardasi, E., K. Beegle, A. Dillon, and P. Serneels 2011. "Do Labor Statistics Depend on How and to Whom the Questions Are Asked? Results from a Survey Experiment in Tanzania." *World Bank Economic Review* 25(3): 418–47.

Bargain, O., and D. Boutin 2021. "Minimum Age Regulation and Child Labor: New Evidence from Brazil." *World Bank Economic Review* 35(1): 234–60.

Basu, K., S. Das, and B. Dutta 2010. "Child Labor and Household Wealth: Theory and Empirical Evidence of an Inverted-U." *Journal of Development Economics* 91(1): 8–14.

Beegle, K., C. Carletto, and K. Himelein 2012. "Reliability of Recall in Agricultural Data." *Journal of Development Economics* 98(1): 34–41

Beegle, K., R. Dehejia, and R. Gatti 2009. "Why Should We Care about Child Labor? The Education, Labor Market, and Health Consequences of Child Labor." *Journal of Human Resources* 44(4): 871–89.

Bhalotra, S., and C. Heady 2003. "Child Farm Labor: The Wealth Paradox." *World Bank Economic Review* 17(2): 197–227.

Blair, G., A. Coppock, and M. Moor 2020. "When to Worry about Sensitivity Bias: A Social Reference Theory and Evidence from 30 Years of List Experiments." *American Political Science Review* 114(4): 1297–315.

Blair, G., and K. Imai 2012. "Statistical Analysis of List Experiments." *Political Analysis* 20(1): 47–77.

Borgers, N., E. De Leeuw, and J. Hox 2000. "Children as respondents in survey research: Cognitive development and response quality 1." *Bulletin of Sociological Methodology* 66(1): 60–75.

Boutin, D., and M. Jouvin 2022. "Child Labour Consequences on Education and Health: A Review of Evidence and Knowledge Gaps." Bordeaux Economics Working Papers No. 2022-14.

Buono, C., and A. Babo 2013. "Travail des Enfants Dans les Exploitations de Cacao en Côte d'Ivoire. Pour une Réconciliation entre Normes Locales et Normes Internationales Autour du bic du Balai et de la Machette." *Mondes en Developpement* 41(3): 69–84.

Çarkoğlu, A., and S. E. Aytaç 2015. "Who Gets Targeted for Vote-Buying? Evidence from an Augmented List Experiment in Turkey." *European Political Science Review* 7(4): 547–66.

Chen, L., X. Zhao, O. Tang, L. Price, S. Zhang, and W. Zhu 2017. "Supply Chain Collaboration for Sustainability: A Literature Review and Future Research Agenda." *International Journal of Production Economics* 194(1): 73–87.

Chuang, E., P. Dupas, E. Huillery, and J. Seban 2021. "Sex, Lies, and Measurement: Consistency Tests for Indirect Response Survey Methods." *Journal of Development Economics* 148(1):102582.

Cigno, A. 2009. "The Economics of Child Labor."In *Labor and Employment Law and Economics*. Edward Elgar Publishing.

Co Man, K. B., L. C. Co Man, and M. M. E. Keith 2017. "The Size of the LGBT Population and the Magnitude of Antigay Sentiment Are Substantially Underestimated." *Management Science* 63(10): 3168–86.

Comşa, M., and C. Postelnicu 2013. "Measuring Social Desirability Effects on Self-Reported Turnout Using the Item Count Technique." *International Journal of Public Opinion Research* 25(2): 153–72.
Corstange, D. 2010. "Vote Buying under Competition and Monopsony: Evidence from a List Experiment in Lebanon." *Paper presented at the Annual Conference of the American Political Science Association*, Washington, DC.
Cullen, C., 2020. *Method matters: Underreporting of intimate partner violence in Nigeria and Rwanda. Policy research working paper 9274*. Washington DC: World Bank.
De Cao, E., and C. Lutz 2018. "Sensitive Survey Questions: Measuring Attitudes regarding Female Genital Cutting through a List Experiment." *Oxford Bulletin of Economics and Statistics* 80(5): 871–92.
de Janvry, A., C. McIntosh, and E. Sadoulet 2015. "Fair Trade and Free Entry: Can a Disequilibrium Market Serve as a Development Tool?" *Review of Economics and Statistics* 97(3): 567–73.
Dillon, A. 2010. Measuring Child Labor: Comparisons between Hours Data and Subjective Measures.In *Child labor and the transition between school and work*, 135–159. Emerald Group Publishing Limited.
Dumas, C. 2013. "Market Imperfections and Child Labor." *World Development* 42(1): 127–42.
Edmonds, E. V., and M. Shrestha 2012. "The Impact of Minimum Age of Employment Regulation on Child Labor and Schooling." *IZA Journal of Labor Policy* 1(1): 1–28.
Fafchamps, M., and J. Wahba 2006. "Child Labor, Urban Proximity, and Household Composition." *Journal of Development Economics* 79(2): 374–97.
Fekadu, D., A. Alem, and B. Hägglöf 2006. "The Prevalence of Mental Health Problems in Ethiopian Child Laborers." *Journal of Child Psychology and Psychiatry and Allied Disciplines* 47(9): 954–59.
Fountain, A., and F. Huetz-Adams 2020. *Cocoa Barometer 2020*. SÜDWIND eV-Institut für Ökonomie und Ökumene.
Fuchs, M. 2009. "The reliability of children's survey responses. The impact of cognitive functioning on respondent behavior". In *Symposium 2008: Data Collection: Challenges, Achievements and New Directions*. Statistics Canada, Ottawa: Statistics Canada.
Gaddis, I., G. Oseni, A. Palacios-Lopez, and J. Pieters 2021. "Measuring Farm Labor: Survey Experimental Evidence from Ghana." *World Bank Economic Review* 35(3): 604–34.
Galdo, J., A. C. Dammert, and D. Abebaw 2018. "Child Labor Measurement in Agricultural Households: Seasonality, Proxy Respondent and Gender Information Gaps in Ethiopia." GLM | LIC Working Paper No. 47.
———. 2021. "Gender bias in agricultural child labor: Evidence from survey design experiments." *The World Bank Economic Review* 35(4): 872–891.
Glynn, A. N. 2013. "What Can We Learn with Statistical Truth Serum?" *Public Opinion Quarterly* 77(S1): 159–72.
Gonzalez-Ocantos, E., C. K. de Jonge, C. Meléndez, J. Osorio, and D. W. Nickerson 2012. "Vote Buying and Social Desirability Bias: Experimental Evidence from Nicaragua." *American Journal of Political Science* 56(1): 202–17.
Guarcello, L., I. Kovrova, S. Lyon, M. Manacorda, and F. C. Rosati 2010. *"Towards Consistency in Child Labour Measurement: Assessing the Comparability of Estimates Generated by Different Survey Instruments."* UCW Working Paper 54.
Guarcello, L., S. Lyon, and F. Rosati 2004. *"Impact of Working Time on Children's Health."* UCW Working Paper 43916.
Guarcello, L., F. Mealli, and F. C. Rosati 2009. "Household Vulnerability and Child Labor: The Effect of Shocks, Credit Rationing, and Insurance." *Journal of Population Economics* 23(1): 169–98.
Holbrook, A. L., and J. A. Krosnick 2010. "Social Desirability Bias in Voter Turnout Reports: Tests Using the Item Count Technique." *Public Opinion Quarterly* 74(1): 37–67.
Imai, K. 2011. "Multivariate Regression Analysis for the Item Count Technique." *Journal of the American Statistical Association* 106(494): 407–16.
Ingram, V., F. van Rijn, Y. Waarts, and H. Gilhuis 2018. "The Impacts of Cocoa Sustainability Initiatives in West Africa." *Sustainability (Switzerland)* 10(11): 4249.
Janzen, S. A. 2018. "Child Labour Measurement: Whom Should We Ask?" *International Labour Review* 157(2): 169–91.
Knößlsdorfer, I., J. Sellare, and M. Qaim 2021. "Effects of Fairtrade on Farm Household Food Security and Living Standards: Insights from Côte d'Ivoire." *Global Food Security* 29(1):100535.
Krebs, C. P., C. H. Lindquist, T. D. Warner, B. S. Fisher, S. L. Martin, and J. M. Childers 2011. "Comparing Sexual Assault Prevalence Estimates Obtained with Direct and Indirect Questioning Techniques." *Violence against Women* 17(2): 219–35.

Lax, J. R., J. H. Phillips, and A. F. Stollwerk 2016. "Are Survey Respondents Lying about Their Support for Same-Sex Marriage? Lessons from a List Experiment." *Public Opinion Quarterly* 80(2): 510–33.

Lépine, A., C. Treibich, C. T. Ndour, K. Gueye, and P. Vickerman 2020. "HIV Infection Risk and Condom Use among Sex Workers in Senegal: Evidence from the List Experiment Method." *Health Policy and Planning* 35(4): 408–15.

Lonie, S., M. Martinez, R. Oulai, and C. Tullis 2018. *Opportunities for Digital Financial Services in the Cocoa Value Chain in Côte d'Ivoire*. World Bank.

Meemken, E. M. 2020. "Do Smallholder Farmers Benefit from Sustainability Standards? A Systematic Review and Meta-Analysis." *Global Food Security* 26(1):100373.

Moseson, H., M. Massaquoi, C. Dehlendorf, L. Bawo, B. Dahn, Y. Zolia, E Vittinghoff *et al.* 2015. "Reducing Underreporting of Stigmatized Health Events Using the List Experiment: Results from a Randomized, Population-Based Study of Abortion in Liberia." *International Journal of Epidemiology* 44(6): 1951–58.

Nkamleu, G. B., and A. Kielland 2006. "Modeling Farmers' Decisions on Child Labor and Schooling in the Cocoa Sector: A Multinomial Logit Analysis in Côte d'Ivoire." *Agricultural Economics* 35(3): 319–33.

Oya, C., F. Schaefer, and D. Skalidou 2018. "The Effectiveness of Agricultural Certification in Developing Countries: A Systematic Review." *World Development* 112(1):282–312.

Putnick, D. L., and M. H. Bornstein 2015. "Is Child Labor a Barrier to School Enrollment in Low- and Middle-Income Countries?" *International Journal of Educational Development* 41(1):112–20.

Raghavarao, D., and W. T. Federer 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.

Roggero, P., V. Mangiaterra, F. Bustreo, and F. Rosati 2007. "Erratum: The Health Impact of Child Labor in Developing Countries: Evidence from Cross-Country Data (American Journal of Public Health (2007) 97, (271-275) DOI: 10.2105/AJPH.2005.066829." *American Journal of Public Health* 97(3): 393.

Sadhu, S., K. Kysia, L. Onyango, C. Zinnes, S. Lord, A. Monnard, and I. R. Arellano 2020. *Assessing Progress in Reducing Child Labor in Cocoa Production in Cocoa Growing Areas of Côte d'Ivoire and Ghana*. NORC: University of Chicago.

Sellare, J., E. M. Meemken, C. Kouamé, and M. Qaim 2020. "Do Sustainability Standards Benefit Smallholder Farmers Also When Accounting for Cooperative Effects? Evidence from Côte d'Ivoire." *American Journal of Agricultural Economics* 102(2): 681–95.

Tadesse, G., G. T. Abate, and T. Zewdie 2020. "Biases in Self-Reported Food Insecurity Measurement: A List Experiment Approach." *Food Policy* 92(1): 101862.

Tourangeau, R., and T. Yan 2007. "Sensitive Questions in Surveys." *Psychological Bulletin* 133(5): 859–83.

Trinh, T. A. 2020. "Mental Health Impacts of Child Labour: Evidence from Vietnam and India." *Journal of Development Studies* 56(12): 2251–65.

Tsuchiya, T., Y. Hirai, and S. Ono 2007. "A Study of the Properties of the Item Count Technique." *Public Opinion Quarterly* 71(2): 253–72.

Udry, C. 2006. "Child Labor."In *Understanding Poverty*. A. V. Banerjee, , R. Bénabou and D. Mookherjee. 243 –58. Oxford: Oxford University Press.

UNICEF. 2018. *Children's Rights in the Cocoa-Growing Communities of Côte d'Ivoire*. Geneva: United Nations.

Wolter, F., and B. Laier 2014. "The Effectiveness of the Item Count Technique in Eliciting Valid Answers to Sensitive Questions. An Evaluation in the Context of Self-Reported Delinquency." *Survey Research Methods* 8(3): 153–68.

World Bank Group. 2019. *Economic Update for Cote d'Ivoire: Cacaoland-How to Transform Cote d'Ivoire*. World Bank.

# Supplementary Online Appendix

## Addressing Social Desirability Bias When Measuring Child Labor Use: An Application to Cocoa Farms in Côte d'Ivoire

**Marine Jouvin**

## Appendix S1. Additional Tables

**Table S1.1.** Descriptive Statistics of Basic Farmer Characteristics

| | *N* | Proportion (%) |
|---|---|---|
| **Age** | | |
| Less than 35 years | 1,237 | 27.75 |
| Between 35 & less than 50 years | 2,195 | 49.24 |
| 50 years and above | 1,026 | 23.01 |
| **Marital status** | | |
| Single | 377 | 8.46 |
| Divorced | 24 | 0.54 |
| Married/in a relationship | 3,975 | 89.17 |
| Widowed | 82 | 1.84 |
| **Education** | | |
| No education | 2,524 | 56.62 |
| Primary education | 1,163 | 26.09 |
| Middle school | 574 | 12.88 |
| High school & higher | 197 | 4.42 |
| **Nationality** | | |
| Ivorian | 3,297 | 73.96 |
| Burkinabe | 1,093 | 24.52 |
| Other | 68 | 1.52 |
| **Children in the household** | | |
| Yes | 4,002 | 89.77 |
| No | 456 | 10.23 |
| **Polygamous household** | | |
| Yes | 727 | 16.31 |
| No | 3,731 | 83.69 |
| **Cocoa income** | | |
| Less than 500,000 XOF | 1,276 | 28.62 |
| Between 500,000 & 1,000,000 XOF | 1,637 | 36.72 |
| More than 1,000,000 XOF | 1,545 | 34.66 |
| **Cocoa farm size** | | |
| Less than 1.5 ha | 1,321 | 29.63 |
| Between 1.5 ha & 3 ha | 1,919 | 43.05 |
| More than 3 ha | 1,218 | 27.32 |
| **Other crop** | | |
| Yes | 2,282 | 51.25 |
| No | 2,171 | 48.75 |
| **Region** | | |
| Agnéby-Tiassa | 195 | 4.37 |
| Cavally | 1,157 | 25.95 |
| Gôh | 322 | 7.22 |
| Grands Ponts | 459 | 10.30 |
| Guémon | 624 | 14.00 |
| Lôh-Djiboua | 776 | 17.41 |
| Nawa | 490 | 10.99 |
| Tonkpi | 435 | 9.76 |

*Source:* Author's calculation.

*Note*: The table reports basic descriptive statistics on the farmers surveyed, about their age, marital status, level of education, nationality, region of residence, number of children in their household, polygamous nature of their household, income from cocoa, size of their cocoa farms, and whether they grow another crop in addition to cocoa. The sample on which each category is calculated and the proportion of farmers belonging to each category are reported.

**Table S1.2.** Basic Characteristics of Respondents across Groups—2022-V1

| | Control | Treatment | Placebo | *p*-value | Obs. |
|---|---|---|---|---|---|
| Age | 44.53 | 44.38 | 44.14 | 0.419 | 7,447 |
| Married | 0.93 | 0.93 | 0.92 | 0.232 | 7,447 |
| No education | 0.64 | 0.64 | 0.62 | 0.430 | 7,447 |
| Ivorian | 0.70 | 0.71 | 0.72 | 0.430 | 7,447 |
| Number of children | 2.35 | 2.35 | 2.32 | 0.767 | 7,447 |
| Number of adults | 3.11 | 3.14 | 3.10 | 0.481 | 7,447 |
| Cocoa income (XOF) | 1,758,013 | 1,737,334 | 1,738,668 | 0.786 | 7,447 |
| Other crop | 0.68 | 0.67 | 0.68 | 0.519 | 7,447 |

*Source:* Author's calculation.
*Note:* The *p*-value comes from the *F*-test of equality of the means across the three groups.

**Table S1.3.** Basic Characteristics of Respondents across Groups—2022-V2

| | Control | Treatment | Placebo | *p*-value | Obs. |
|---|---|---|---|---|---|
| Age | 44.46 | 44.71 | 44.48 | 0.613 | 5,593 |
| Married | 0.94 | 0.94 | 0.95 | 0.633 | 5,593 |
| No education | 0.54 | 0.54 | 0.52 | 0.531 | 5,593 |
| Ivorian | 0.85 | 0.85 | 0.85 | 0.889 | 5,593 |
| Number of children | 2.52 | 2.49 | 2.49 | 0.722 | 5,593 |
| Number of adults | 3.16 | 3.21 | 3.17 | 0.323 | 5,593 |
| Cocoa income (XOF) | 1,613,451 | 1,617,350 | 1,622,078 | 0.964 | 5,593 |
| Other crop | 0.68 | 0.69 | 0.69 | 0.668 | 5,593 |

*Source:* Author's calculation.
*Note:* The *p*-value comes from the *F*-test of equality of the means across the three groups.

**Table S1.4.** Farmers' Use of Labor (Using Direct Questions)

| | Child labor | | | Any type of labor | |
|---|---|---|---|---|---|
| | Harvesting & pod-breaking | Preparing the farm | Employed child labor | Household labor (any tasks) | Employed labor (any tasks) |
| **2019** | | | | | |
| % of yes | 12.76 | 11.28 | — | 76.96 | 66.69 |
| *N* | 2,304 | 2,154 | | 4,458 | 4,458 |
| **2022-V1** | | | | | |
| % of yes | 5.85 | 2.71 | 1.11 | 92.46 | 72.28 |
| *N* | 7,447 | 7,447 | 7,447 | 7,447 | 7,447 |
| **2022-V2** | | | | | |
| % of yes | 3.97 | 0.99 | 0.59 | 92.01 | 76.70 |
| *N* | 3,750 | 3,750 | 3,750 | 5,593 | 5,593 |

*Source:* Author's calculation.
*Note:* In 2019 and 2022-V2, the recall period used for direct questions about child labor is 12 months, while in 2022-V1, the recall period is set at 6 months in order to match that of the list experiment. For columns 4 & 5, recall periods are 12 months.

**Table S1.5.** List Experiments in 2022-V1

| | |
|---|---|
| Harvesting & pod-breaking | Since the beginning of this main crop, for harvesting and breaking the pods:<br>(1) I worked alone, no one helped me.<br>(2) At least one adult in my family helped me.<br>**(3) One or more people under 16 helped me.**<br>(4) The community leader helped me.<br>In total, how many statements apply to your case? |
| Farm Preparation | For this main crop, to prepare my plot (applying pesticides, clearing, felling of trees, burning of plots):<br>(1) I worked alone, no one helped me.<br>(2) At least one adult in my family helped me.<br>**(3) One or more people under 16 helped me.**<br>(4) The community leader helped me.<br>In total, how many statements apply to your case? |
| Employed child labor | Since the beginning of this main crop, to do some work on my farm (for all types of activities):<br>(1) I worked alone, no one helped me.<br>(2) At least one adult in my family helped me.<br>**(3) I have employed and paid one or more people under 16 who are not members of my family.**<br>(4) I employed the community leader.<br>In total, how many statements apply to your case? |

*Source:* Author.

*Note:* For each list experiment, the bolded item is the sensitive one, related to child labor. It is only read to cocoa farmers who belong to the treatment group for that list experiment.

**Table S1.6.** List Experiments in 2022-V2

| | |
|---|---|
| Harvesting & pod-breaking | During the last campaign, for harvesting and breaking the pods:<br>(1) No adult women helped me.<br>(2) At least one adult in my family helped me.<br>**(3) One or more people under 16 helped me.**<br>(4) A group of communal laborers helped me.<br>In total, how many statements apply to your case? |
| Farm Preparation | During the last campaign, to prepare my plot (applying pesticides, clearing, felling of trees, burning of plots):<br>(1) No adult women helped me.<br>(2) At least one adult in my family helped me.<br>**(3) One or more people under 16 helped me.**<br>(4) A group of communal laborers helped me.<br>In total, how many statements apply to your case? |
| Employed child labor | During the last campaign, to do some work on my farm (for all types of activities):<br>(1) I have not employed any adult women.<br>(2) At least one adult in my family helped me.<br>**(3) I have employed and paid one or more people under 16 who are not members of my family.**<br>(4) I employed a group of communal laborers.<br>In total, how many statements apply to your case? |

*Source:* Author.

*Note:* For each list experiment, the bolded item is the sensitive one, related to child labor. It is only read to cocoa farmers who belong to the treatment group for that list experiment.

**Table S1.7.** Distribution of Farmers according to Their Answer to the List Experiment—"Harvesting and Pod-Breaking"

| | 2019 | | 2022-V1 | | 2022-V2 | |
|---|---|---|---|---|---|---|
| | Control | Treatment | Control | Treatment | Control | Treatment |
| 0 | 3.43 | 2.41 | 2.13 | 2.42 | 0.22 | 0.27 |
| 1 | 29.56 | 25.77 | 87.82 | 82.05 | 28.45 | 27.17 |
| 2 | 52.86 | 49.30 | 9.86 | 15.02 | 65.96 | 66.43 |
| 3 | 14.15 | 12.58 | 0.19 | 0.52 | 5.37 | 6.13 |
| 4 | — | 9.94 | — | 0 | — | 0 |
| *N* | 2,304 | 2,154 | 2,586 | 2,484 | 1,842 | 1,844 |

*Source:* Author's calculation.

*Note:* The table shows the distribution of éfarmers according to their response to the list experiment concerning harvesting and pod-breaking, for each survey: 2019, version 1 (V1) of 2022 and version 2 (V2) of 2022. For example: 3.43% of farmers belonging to the control group in 2019 had answered "yes" to 0 item when being asked the list experiment question.

**Table S1.8.** Distribution of Farmers according to Their Answer to the List Experiment—"Farm Preparation"

| | 2019 | | 2022-V1 | | 2022-V2 | |
|---|---|---|---|---|---|---|
| | Control | Treatment | Control | Treatment | Control | Treatment |
| 0 | 3.02 | 2.91 | 2.52 | 2.94 | 0.31 | 0.27 |
| 1 | 31.66 | 28.04 | 87.21 | 84.57 | 26.64 | 27.58 |
| 2 | 45.17 | 43.01 | 10.22 | 12.41 | 65.44 | 64.60 |
| 3 | 20.15 | 14.02 | 0.04 | 0.08 | 7.60 | 7.55 |
| 4 | — | 12.02 | — | 0 | — | 0 |
| *N* | 2,154 | 2,304 | 2,377 | 2,586 | 1,907 | 1,842 |

*Source:* Author's calculation.

*Note:* The table shows the distribution of éfarmers according to their response to the list experiment concerning farm preparation, for each survey: 2019, version 1 (V1) of 2022 and version 2 (V2) of 2022. For example: 3.02% of farmers belonging to the control group in 2019 had answered "yes" to 0 item when being asked the list experiment question.

**Table S1.9.** Distribution of Farmers According to Their Answer to the List Experiment - "Employed Child Labor"

| | 2019 | | 2022-V1 | | 2022-V2 | |
|---|---|---|---|---|---|---|
| | Control | Treatment | Control | Treatment | Control | Treatment |
| 0 | 3.43 | 2.74 | 2.78 | 1.64 | 0.38 | 0.16 |
| 1 | 26.65 | 26.42 | 85.55 | 85.70 | 26.63 | 23.75 |
| 2 | 48.96 | 41.27 | 11.63 | 12.54 | 61.44 | 63.56 |
| 3 | 20.96 | 14.16 | 0.04 | 0.13 | 11.55 | 12.48 |
| 4 | | 15.41 | | 0 | | 0.05 |
| N | 2,304 | 2,154 | 2,484 | 2,586 | 1,844 | 1,907 |

*Source:* Author's calculation.

*Note:* The table shows the distribution of éfarmers according to their response to the list experiment concerning the employment of children, for each survey: 2019, version 1 (V1) of 2022 and version 2 (V2) of 2022. For example: 3.43% of farmers belonging to the control group in 2019 had answered "yes" to 0 item when being asked the list experiment question.

**Table S1.10.** Wording of the List Experiments Tests and Proposed Items

| 1st test | 2nd test |
|---|---|
| During the past week : | During the past week : |
| (1) You harvested pods on your cocoa farm. | (1) You bought pesticides. |
| (2) You helped a neighbor on his cocoa farm. | (2) You received shade trees. |
| (3) You cleared your cocoa farm. | (3) You bought some fertilizer. |
| In total, how many statements apply to your case? | In total, how many statements apply to your case? |

*Source:* Author.

*Note:* The equivalent direct questions followed the same wording as the proposed items for each list experiment. For example, regarding the 1st item of the 1st list experiment test, the direct question was : "During the past week, have you harvested pods on your cocoa farm?".