Measuring Child Labor: the Who's, the Where's, the When's, and the Why's

Guilherme Lichand and Sharon Wolf*

February 15, 2023

Abstract: Measuring child labor accurately is a major challenge. While parents' and children's reports tend to differ dramatically, there is typically no way to verify whose reports are truthful (if any). To overcome this challenge, this paper uses novel data from a cocoa certifier in Côte d'Ivoire that draws on satellite imagery to minimize under-reporting. Concretely, aerial photos allow them (1) to select remote and hard-to-reach communities, where parents typically have not been sensitized by government or NGOs, averting social desirability biases; and (2) to visit these communities while cocoa is being harvested, precisely when children in employment are very visible, making it easier for enumerators to impute it if parents still fail to report it. We compare their figures with those obtained from business-as-usual surveys with parents and children in these regions. We find that adults dramatically under-report child labor relative to the certifier data, by a factor of at least 60%; in turn, children self-reports are statistically identical to the latter. Taking advantage of an experiment that randomly assigned a text-message campaign to discourage child labor, we further show that parents' reports not only underestimate its prevalence, but can even lead to the wrong conclusions about the effects of policy interventions.

*Lichand: Department of Economics, University of Zurich (guilherme.lichand@econ.uzh.ch); Wolf: Graduate School of Education, University of Pennsylvania (wolfs@upenn.edu). Study conducted in partnership with Transforming Education in Cocoa Communities (TRECC) and generously funded by the Jacobs Foundation Science of Learning initiative. Any views and opinions contained in this paper are those of the authors and do not necessarily reflect the views or opinions of TRECC or the Jacobs Foundation. We thank Maxime Deroubaix and Michael Kra from ENVERITAS for explaining the survey methodology in detail and kindly granting access to their geo-referenced child labor data. Guilherme is a partner and chairman at Movva, the implementing partner of the intervention evaluated as part of this study. We acknowledge helpful comments from Andrew Dillon and Megan Passey (International Cocoa Initiative), excellent research assistance by Laura Ogando and Gabriel de Campos, and field supervision by Nicolo Tomaselli and Innovations for Poverty Action. All remaining errors are ours.

1 Introduction

Child labor is a pervasive practice in agriculture, especially in West Africa – where the global cocoa industry sources roughly half of its produce. 2020 data from the International Labour Organization (ILO) documented that 160 million children worldwide were active workers, – 9.6% of all 5-17 year-old children (ILO and UNICEF, 2021). Strikingly, that figure was over 2-fold in Sub-Saharan Africa, affecting 23.9% of children. In Côte d'Ivoire, the setting of our study, 15% of the cocoa industry employees that year were children (Sadhu et al., 2020). Although already extremely high, those numbers might have been even higher before: according to Save the Children, child labor plummeted by 38% globally between 2000 and 2016, declining from 246 to 152 million (Christian Science Monitor, 2019).

Are these figures and trends, however, reliable? There are two reasons to believe that the answer is negative. The first is conceptual: parents[1] have no incentives to truthfully disclose children in employment. They might fear that admitting to child labor, actively discouraged since decades by several NGOs and multilateral organizations, could adversely impact their livelihoods, either directly (e.g., if they face legal action from child protective services) or indirectly (e.g., if the companies sourcing from them are punished for child labor in their supply chain, which could trickle down to lower prices or even to discontinuation of their income); see Lépine et al. (2022). Since child labor is *contained* in children in employment according to the ILO methodology (International Labour Organization, 2017), if parents in fact under-report children in employment, then child labor will necessarily be under-estimated in official statistics.

The second reason is empirical: there are striking differences between children's and adults' reports in the few settings where both have been asked independently about children in employment. According to NORC (a research institute based at the University of Chicago), which surveys children directly about the number of hours worked in cocoa fields, 38% of 5-17 year olds in Côte d'Ivoire reported to have worked in 2018-19 (Sadhu et al., 2020); in contrast, the ILO figure for 2016 – based on adult reports for children employment – was only 23% (ILO and UNICEF, 2021). Even worse, different sources tell very different stories about child labor's recent trends. While ILO data indicate a 38% decrease in child labor worldwide since 2000 (Christian Science Monitor, 2019), NORC data record a nearly 65% *increase* in child labor since 2008-09 (Sadhu et al., 2020).

While these differences are suggestive that official statistics based on adult reports are biased, it is hard to be sure.[2] In particular, in the absence of verification, it is unclear whether children's self-reports do not suffer from reporting biases too. This paper uses novel data from a cocoa certifier in Côte d'Ivoire to overcome this challenge. ENVERITAS, a global NGO that certifies smallholder coffee and cocoa farms when it comes to agricultural best practices – including the absence of child labor –, draws on satellite imagery to minimize under-reporting. Concretely, aerial photos allow them (1) to select remote and hard-to-reach communities, where parents typically have not been sensitized by government or NGOs, averting social desirability biases; and (2) to visit these communities while cocoa is being harvested, precisely when child labor is very visible, making it easier for enumerators to mark it as present even if parents still fail to report it. We compare their figures with those obtained from business-as-usual surveys with parents and children in these regions.

[1]Throughout the text, we use the term 'parent' to refer to the child's primary adult caregiver.

[2]Besides differences in whom is surveyed, discrepancies across surveys might also accrue to differences in their geographical coverage or in the timing of data collection, or to other methodological differences.

Our main contribution is to document first-hand that adult surveys indeed under-report the prevalence of child labor, and the extent of under-reporting. In regions with subsequent third-party verification, 45.5% of children reported having worked in cocoa plantations in the previous month, matching almost exactly the 44.4% prevalence indicated by the certifier; in contrast, only 16.2% of parents in those regions reported children in employment – a nearly 2/3 reporting gap. Across regions, under-reporting ranged from 60% to 85%.

We also provide supporting evidence to the certifier claim that more remote communities are less sensitive to reporting biases. Taking advantage of households' GPS location, we show that while distance to the school (our measure of *remoteness*) is not systematically associated with children's self-reports, it significantly decreases discrepancies between children's and parent's reports: every extra km away from school decreases that difference by about 7% of the average discrepancy.

Previous studies were not able to document the extent of under-reporting by parents. These papers rely on three types of comparisons. First, comparisons between parents' reports under different conditions for social desirability bias (e.g., Jouvin, 2021), to document that such biases exist – but unable to pin down their magnitude. Second, comparisons between adults' and children's reports (Dillon, 2010; Galdo et al., 2020), to document that discrepancies exist – but unable to pin down what the 'ground truth' is. Third, comparisons between an objective measure of children in employment and children's self-reports (Dillon et al., 2017), to document that the latter are accurate – but unable to pin down under-reporting by parents. Similarly to ours, the latter study documents that children accurately report the number of hours they work, using logs from GPS trackers worn by different household members to verify data from surveys and activity diaries. The study, however, actively refrained from having enumerators ask adults about child labor. In the absence of parents' reports, it could not document whether parents under-report children in employment and, if so, by how much.

Our second contribution is to document that basing child labor accounts on surveys with parents not only underestimates its prevalence, but can also bias evaluations of how it responds to policy interventions. Taking advantage of a randomized control trial that assigned some Ivorian parents to messages discouraging child labor in cocoa fields to study whether the estimated impact of the intervention on child labor depends on how the latter is measured, we find that while messages had no effect on children in employment according to children themselves, they significantly increased children in employment according to parents (by 55.1%). Presumably, the reason for such discrepancy is that the intervention tried to foster investments in children that would reduce children's participation in labor activities *without explicitly condemning child labor* – partially deterring social desirability biases. Once again consistent with the claim that such biases are stronger in communities that have been previously sensitized, treatment effects on child labor according to parents sharply decay with our measure of remoteness.

These findings have key implications for how child labor should be measured. One possibility is to survey children directly (the '*who*'). Our results indicate that this would most likely yield accurate estimates of children in employment, but it might also involve complex technical and ethical dilemmas. It might be hard to ensure a unified understanding of what characterizes employment, especially among younger respondents. Most importantly, participation might put children at risk if it triggers backlash by parents. An alternative is to survey adults, as usual, but to focus on hard-to-reach communities (the '*where*') during harvest season (the '*when*') – leveraging technological advances such as satellite imagery. We further discuss the implications and limitations of our findings in Section 6.

2 Background

2.1 How child labor is measured

2.1.1 ILO methodology

The International Labour Organization (ILO) follows the Convention on the Rights of the Child, the ILO Minimum Age for Admission to Employment Convention (No. 138), and the ILO Worst Forms of Child Labour Convention (No. 182) to define child labor (ILO and UNICEF, 2021). According to these conventions, whether children in employment characterizes child labor depends on the child's age, the number of hours dedicated to work, and the work conditions. For children less than 11 years old, any employment characterizes child labor. For those between 12 to 14 years old, 15 or more weekly work hours or hazardous work conditions characterize child labor. Last, for those between 15 to 17 years old, child labor applies in case of 43 or more weekly work hours or hazardous work conditions. Appendix A illustrates all conditions used by ILO to define child labor for children of different age groups.

Statistics on children in employment, number of hours and work conditions come from different surveys around the globe. ILO does not collect the data itself, but rather harmonizes data from these different sources to compute the prevalence of child labor according to its methodology. All leading international organizations follow ILO's methodology strictly or closely. For example, UNICEF adapted its Multiple Indicator Cluster surveys after 2013 to match ILO guidelines. The World Bank also tracks child labor following the same methodology, only for 7-14 year-olds rather than 5-17 year-olds.

Child labor is computed based on surveys with adults and children. Importantly, ILO's methodology *only uses the adult questionnaire* to compute children in employment. In turn, the children questionnaire is used to assess hazardous work conditions *only among those who are employed* ("[a]s in the previous rounds, the current round of the Global Estimates of Child Labor uses data obtained from the adult questionnaire, except for conditions of work, where the information from the child questionnaire is deemed to be more reliable"; International Labour Organization, 2017, p. 59). In other words, child labor is a subset of children in employment – which is exclusively based on parents' reports. As such, official statistics on child labor depend crucially on the accuracy of the latter.

2.1.2 NORC methodology

NORC, a research institution at the University of Chicago, has tracked child labor in the cocoa industry for Ghana and Côte d'Ivoire since 2015, building on Tulane University's work in the region dating back to 2008. It reports statistics associated with child labor, for cocoa production in particular and for agricultural activities more broadly (Sadhu et al., 2020). NORC defines child labor based on the number of work hours and work conditions for children of different age groups, following ILO's methodology exactly (see Appendix A). Different from ILO, however, NORC surveys children about the number of hours worked *directly*, and defines child labor based on *children's self-reports* ("[u]sing the responses of children relating to engagement in cocoa production, we generated estimates of children's engagement in child labor and in hazardous child labor in cocoa production-related activities"; Sadhu et al., 2020, p.61).

Comparing the different data sources – which differ according to the reporting sources used for computing children in employment – is telling. 2016 ILO data for Côte d'Ivoire (`https://ilostat.ilo.org/topics/child-labour/`) indicated that 17.5% of 5-17 year-old children engaged in economic activity and household chores. In contrast, NORC data indicated that, as recently as 2018-19, 64% of all 5-17 year-olds in cocoa-producing regions of Côte d'Ivoire worked in the past 7 days, and 78% in the past 12 months.

While differences are striking, the NORC surveys cover different geographies and years than those used to compute official statistics, making it hard to attribute the gaps to under-reporting by adults in the ILO, UNICEF and World Bank data. Moreover, even if one would accept that parents report children in employment to a lesser extent than children themselves (see, for instance, Dillon, 2010; Galdo et al., 2020) or that parents' reports are prone to social desirability biases (Jouvin, 2021), it could be that children's self-reports are similarly unreliable. Without additional data to *verify the actual prevalence* of children in employment, one simply cannot tell.

2.1.3 ENVERITAS methodology

ENVERITAS is a not-for-profit NGO that certifies coffee (and, more recently, cocoa) companies by verifying farming practices in their supply chain – from chemical usage to child labor (see, for instance, Tran et al., 2021b). Since 2020, their methodology has been aligned with the ILO standards.

The certifier has specialized in reaching small, hard-to-reach and remote farmers, who often cannot even be located by traditional certifiers. To do that, ENVERITAS relies on partnerships to access fresh acquisitions of 50cm-resolution satellite data filtered for quality (with alleged maximum cloud covers of 15%, "crucial for finding (...) farms in cloudy equatorial regions"; Enveritas, 2020), combined with machine learning models to identify specific crops. Additional details on how such models have been successfully used to map coffee-growing households in different geographies can be found in Tran et al. (2021a). In Côte d'Ivoire, the certifier has applied this methodology to cocoa farmers to identify constraints to quality education and early childhood development – including child labor – in cocoa-growing communities (TRECC, 2021).

The combination of satellite imagery with machine learning models generates GPS coordinates for each ENVERITAS field team. Pins in Google Maps assign plots to be surveyed to each enumerator. Michael Kra, country lead for ENVERITAS, points out that ''*When one surveys farmers, sometimes they do not tell you the truth. The pins will tell us to go where the truth is"*. The idea is that farmers in hard-to-reach communities are less sensitive to social desirability biases. Michael Kra explains: *"We work in very rural communities. The road is often not good, requiring enumerators to use boats to cross rivers, motorcycles, etc. These communities being much less accessible, farmers have been rarely sensitized about child labor. This is why it is easier for them to openly disclose child labor. We can only find these communities because of satellite imagery"*. Appendix A includes a sample satellite picture to illustrate how they are able to locate cocoa plantations based on aerial imagery. The appendix also compiles pictures from enumerators in the field as they survey some of these remote communities, documenting how they often have to cross dirt roads and flooded paths in order to reach them.

ENVERITAS further monitors cocoa harvests in the country with the help of satellite imagery, surveying farmers specifically during harvest season. With harvesting activities ongoing at the time of the survey, the potential presence of child labor becomes apparent: not only it is much harder for parents to falsely deny children in employment when it is more visible, but also, ENVERITAS complements survey data with direct field observation (Tran et al., 2021a). Michael Kra explains how surveying farmers during harvest time make it harder to under-report child labor: *"Enumerators can mark child labor as present even if a parent fails to report it (although this happens in less than 25% of surveys). They can add comments with additional information they have observed in the farm: e.g., 'child is* *helping, climbing cocoa trees' "*.

2.2 Background for this study

2.2.1 Study sample

Our study takes place in the cocoa-producing regions of Aboisso and Bouafle in Côte d'Ivoire. Along with Ghana, the country hosts almost 2/3 of the world's cocoa production. This has been linked to one of the highest incidences of child labor worldwide, with nearly 1.6 million children employed in cocoa fields.[3]

We collected child labor data in the context of a broader research project, focused on evaluating different communication interventions to prevent student dropouts (see Wolf and Lichand, 2023). As such, we focus on a sample of primary school children, all of whom were enrolled at baseline. We discuss the implications of that sampling restriction for the generalizability of our findings in Section 6.

Appendix C provides descriptive statistics for our study sample. Almost all (92%) of participating children are 5-11 years old – for whom any form of employment is considered child labor according to ILO guidelines. Half of children in our sample are girls and live in rural areas (defined according to their parents' income source), and slightly over half of them are enrolled in CP2. Nearly a quarter (22%) of households in our sample are extremely poor – at baseline, they made at most a little over 1 USD a day. For only 18% of households income from all sources was more than 6 USD a day at baseline.

2.2.2 Campaign to discourage child labor

Wolf and Lichand (2023) evaluated an communication intervention (Eduq+, implemented by Brazilian EdTech Movva) that delivered nudges directly to parents' and/or teachers' mobile phones. The intervention was implemented over the entire 2018-19 school year. Nudges were organized in thematic sequences – comprised of four messages –, with two messages delivered each week. Content was catered to each students' age group. Messages tried to encourage parents to participate more actively in their children's school life. There was some emphasis on showing up to school, especially in the context of parent-teacher meetings, and on discouraging harmful practices like corporal punishment as a disciplining strategy.

Several sequences explicitly discouraged child labor in cocoa fields, describing how it might detract from child development and learning. The language was careful, in an attempt to openly discuss the issues without creating stigma or setting social expectations that ultimately make it harder to track whether children work on the fields. To illustrate the approach, during the intervention, parents received a text message stating "*It is important that you child complete her/his education! School can provide a better future not only for her/him, but for your whole family*". A few days later, another text encouraged them to "*Talk to your child about the importance of focusing on her/his education. Equally important as learning family traditions is learning the values and skills that only the school can teach*". See Lichand and Wolf (2022) for additional details of the intervention, including additional examples.

In the experiment, nudges to parents were cross-randomized with nudges to teachers, aimed at increasing their attendance and time-on-task while teaching. For simplicity, in the main text, we focus on discrepancies between parents' and children's reports between the treatment condition that had only parents nudged and the control group (which did

[3] https://foodtank.com/news/2021/02/norc-report.

not receive any messages). Appendix E compiles results for the remainder treatment assignments.

3 Data and outcomes

3.1 Survey data

Our study comprises 198 CP2 and CE2 classrooms (second and fourth grades, respectively) across 99 Ivorian public schools in the cocoa-producing regions of Aboisso and Bouafle. Within each school, we randomly drew 13 CP2 students and 12 CE2 students to be surveyed at baseline (at the beginning of the school year, in October 2018) and end line (at the end of the school year, in June 2019). Importantly, children and parents were surveyed independently. Enumerators ensured that this was the case, especially since we also tested children's numeracy and literacy skills as part of the broader project, which required them to sit by themselves – only accompanied by our survey team.

We also conducted an additional follow-up survey at the beginning of the following school year (in October 2019). In this follow-up, we surveyed all teachers and only one parent per classroom, but no children. This follow-up data focused on collecting additional information about work conditions for children in employment, but it also provides a measure of children in employment according to parents already after the growing season (Yoroba et al., 2019) and much closer to the timeline of the ENVERITAS data collection.

The sample comprises 1,285 CP2 students and 1,190 CE2 students surveyed at baseline along with their primary parents, in addition to their 198 teachers. We were able to track all teachers, 1,157 CP2 students (90.0%) and 1,086 CE2 students (91.3%) at end line. We assigned replacement households in case the ones drawn could not be tracked by enumerators. At the end line, no children – and less than 3.5% of their parents – refused to answer about their employment conditions. We discard child-parent pairs involving refusals, focusing on the 2,500 observations for whom we can compare children's and parents' reports. Out of those, we have information on all baseline characteristics that we use as controls in some of our specifications for 2,246 observations. Furthermore, we have information on home GPS location (which we use to compute a measure of remoteness) for 1,790 of those. Missing data for the latter often involves parents and children whom were surveyed at the school or in some other location.

Appendix B compiles the survey questions related to children in employment in each wave. As indicated, we asked parents and children the same question about children in employment in cocoa fields at both baseline and end line ("*In the last month, have you [any of your children] engaged in one or more of the following activities, for one hour or more? Work in a cocoa plantation*"). This aligned with the data collected by ENVERITAS and NORC, and follows ILO's methodology for measuring children in employment. Since parents are not asked *specifically* about the child from whom we elicited employment information independently, that should decrease our ability to detect under-reporting by parents (e.g., if some answer affirmatively about older children, or about children who are already out of school).

Even though we also surveyed parents and children about other forms of employment (e.g., domestic work and construction), we do not analyze these data in this paper because we have no way to verify such reports. These additional measures of children in employment are described and analyzed in Wolf and Lichand (2023).

3.2 Certifier data

We use data on children in employment collected by ENVERITAS over January 2020, during the harvest season in Aboisso and in two sub-regions of Bouafle (Bouafle 2 and Tiapoum Adiake). ENVERITAS's sampling frame relied on geographical units of 10,000 farmers, identified via satellite imagery. They randomly drew 125 farmers to be surveyed in each unit. 8,150 households were approached by ENVERITAS, 7,402 of which were successfully surveyed. The certifier also surveyed schools in regions outside of our study sample; see Appendix A. We do not use data in these other regions, or collected prior to 2020 (before the survey instrument was consistent with ILO's methodology regarding the definition of child labor).

Adults surveyed by ENVERITAS were asked "*Do any of your children between 6 and 16 years old help you work on the cocoa farm?*". Children were not surveyed directly by ENVERITAS; thus far, they have only piloted surveys with children in Tonkpi, a region outside our study sample.

4 Empirical strategy

4.1 Assessing the accuracy of different reporting sources

We assess the accuracy of children in employment reports according to parents and children by comparing it to the certifier data, both within each region and in the aggregate, considering the three geographical units for which the different data overlap.

We report p-values from tests of differences in proportions of children in employment according to each source, considering equal population variances when comparing parents' and children's reports (through an Ordinary Least Squares regression, given the paired design), and unequal population variances when comparing any of them to certifier data – since the latter was collected from a different sample. We cluster standard errors at the regional level.

4.2 Assessing the certifier claim that remoteness is inversely related to reporting biases

Next, to assess whether remoteness is indeed inversely related to reporting biases, we estimate how children's and parents' reports vary with their household's distance to the school (in km). Appendix C includes the histogram for this measure, computed as the linear distance between the school and home GPS locations. 70% of households with valid home GPS data are within 1km of the school. While some are as far as 20km from the latter, the vast majority are within a 2.5 km radius of it.

We take advantage of natural variation in distance to the school by estimating a linear model for the association between distance and children's reports, and between the former and the discrepancy between children's and parents' reports. When analyzing discrepancies, we also control for the school-level share of children in employment, and allow its coefficient to vary with whether the child has reported to work – all of which might influence the extent of reporting biases.

4.3 Assessing the sensitivity of treatment effects estimates to different reporting sources

To document whether different reporting sources might lead to bias in evaluating the impacts of interventions to discourage child labor, we contrast effect sizes of nudges to parents on the prevalence of children in employment in cocoa fields based on children's self-reports or on adult reports. We also allow treatment effects based on parents' reports to vary with the distance to the school, to assess whether results are consistent with the spatial patterns we document for parents' reporting biases.

Last, Appendix E contrasts the relationship between the classroom-level prevalence of child labor and average standardized test scores comparing, within each treatment cell of the experiment, the slope of the linear association between the two outcomes when child labor is based on children's self-reports and when it is based on parents' reports instead.

5 Results

5.1 Descriptive statistics based on children's self-reports

Appendix C documents the aggregate prevalence of children in employment, and that by student characteristics, according to children's self-reports at baseline. 38.1% of children reported to have worked at least one hour in cocoa fields over the previous month. As a point of comparison, by the end of the school year (closer to harvest season), this figure was up to 50% (see Appendix D). The prevalence of child labor was only slightly higher among fourth graders than that among second graders (39.6% vs. 36.8% at baseline, and 51% vs. 48% at end line). Boys were nearly 50% more likely to work in cocoa fields at baseline than girls (44.4% vs. 31.2%). The baseline prevalence of children in employment was higher for the bottom income bracket (38.9%), but not low even for the top income bracket (23.9%). Naturally, child employment in cocoa fields was much higher in rural areas (52.4% vs. 23.9%). For an account of adult reports about children's work conditions, elicited in the follow-up survey, see Appendix D.

Appendix C also illustrates correlations between classroom-level prevalence of child labor and educational outcomes. Consistent with common sense, classrooms with a higher share of children in employment at baseline feature lower test scores by the beginning of the school year, and higher dropout rates over the course of the school year. Estimating a linear relationship between the variables in each case suggests that moving from 0% to 20% children in employment is associated with about 0.08 s.d. lower test scores – what children tend to learn in one school quarter, and the magnitude of effect sizes of many educational interventions, such as nudges to parents evaluated in this setting (Lichand and Wolf, 2022). Similarly, moving from 0% to 40% children in employment is associated with roughly doubling dropout rates.

While these associations are not causal, they help understand the centrality of the issue for governments and international organizations monitoring children's rights. This is why accurate measurement is key.

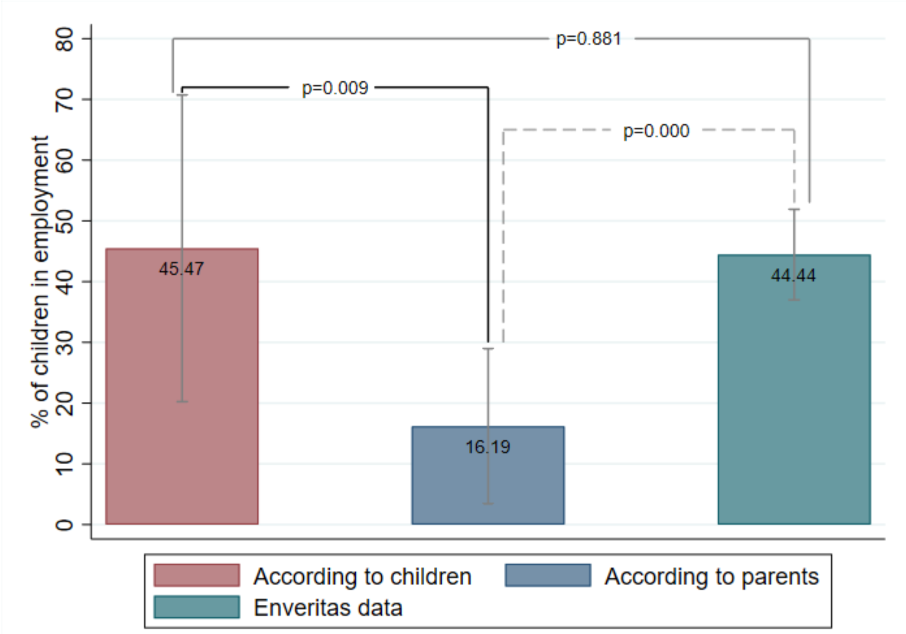
5.2 Validating survey data with certifier data

We now turn to the comparisons between independent reports by parents and children in our end-line survey data, and ENVERITAS data. Figure 1 presents the prevalence of children in employment according to each reporting source, along with p-values for pair-wise statistical tests of differences in proportions. Panel A considers the aggregate prevalence figures for the regions where both data sources overlap, and Panel B documents comparisons within each region.

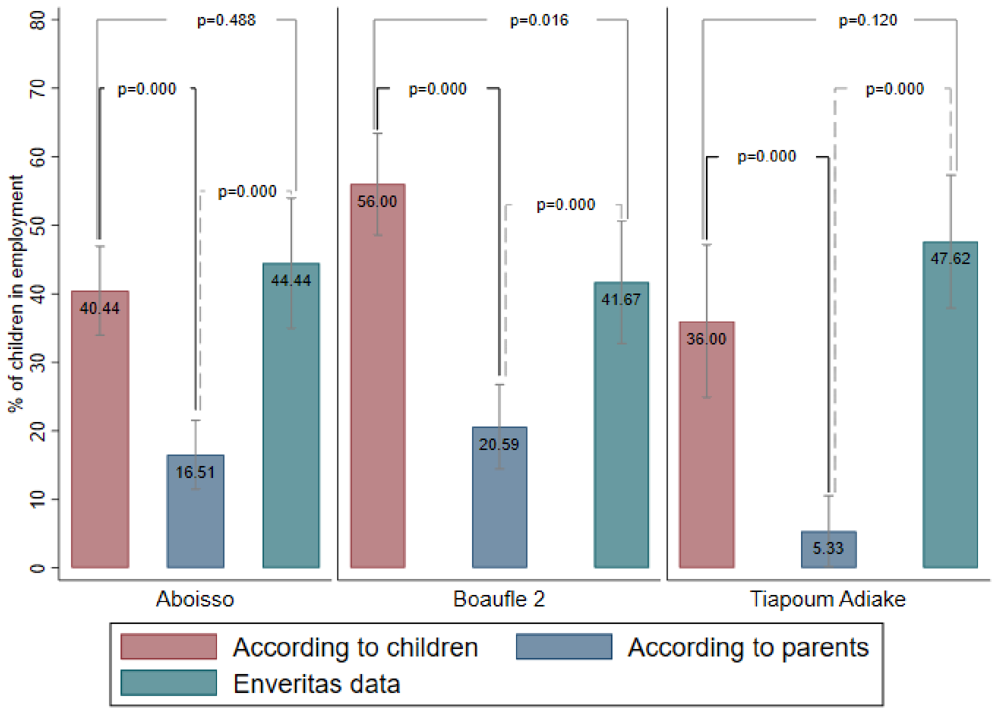
In the absence of under-reporting, the rate of children in employment according to either parents' or children' reports should be identical (in fact, since the question directed to parents was about *any* of their children, the former should be weakly greater than the latter). This is, however, strictly at odds with what we find. In Figure 1, Panel A documents that, in regions with subsequent verification, 45.5% of children reported to have worked in cocoa plantations in the previous month, matching almost exactly the 44.4% prevalence indicated by the certifier (p-value of the difference = 0.881). In contrast, only 16.2% of parents in those regions reported employing children – a nearly 2/3 reporting gap (p = 0.000). Panel B shows that, across regions, under-reporting by adults was striking, ranging from 60% to 85% (p = 0.000 in each case). Children's self-reports, in turn, range from 75.6% to 134% of ENVERITAS figures, and are not statistically different from ENVERITAS data in two out of three regions ($p = 0.488$ in Aboisso, $p = 0.016$ in Bouafle 2, and $p = 0.120$ in Triapoum Adiake).

Figure 1: Validation of child labor measures using third-party data

Panel A: Share of students who worked in cocoa plantations, by reporting source



Panel B: Share of students who worked in cocoa plantations, by source and region



Notes: Bars show the share of children who worked at least an hour in cocoa fields over the previous month, according to children (in red), parents (in blue) and ENVERITAS (in green). Panel A reports the average prevalence across all regions for which survey data overlaps with ENVERITAS data. Panel B breaks down prevalence by region. Children answered the following question at end line: "In the last month, have you engaged in one or more of the following activities, for one hour or more? Work in a cocoa plantation", as described in Table B1. Parents answered the following question at end line: "I will now ask you some questions about activities that your children might have recently performed. In the last month, has any of your children engaged in one or more of the following activities, for one hour or more? Work in a cocoa plantation", as described in Table B1. In the survey conducted by ENVERITAS during the harvest season (identified through satellite imagery), farmers answered the following question: "Do any of your children between 6 and 16 years old help you work on the cocoa farm?". For the measures reported by children and parents, observations are restricted to the control group of the communication intervention. P-values from tests of proportions with unequal population variances (when children's or parents' reports are compared to ENVERITAS data; accounting for each source's intra-cluster correlation computed at the regional level, in Panel A), and from tests of proportions with equal population variances (when comparing children's and parents' reports) through Ordinary Least Squares regressions (clustering standard errors at regional level, in Panel A).

Appendix F in Supplementary Materials uses our estimates to calibrate predictions of the bias-adjusted prevalence of child labor in Côte d'Ivoire and Ghana. Under the stated assumptions, child labor could affect nearly 6.9 million children in these countries, 50% more than in the World Development Indicators' statistics.

5.2.1 Robustness to differences in the timing of the surveys

One potential caveat of the results above is that our end-line survey (when both children and parents were asked about children in employment) dates from June 2019, almost 6 months prior to the ENVERITAS data collection. Here we assess whether such difference matters for our findings.

To do that, we take advantage of our follow-up wave, conducted in October 2019 (much closer to January 2020, when ENVERITAS conducted its survey). While children were not surveyed in that wave, we can compare parents' reports across the end-line and follow-up surveys to gauge whether adult reports converged to children's self-reports as harvest season was approaching. Appendix D documents that this was *not* the case. At the follow-up wave, only 28% of parents admitted that children worked at all in cocoa fields during the school year. This figure was still only about half that reported by children at the end-line survey – even though, at the follow-up wave, we asked parents to report on work performed by children in cocoa fields *at any point during the previous school year*. Appendix D further documents that teacher reports of child labor over the course of the previous school year were statistically identical to parents' reports at the end line.

5.2.2 Spatial patterns of under-reporting

Next, we investigate the association between our measure of remoteness – the linear distance from each household to the school where the child participating in the study was enrolled – with under-reporting by parents. If remote communities are indeed less often or less intensely sensitized, a challenge is that remoteness might affect both under-reporting *and* the true prevalence of child labor. To separate these potential effects, Table 1 starts by assessing whether children's self-reports vary systematically with our measure of remoteness (in column 1); after all, our previous results suggest that this outcome most likely matches the 'true prevalence' of children in employment. Next, columns (2) to (4) assess how the discrepancy between children's and parents' reports varies with that measure. All columns control for children's self-reports; column (3) further controls for the school-level prevalence of child labor (according to children) and its interaction with each child's self-report, as those might also affect the extent of reporting biases; column (4) additionally controls for baseline characteristics. Because we include a school-level measure of prevalence, we cluster standard errors at the school level.

In Table 1, column (1) documents that distance to school is not associated with children's self-reports. The average end-line prevalence of children in employment within the sub-sample with valid home GPS data was 51.8%; such prevalence did not systematically increase with the household's distance to the school. This is consistent with the idea that more remote communities do not necessarily feature more children working in cocoa fields. In turn, columns (2) to (4) document that such distance was systematically associated with lower discrepancies between children's and parents' reports. While, in this sub-sample, the average discrepancy between children's and parent's reports was 22.5 p.p., we estimate that it decreased by 1.6 p.p. (significant at the 5% level) with every km away from school, about 7% of the average discrepancy. Incidentally, other estimated coefficients are also informative: consistent with our previous discussion, the more common children in employment is at the school, the lower the reporting discrepancies – consistent with the role of social expectations. Results support the certifier claim that remoteness is inversely associated with reporting biases, without necessarily being associated with children in employment itself.

Table 1: Association between distance to school, children's self-reports, and adult under-reporting

	Children in employment (self-report)	Discrepancy (child - parent)		
	(1)	(2)	(3)	(4)
Distance to school (km)	0.005	-0.008	-0.015*	-0.016**
	(0.015)	(0.013)	(0.008)	(0.008)
Children in employment (self-report)		0.630***	1.063***	1.071***
		(0.035)	(0.064)	(0.069)
School-level % children in employment (self-reports)			-0.396***	-0.390***
			(0.061)	(0.067)
Children in employment (self-report) × School-level % children in employment (self-reports)			-0.505***	-0.504***
			(0.114)	(0.124)
Dep. variable mean	0.518	0.225	0.225	0.225
Controls (baseline survey)	No	No	No	Yes
R-squared	0.000	0.365	0.454	0.463
Observations	1,790	1,789	1,789	1,608

Notes: In column (1), children in employment (self-report) = 1 if the child answered affirmatively to the question: "In the last month, have you engaged in one or more of the following activities, for one hour or more? Work in a cocoa plantation"; and 0 otherwise. In columns (2) to (4), discrepancy = self-report - parent's report, whereby the latter = 1 if the parent answered affirmatively to the question "I will now ask you some questions about activities that your children might have recently performed. In the last month, has any of your children engaged in one or more of the following activities, for one hour or more? Work in a cocoa plantation"; and 0 otherwise. Distance from school (km) is the linear distance between the home and their child's school GPS locations, in kilometers. School-level % children in employment averages self-reports within each school. Baseline controls include child gender, grade indicators, standardized test scores (averaged across numeracy and literacy); and summary measures of parental engagement, student effort, socio-emotional skills, working memory, visual attention, impulsivity, self-esteem, and mindset (see **?**. Summary measures computed following Kling et al. (2007), standardizing each component by normalizing values by the mean and standard deviation of the control group at the baseline survey within each grade, and then averaging over all standardized components. Sample restricted to observations with valid home GPS coordinates. All regressions estimated through Ordinary Least Squares with standard errors clustered at the school level. *** p<0.01, ** p<0.05, * p<0.10

5.3 How adult reports can also bias policy evaluations

Last, we present evidence that relying on parents' reports for children in employment might not only lead to inaccurate estimates about the prevalence of child labor, but also, to potentially incorrect conclusions about the effects of interventions to discourage it. Table 2 estimates treatment effects of nudges to parents on children in employment at end line, relative to the control group. Column (1) uses children's self-reports as the dependent variable, while columns (2) to (4) use parents' reports. Columns (1) and (2) use the full sample for whom we have children's and parents' reports; columns (3) and (4) restrict attention to those with valid home GPS information. Column (4) allows treatment effects to vary with our measure of remoteness – the linear distance from the household to the school where the participating child was enrolled. The idea is that part of the effects of the intervention on parents' reports might play out through its effects on reporting biases (especially since its goal was to openly address the issue without creating stigma). All columns control for baseline characteristics. Because the over-arching intervention evaluated in Wolf and Lichand (2023) was centered around teachers, we cluster standard errors at the classroom level.

Table 2: Treatment effects of nudges and distance to school

	Children in employment (self-report)	Children in employment (parent's report)		
	(1)	(2)	(3)	(4)
Nudge to parents	0.054	0.089*	0.120**	0.145**
	(0.040)	(0.045)	(0.056)	(0.061)
Distance to school (km)				0.016
				(0.012)
Nudge to parents × Distance				-0.061*
				(0.031)
Control mean	0.498	0.221	0.230	0.230
Controls (baseline survey)	Yes	Yes	Yes	Yes
R-squared	0.115	0.091	0.094	0.096
Observations	2,246	2,176	1,608	1,608

Notes: In column (1), children in employment (self-report) = 1 if the child answered affirmatively to the question: "In the last month, have you engaged in one or more of the following activities, for one hour or more? Work in a cocoa plantation"; and 0 otherwise. In columns (2) to (4), children in employment (parent's report) = 1 if the parent answered affirmatively to the question "I will now ask you some questions about activities that your children might have recently performed. In the last month, has any of your children engaged in one or more of the following activities, for one hour or more? Work in a cocoa plantation"; and 0 otherwise. Distance from school (km) is the linear distance between the home and their child's school GPS locations, in kilometers. School-level % children in employment averages self-reports within each school. Baseline controls include child gender, grade indicators, standardized test scores (averaged across numeracy and literacy); and summary measures of parental engagement, student effort, socio-emotional skills, working memory, visual attention, impulsivity, self-esteem, and mindset (see **?**. Summary measures computed following Kling et al. (2007), standardizing each component by normalizing values by the mean and standard deviation of the control group at the baseline survey within each grade, and then averaging over all standardized components. In columns (3) and (4), sample restricted to observations with valid home GPS coordinates. All regressions estimated through Ordinary Least Squares with standard errors clustered at the classroom level. *** p<0.01, ** p<0.05, * p<0.10

In Table 2, column (1) documents that the intervention did not systematically affect children in employment according to children's self-reports. In contrast, based on parents' reports, one would have concluded that the intervention *increased* the prevalence of children in employment (column 2). The effect is sizeable: a 40.3% increase (8.9 p.p., significant at the 10% level) relative to an end-line prevalence of 22.1% in the control group. Column (3) documents that the same holds within the sub-sample with valid home GPS information. Most importantly, column (4) documents that this is likely an artifact of the interplay of the intervention with social desirability biases in communities that have been previously sensitized: very close to the school, the estimated effect sizes is even larger (a 14.5 p.p. increase, significant at the 5% level), but it sharply decreases with distance to school, fading out at an approximately 3km radius, based on our linear estimate.

6 Concluding Remarks

Data on the prevalence of child labor across space and over time is critical for governments and international organizations committed to ensuring children's rights. In the cocoa industry, particularly intensive in child labor given its low rate of mechanization, data issues have posed important challenges to monitoring and enforcement by international organizations and policymakers over the years.

Our finding that child labor statistics following the ILO survey methodology not only misrepresent the prevalence of child labor, but also mischaracterize its trends (especially where interventions have been put in place), raises critical concerns. While the general sentiment of the literature on child labor is that substantial progress has been achieved in recent decades ("[a]n important lesson from all the literature reviewed herein is that child labor can change dramatically and quickly in countries as a result of changes in the economic and policy environment."; Edmonds and Theoharides, 2021, pp. 27-28) –, our results call that sentiment into question.

Based on our results, asking children independently about whether they work (and, if so, how many hours) could yield child labor indicators consistent with costly-verification data. There are, however, technical challenges involved in interviewing children, including whether children are asked inside our outside the household, and variations in the understanding of what exactly characterizes 'work' by children in different countries (see Guarcello et al., 2010 and Dillon, 2010 for a broader discussion of different framing issues). While the experiences of NORC and Tulane University in Côte d'Ivoire and Ghana since 2008-09 could inform the replication of self-reports elicitation of child labor by other agencies moving forward, surveying children directly involves some ethical challenges as well. In particular, children might be exposed to violence if interviews trigger backlash from parents, even when they are conducted in a different setting (e.g., schools). These concerns also limit the potential of technologies worn by children, such as GPS trackers, despite their validated accuracy (Dillon et al., 2017).

For those reasons, finding ways to limit reporting biases in adult surveys seems like a superior (and necessary) alternative. Progress in this space has, however, been slow. Most studies focus on indirect elicitation methods (in particular list experiments, in various forms), but these methods generate estimates that are not only typically imprecise, but also, not necessarily closer to the 'ground truth'.[4] In contrast, leveraging technologies to survey adults in communities less subject to previous sensitization and during periods when it is easier to observe child labor might be a much better way forward.

[4]See, for instance, https://statmodeling.stat.columbia.edu/2014/04/23/thinking-list-experiment-heres-list-reasons-think/).

Naturally, that also involves its own challenges. Although we have documented that remoteness is not systematically associated with the true prevalence of child labor (based on children's self-reports) in the close vicinity of schools, it might be that the communities ultimately surveyed based on satellite imagery are *not* representative of the territories or the populations of interest, leading to biased aggregate estimates of children in employment and child labor. Moreover, if remote sensing through satellite imagery is imperfect (e.g., if the geographical coverage of the available data is selective either when it comes to which regions it covers or which farms it surveys within each region), then the estimates will again be biased. There are also cost considerations. Obtaining access to high-frequency data that can inform prediction models (and hiring staff or external vendors to train and update such models) might be expensive. Such requirements might limit the ability to learn timely about how child labor evolves across space, particularly in response to policy interventions.

While we expect our contributions to generalize beyond the specifics of the Ivorian cocoa industry, a limitation is that our sample consists entirely of school children – all of whom were enrolled at the time of our surveys. This excludes those who had dropped out (or never enrolled in the first place), presumably more likely to work in cocoa fields. While their parents might be less sensitive to social desirability biases (since their children are not in school), there are many other potential sources of social pressure to under-report children in employment – particularly economic pressures linked to restrictions to child labor in global supply chains. Investigating how parents' under-reporting changes with children's school enrollment remains a promising avenue for future work.

References

Christian Science Monitor (2019). Child labor in decline: Life gets better for 94 million children. Last accessed February 3, 2022.
URL: *https://www.csmonitor.com/World/Points-of-Progress/2019/0828/Child-labor-in-decline-Life-gets-better-for-94-million-children*

Dillon, A. (2010). Measuring child labor: Comparisons between hours data and subjective measures, *Research in Labor Economics* (31): 135–169.

Dillon, A., Hanicotte, H. and Valenza, M. (2017). Pilot data analysis & methodological insights on labor measurement, *Technical report*, Innovations for Poverty Action.

Edmonds, E. V. and Theoharides, C. (2021). Child labor and economic development, *Handbook of Labor, Human Resources and Population Economics*, Springer, pp. 1–29.

Enveritas (2020). Supply chain verification: Finding small-holder coffee farms. Last accessed 8 February 2020.
URL: *https://storymaps.arcgis.com/stories/10ef0006e4d94b86b228748c7082cf85/print*

Galdo, J., Dammer, A. and Abebaw, D. (2020). Gender bias in agricultural child labor: Evidence from survey design experiments, (13826).

Guarcello, K., Kovrova, I., Lyon, S., Manacorda, M. and Rosatti, F. (2010). Towards consistency in child labourmeasurement: assessing the comparability of estimates generated by different survey instruments, (54).

ILO and UNICEF (2021). Child labour: Global estimates 2020, trends and the road forward.

International Labour Organization (2017). Methodology of the global estimates of child labour, 2012-2016.

Jouvin, M. (2021). Addressing social desirability bias in child labor measurement : an application to cocoa farms in côte d'ivoire, (2021-08).

Kling, J., Liebman, J. and Katz, L. (2007). Experimental analysis of neighborhood effects, *Econometrica* **75**(1): 83–119.

Lichand, G. and Wolf, S. (2022). Arm-wrestling in the classroom: the non-monotonic effects of monitoring teachers, *University of Zurich, Department of Economics, Working Paper* (357).

Lépine, A., Toh, W. and Treibich, C. (2022). Estimating the prevalence of child labour in the cocoa industry via indirect elicitation methods: A mixed-methods study.

Sadhu, S., Kysia, K., Onyango, L., Zinnes, C., Lord, S., Monnard, A. and Arellano, I. (2020). Norc final report: Assessing progress in reducing child labor in cocoa production in cocoa growing areas of côte d'ivoire and ghana, *Technical report*.

Tran, H., Nathan, S., Ilmma, A., Burkiewicz, M. and Wisana, D. G. K. (2021a). Identifying limiting factors for feasible productivity improvement for smallholder farmers in coffee sector in indonesia, *Asian Journal of Agriculture* **5**(2).

Tran, H., Nathan, S., Ilmma, A., Burkiewicz, M. and Wisana, I. (2021b). Identifying limiting factors for feasible productivity improvement for smallholder farmers in coffee sector in indonesia, *Asian Journal of Agriculture* **5**(2): 53–60.

TRECC (2021). Transforming education in côte d'ivoire – reflections, lessons learned, next steps. Last accessed March 4, 2022.
URL: *https://www.cocoainitiative.org/sites/default/files/resources/TRECC-5-YEAR-REPORT-4.pdf*

Wolf, S. and Lichand, G. (2023). Nudging parents and teachers to improve learning and reduce child labor in côte d'ivoire, *mimeo* .

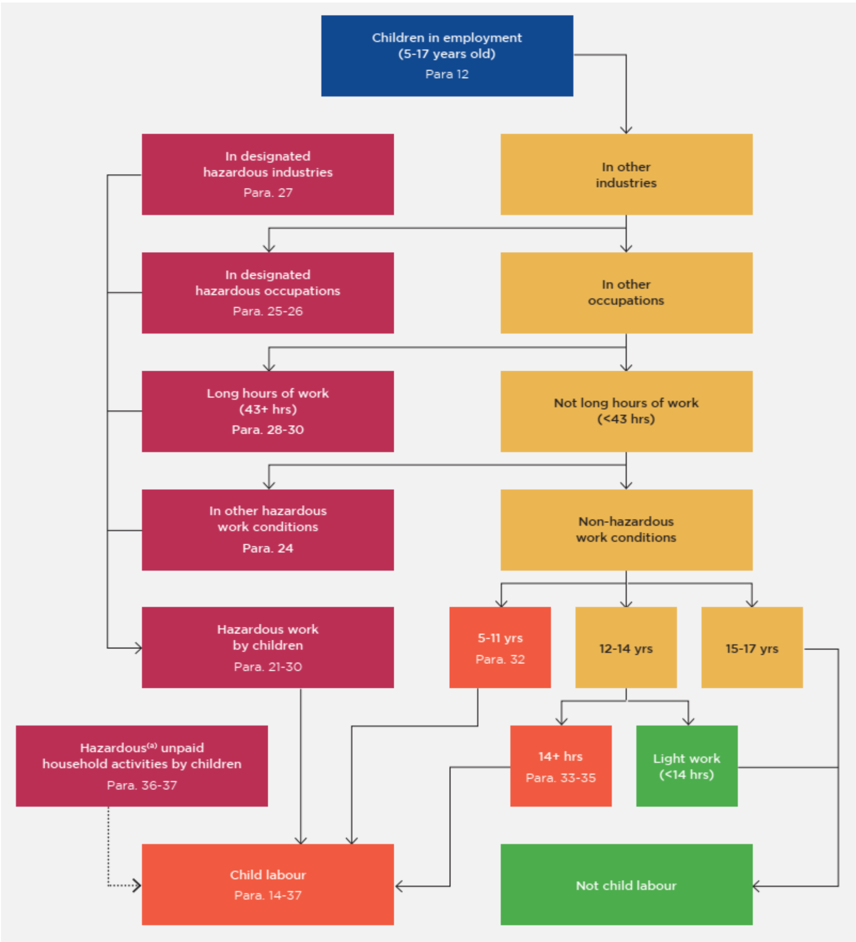
Yoroba, F., Kouassi, B. K., Diawara, A., Yapo, L. A., Kouadio, K., Tiemoko, D. T., Kouadio, Y. K., Kone, I. D. and Assamoi, P. (2019). Evaluation of rainfall and temperature conditions for a perennial crop in tropical wetland: a case study of cocoa in côte d'ivoire.

A Supp. Materials – Additional details on classification methodologies

This Appendix compiles additional details on how the International Labor Organization (ILO), NORC and ENVERITAS compile data on children' work hours and work conditions, and on how these data is used to compute children in employment and child labor statistics in each case.

A.1 ILO

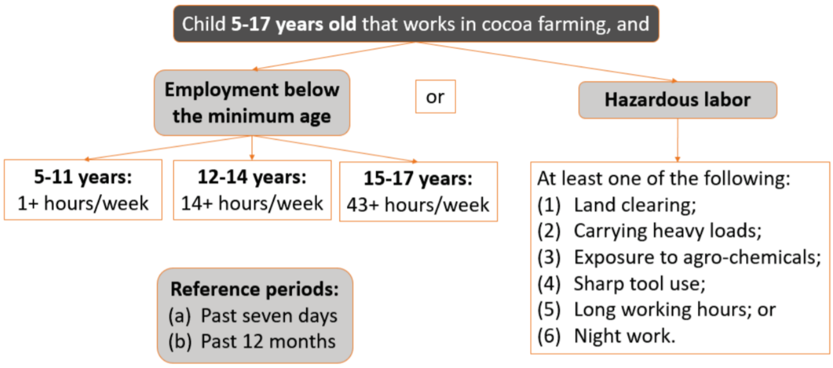
Figure A1: Conceptual framework of the ILO global estimation of child labor



Notes: extracted from ILO and UNICEF (2021).

A.2 NORC

Figure A2: Conceptual framework of the NORC estimation of child labor in cocoa farming



Notes:

A.3 ENVERITAS

Figure A3: Geographical coverage of ENVERITAS data

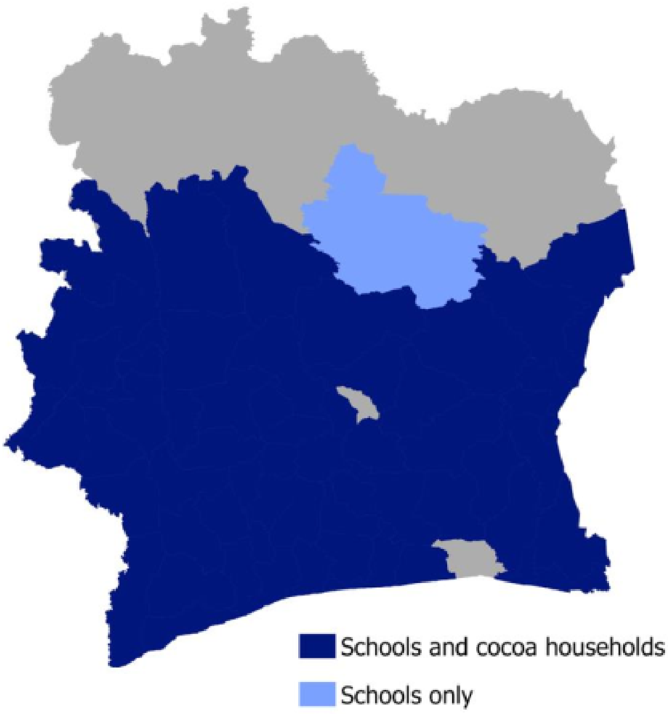


Figure A4: Satellite picture of a community visited by ENVERITAS

Notes: Satellite picture of Baho-Brousse, a community visited by ENVERITAS on Jan/2020. The surrounding cocoa fields are clearly visible in the picture (those not arranged in rows, which, in turn, are rubber plantations).

Figure A5: Enumerators surveying hard-to-reach communities (Jan/2020)

Figure A6: Enumerators surveying hard-to-reach communities (Jan/2020)

B Supp. Materials – Survey instruments

This Appendix compiles the questions used to assess children in employment through our business-as-usual surveys, organized by wave and by whom was asked in each case. For all details on the survey instruments used in the context of the over-arching project, see Lichand and Wolf (2022).

Table B1: Survey questions about children in employment

Questions	Timeline	Respondent
"In the last month, have you engaged in one or more of the following activities, for one hour or more? Work in a cocoa plantation"	Baseline	Children
"I will now ask you some questions about activities that your children might have recently performed. In the last month, has any of your children engaged in one or more of the following activities, for one hour or more? Work in a cocoa plantation"	Baseline	Parents
"In the last month, have you engaged in one or more of the following activities, for one hour or more? Work in a cocoa plantation"	End line	Children
"I will now ask you some questions about activities that your children might have recently performed. In the last month, has any of your children engaged in one or more of the following activities, for one hour or more? Work in a cocoa plantation"	End line	Parents
"Did (child name) work in cocoa fields more than 10 hours a week in a typical week last year?"	Follow-up	Parents
"Did (child name) work in cocoa fields more than 10 hours a week during holidays last year?"	Follow-up	Parents
"To the best of your knowledge in this list of students, can you point out the students who worked in cocoa fields over 10 hours a week in a typical week last year?"	Follow-up	Teachers
"To the best of your knowledge in this list of students, can you point out the students who worked in cocoa fields over 10 hours a week in a typical week during vacation last year?"	Follow-up	Teachers
"To the best of your knowledge in this list of students, can you point out the students who worked at all in cocoa fields during the school year last year?"	Follow-up	Teachers

Notes:

C Supp. Materials – Descriptive statistics

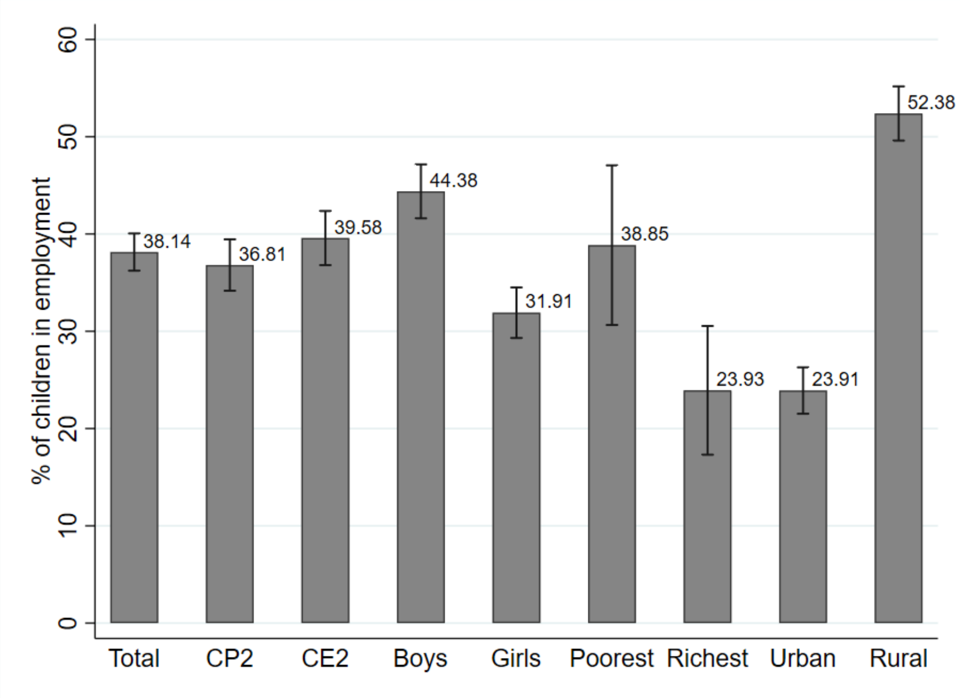
This Appendix compiles descriptive statistics of the study sample, based on our baseline survey. Table C1 provides summary statistics for student and household characteristics. Figure C1 showcases the share of children in employment by student and household characteristics. Next, Figure C2 displays classroom-level correlations between the baseline share of children in employment (based on children's self-reports) and baseline standardized test scores (Panel A) and student dropout rates by the end of the school year (Panel B; restricting attention the control group of the intervention). Last, Figure C3 displays the histogram of distance to the school – the measure of remoteness we explore in the main text.

Table C1: Descriptive statistics (baseline survey)

	Mean	S.D.	Obs
Child is a girl	0.50	0.50	2,475
Child age			2,150
Under 5 years old	0.00	0.04	
5-11 years old	0.92	0.27	
12-14 years old	0.07	0.26	
15 years old and above	0.01	0.08	
Enrolled in 1st primary cycle (CP2)	0.52	0.50	2,475
Rural household	0.50	0.50	2,471
Household monthly income (in 2015 USD)			2,177
Less than USD 19	0.06	0.24	
USD 19-37	0.16	0.37	
USD 37-55	0.16	0.36	
USD 55-92	0.21	0.41	
USD 92-185	0.23	0.42	
USD 185-370	0.11	0.31	
More than USD 370	0.07	0.26	

Notes: CP2 is the second grade for the 1st primary cycle in Côte D'Ivoire education. Rural areas are defined according to parents' main occupation: agricultural or plantation activities are defined as rural. Household monthly income was reported in CFCA and converted to 2015 USD.

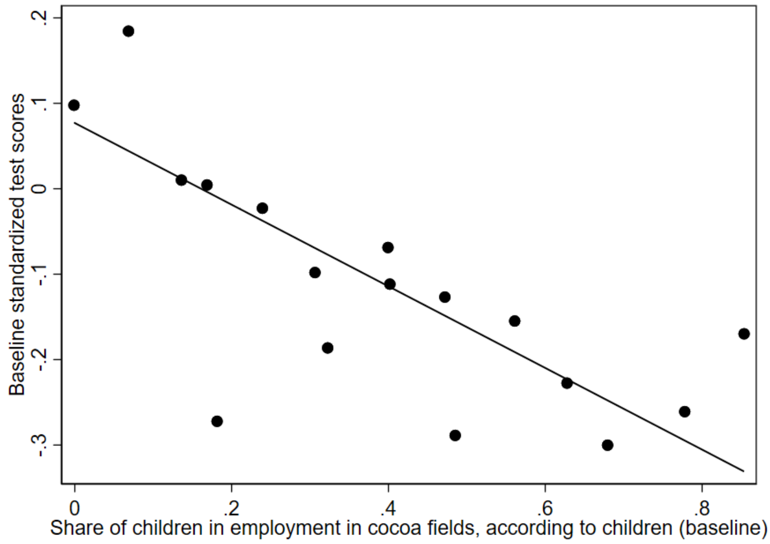
Figure C1: Share of students who worked for at least one hour in cocoa plantations over the last month, according to children



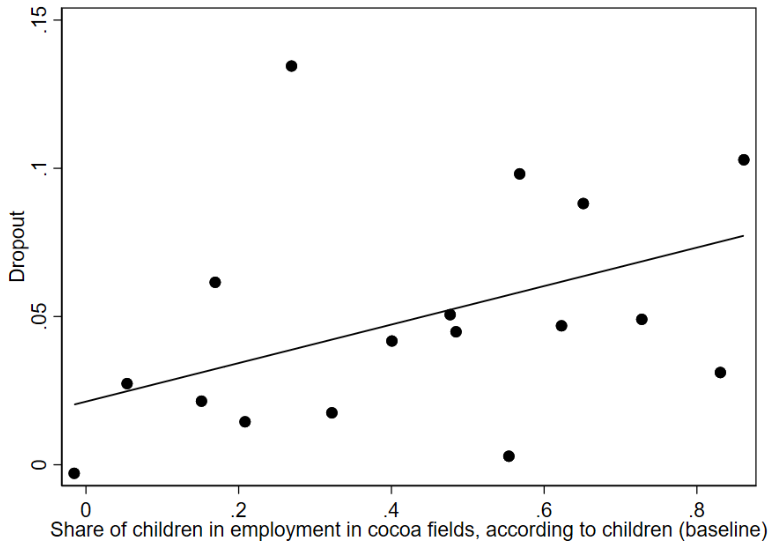
Notes: Figure C1 shows the share of students who report to have worked in cocoa plantations in the last month for one hour or more at baseline, in response to the following question: "In the last month, have you engaged in one or more of the following activities, for one hour or more? Work in a cocoa plantation", as described in Table B1. The first bar comprises the whole sample in Lichand and Wolf (2022); additional bars consider the indicated sub-samples. "Poorest" comprises households with monthly income reported by parents below 10,000 CFA ($\sim$ 19 USD), while "Richest" comprises those with monthly income reported by parents above 200,000 CFA ($\sim$ 372 USD). Rural and urban areas are defined according to parents' main occupation (agricultural or plantation activities are assigned to the former). Samples sizes are the following: (i) Total: 2,475; (ii) CP2: 1,285; (iii) CE2: 1,190; (iv) Boys: 1,237; (v) Girls: 1,238; (vi) Poorest: 139; (vii) Richest: 163; (viii) Urban: 1,234; (ix) Rural: 1,237.

Figure C2: Baseline correlation between child labor and educational outcomes

Panel A: Correlation between test scores and child labor

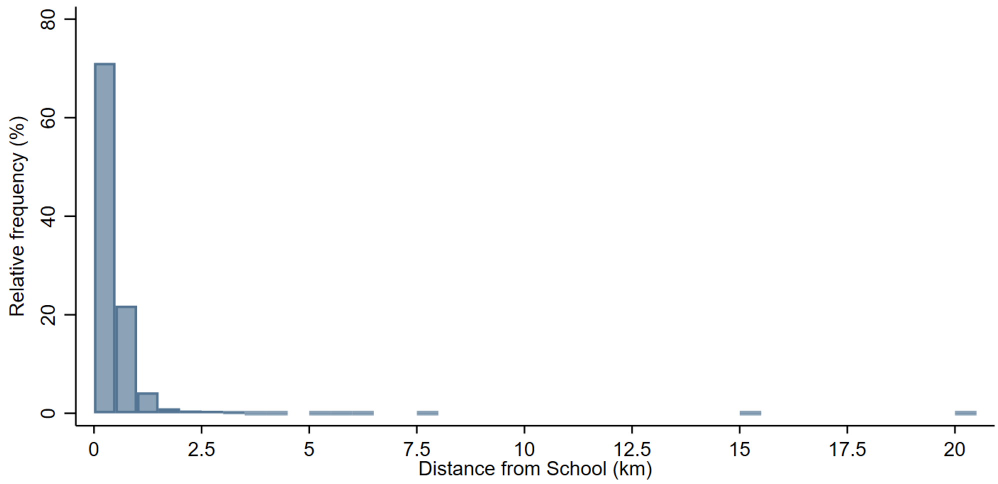


Panel B: Correlation between student dropouts and child labor



Notes: Panel A reports a bin-scatter plot of baseline standardized test scores as a function of baseline children in employment in cocoa fields (self-reported by children). Standardized test scores are a summary measure of numeracy and literacy test scores (averaging across each component, normalized by their mean and standard deviation in the control group), following Kling et al. (2007). Children in employment stands for the baseline share of students who report to have worked in cocoa plantations in the last month for one hour or more, in response to the following question: "In the last month, have you engaged in one or more of the following activities, for one hour or more? Work in a cocoa plantation", as described in Table B1. Test scores and the prevalence of children in employment are averaged at the classroom level. Panel B reports a bin-scatter plot of student dropout rates, based on administrative data (see Lichand and Wolf, 2022), as a function of baseline children in employment in cocoa fields (self-reported by children). Student dropouts and the prevalence of child labor are averaged at classroom level. Because student dropouts are defined at the end line, Panel B restricts observations to the control group of the intervention.

Figure C3: Histogram of the distance from households to the school (in km)

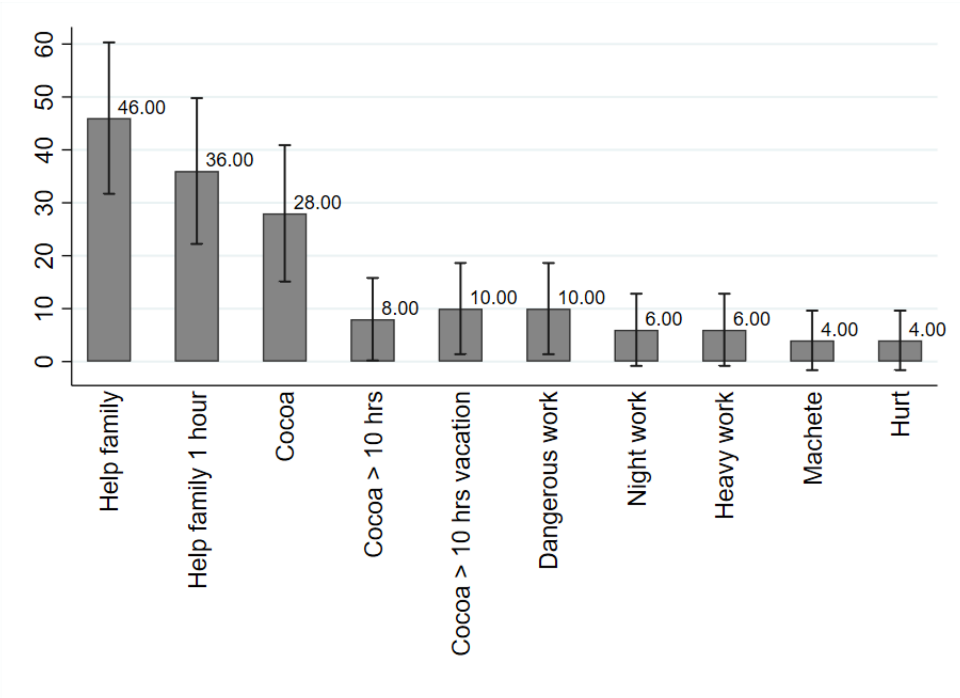


Notes: Distribution for the sub-sample of children with valid GPS coordinates associate with their household location.

D Supp. Materials – Prevalence of children in employment and hazardous work conditions measured in the follow-up survey

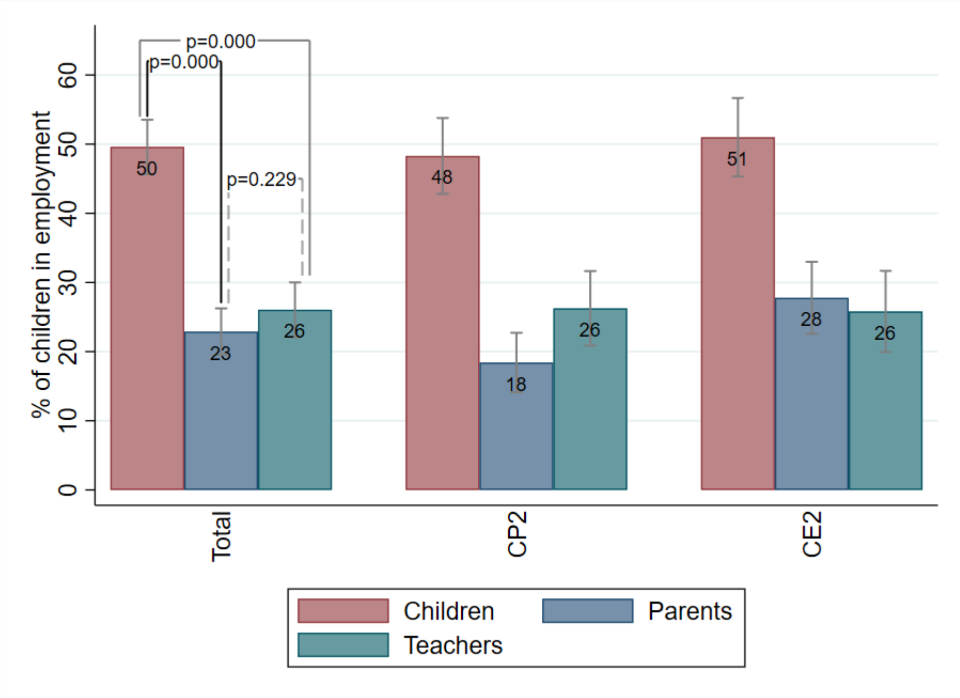
This Appendix compiles results on children's work hours and work conditions based on the follow-up survey (conducted in October 2019). Figure D1 plots the share of affirmative answers to different questions about children in employment. Figure D2 adds to the comparisons in the main text what teachers report about each student in the follow-up survey, when they were asked to report whether each of them had worked in cocoa fields at any point over the course of the previous school year. As the figure shows, teachers' and parents' reports (the latter, collected at the end-line survey) are statistically identical ($p=0.23$), and significantly under-estimate the end-line prevalence of children in employment according to children ($p<0.001$ in each case).

Figure D1: Parents' reports on children in employment and hazardous work conditions at the follow-up wave (October 2019)



Notes: Bars show the share of students who parents have reported to be in the following conditions (yes/no questions): 1) Help family: "Did (child name) work at all to help you around home, assist in a family business or earn pocket money outside school hours under adult supervision during the school year last year?"; 2) Help family 1 hour: "Did (child name) work at all to help you around home, assist in a family business or earn pocket money outside school hours under adult supervision over 1 hour a week during the school year last year?"; 3) Cocoa: "Did (child name) work at all in cocoa fields during the school year last year?"; 4) Cocoa > 10 hrs: "Did (child name) work in cocoa fields more than 10 hours a week in a typical week last year?"; 5) Cocoa > 10 hrs vacation: "Did (child name) work in cocoa fields more than 10 hours a week during vacation last year?"; 6) Dangerous work: "Was (child name) involved in activities in cocoa fields such as clearing of forests and felling of trees, bush burning, manipulating agrochemicals or using sharp tools during the school year last year?"; 7) Night work: "Did (child name) work between 7 p.m. and 7 a.m. during the school year last year?"; 8) Heavy work: "Was (child name) engaged in heavy physical labor in a typical week last year?"; 9) Machete: "Did (child name) used a machete while woking in the fields last year?" 10) Hurt: "Did (child name) get hurt at least once while woking in the fields last year?". All measures were collected at the follow-up surveys (Lichand and Wolf, 2022). Across all bars, the sample is restricted to the control group.

Figure D2: Share of students who worked in cocoa plantations during the school year, according to different sources

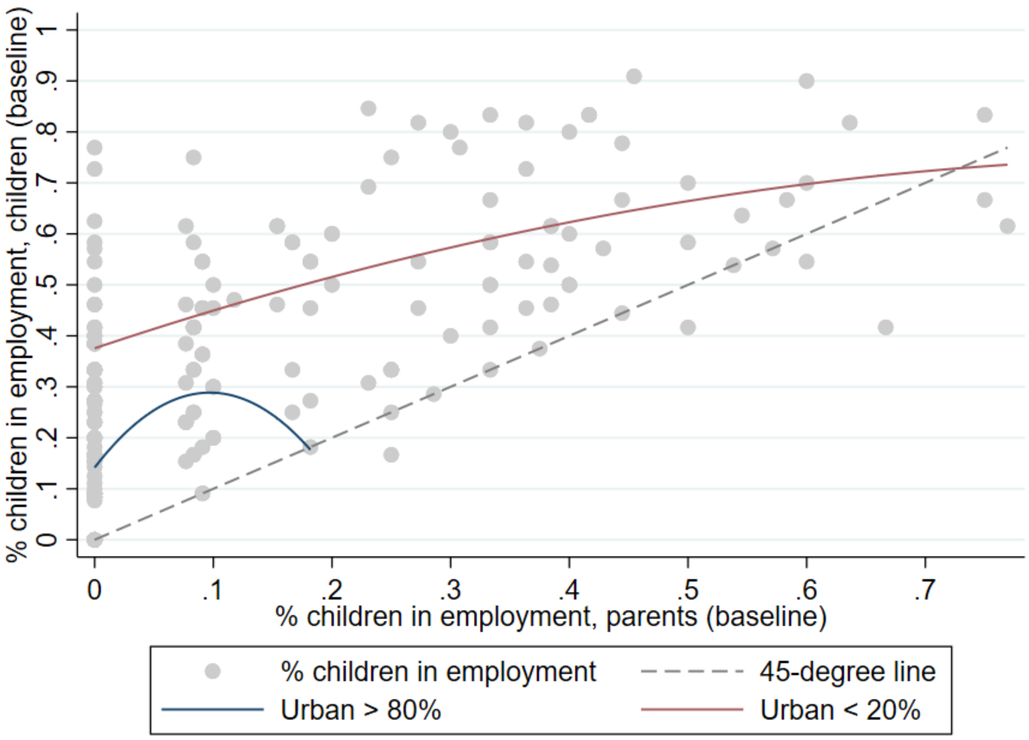


Notes: Bars show the share of students who have worked in cocoa plantations according to children (in red), parents (in blue), or teachers (in green). Children and parents answer at end line if the child engaged in one hour or more in cocoa plantation activities over the last month, as described in Table B1. Teachers answer at follow-up if each of their students worked in cocoa plantation at all over the last school year, as described in Table B1. Across all bars, the sample is restricted to the control group. The first set of bars comprises the whole sample in Lichand and Wolf (2022), while the additional ones split the sample by primary grades (CP2 and CE2). Sample sizes are the following: (i) Total: 625 ; (ii) CP2: 323; (iii) CE2: 302.

E Supp. Materials – Additional results

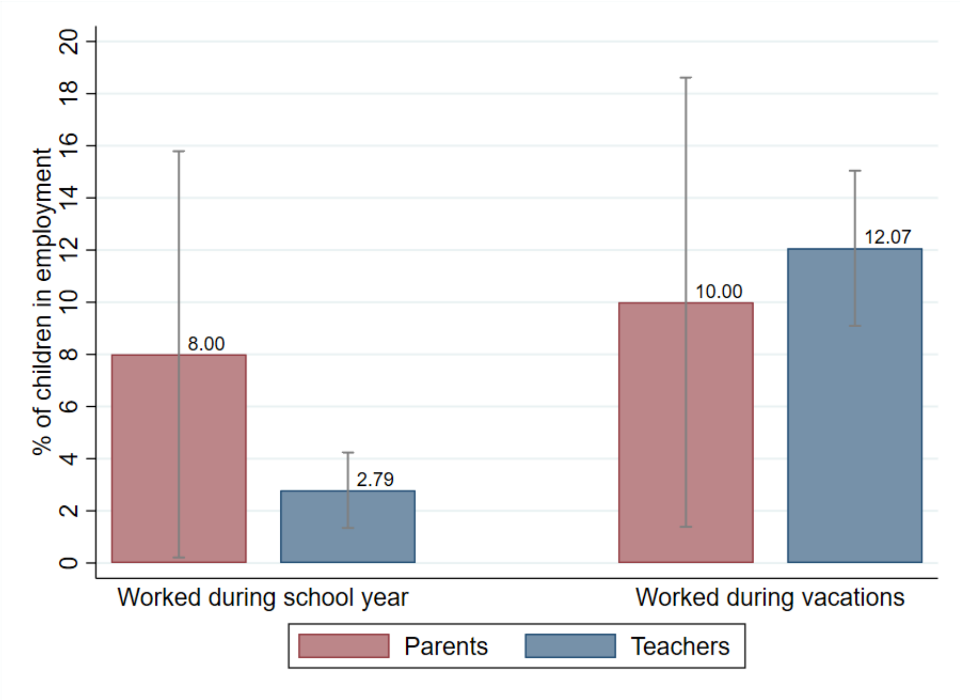
This Appendix compiles additional results. Figure E1 illustrates how under-reporting patterns vary by urban status, at the classroom level. For almost all classrooms, adult reports are higher than children's self-reports (only 6 out of 198 observations are below the 45-degree line). The red line estimates a quadratic relationship for mostly-rural classrooms (those with less than 20% of children living in urban areas), and the blue line, for mostly-urban classrooms (with over 80% of children living win urban areas). No mostly-urban classroom has a prevalence of children in employment greater than 20%. Within that lower range, however, under-reporting can be very high; for instance, the estimated relationship predicts that if parents in these classrooms report a 10% prevalence of children in employment, the bias-adjusted prevalence is actually closer to 30%. For mostly-rural classrooms, under-reporting is predicted to be even the more striking the lower the prevalence in adult reports is. In these classrooms, a 10% reported prevalence would correspond to a nearly 45% bias-adjusted prevalence of children in employment. Naturally, in both cases, there is less room for under-reporting as prevalence according to adult reports increases. Next, Figure E2 documents the share of students reported to work over 10 hours a week at any point during the previous school year, according to both parents and teachers. Teachers were more conservative than parents when the question focused on employment during the school year, but less so if it focused on school holidays. Concretely, teachers identified child labor for only 2.8% of students during the school year (8%, according to parents) but for 12% of students during school holidays (10%, according to parents). Last, Figure E3 documents that, depending on reporting sources, the correlation between children in employment and educational outcomes at end line also systematically differs across treatment cells.

Figure E1: Correlation between parents' and children's answer according to urbanization level



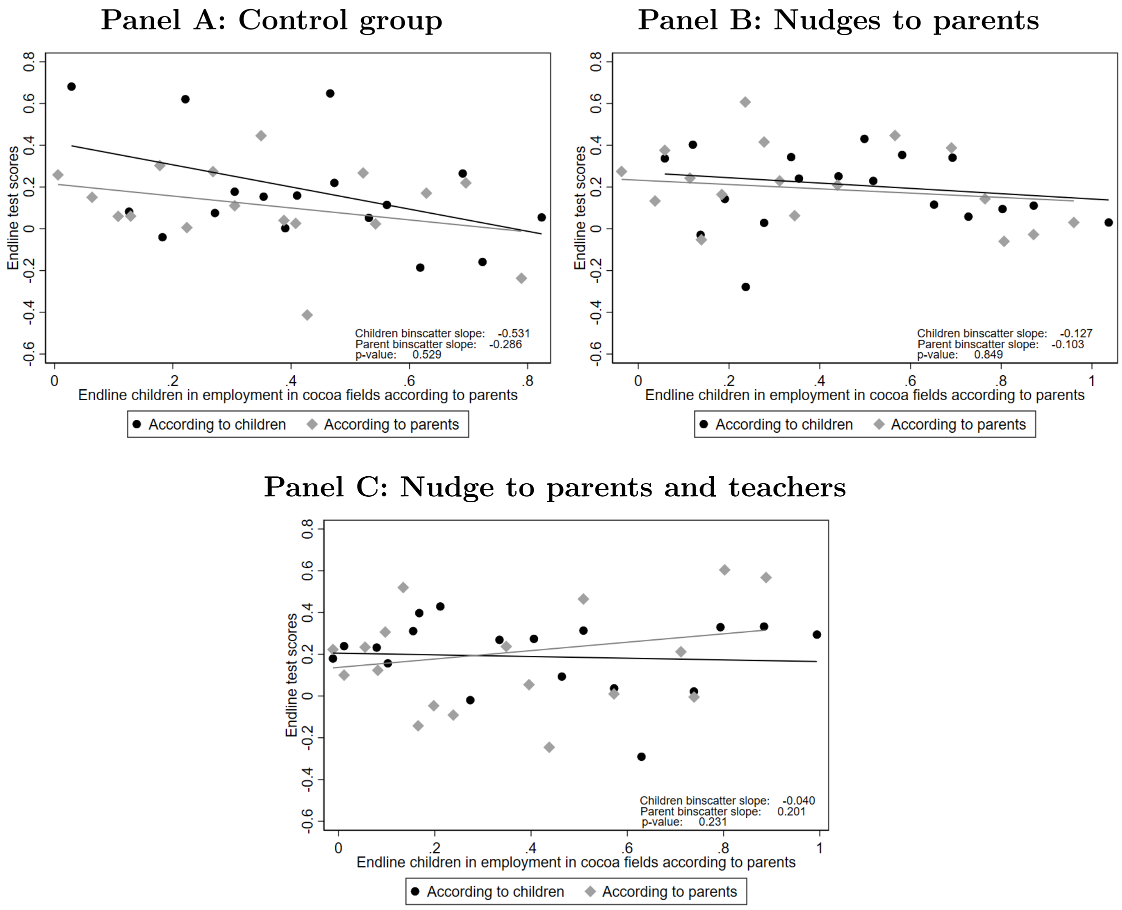
Notes: Data is aggregated at classroom level: average child labor according to children and parents were correlated according to composition on the classroom urbanization. "Urban > 80%" represents the quadratic fit between children's and parent's answer on child labor to classrooms that have more than 80% of their students in the urban category, while "Urban < 20%" represents the quadratic fit between children's and parent's answer on child labor to classrooms that have less than 20% of their students in the urban category, i.e., majority of students are from rural areas. Children answered the following question at baseline: "In the last month, have you engaged in one or more of the following activities, for one hour or more? Work in a cocoa plantation", as described in Table B1. Parents answered the following question at baseline: "I will now ask you some questions about activities that your children might have recently performed. In the last month, has any of your children engaged in one or more of the following activities, for one hour or more? Work in a cocoa plantation", as described in Table B1. Rural and urban areas are defined according to parents' main occupation (agricultural or plantation activities are assigned to the former).

Figure E2: Share of students who worked in cocoa plantations at least 10 hours/week, during the school year and during the school holidays



Notes: Bars show the share of students who worked in cocoa plantations during different periods and according to two different sources. Parents (in red) and teachers (in blue) answered whether the child worked 10 hours/week or more in cocoa fields during a typical week (LHS bars) and whether the child worked 10 hours/week or more in cocoa fields during school holidays in the previous school year, as described in Table B1. All measures were collected at the follow-up surveys (Lichand and Wolf, 2022). Across all bars, the sample is restricted to the control group. Sample sizes are the following: (i) Parents: 200; and (ii) Teachers: 2,500.

Figure E3: Correlation between end-line test scores and child labor, by source and treatment status



Notes: All panels show bin-scatter plots of end-line standardized test scores as a function of children in employment in cocoa fields at end line. Each panel estimates linear relationship between the variables according to different measures of child labor; that based on children's self-reports is shown in black, and that based on parents' reports is shown in gray. Children answered the following question: "In the last month, have you engaged in one or more of the following activities, for one hour or more? Work in a cocoa plantation", as described in Table B1. Parents answered the following question: "I will now ask you some questions about activities that your children might have recently performed. In the last month, has any of your children engaged in one or more of the following activities, for one hour or more? Work in a cocoa plantation", as described in Table B1. Standardized test scores are a summary measure of numeracy and literacy test scores (averaging across each component, normalized by their mean and standard deviation in the control group), following Kling et al. (2007). Panel A restricts the sample to the control group of the intervention; Panel B, to parents who randomly assigned to nudges (via text or or audio messages) in schools where teachers were not assigned to the intervention; and Panel C, to parents assigned to nudges (via text or audio messages) in schools where teachers were also assigned to nudges (via text messages); see Lichand and Wolf (2022).

F Supp. Materials – Predicting the bias-adjusted prevalence of child labor

Informed by the statistical equivalence between ENVERITAS data (treated as the ground truth) and children's reports of the prevalence of children in employment, we can estimate the relationship between adult reports (the ILO standard measure) and children's self-reports not only to compute the extent of under-reporting in our data, but also to predict what accurate reports would have been in geographical units without ENVERITAS data. In this Appendix, we show how estimating this *bias-adjustment factor* for children in employment allows us to compute the bias-adjusted prevalence of child labor for Côte d'Ivoire and Ghana, under the assumption that this relationship is the same conditional on the local characteristics we can measure. We focus on these two countries because 2/3 of the global cocoa production originates within their borders and because our surveys asked about children in employment specifically in cocoa farms.

According to the ILO definition:

$$CL_i = CE_i \text{ x } HC_i, \tag{1}$$

where CL_i is the share of child labor in country i, CE_i is the share of children in employment in country i, and HC_i is the share of children in employment in country i who work in hazardous conditions (= 1 for children under 12, and = % of those working long hours and/or under heavy or dangerous work conditions, otherwise).

Once we determine that children's self-reports are accurate and, as such, can be used as the ground truth, in our data we observe both CE_j, the share of children in employment reported by parents in classroom j, and CE_j^*, the ground truth for children in employment in classroom j. We are interested in predicting CE_i^* for all countries $i \in I$. To do that, we first estimate:

$$CE_j^* = f(CE_j, X_j) + \varepsilon_j, \tag{2}$$

where $f(CE_j, X_j)$ is a function of children in employment reported by parents in classroom j and other classroom characteristics X_j, and ε_j is an error term.

We then use our estimates to compute:

$$\hat{CE}_i^* = \hat{\alpha}_i \text{ x } CE_i, \tag{3}$$

where $\hat{\alpha}_i = \frac{\hat{f}(CE_i, X_i)}{CE_i} \geq 1$.

Those same estimates can be used to recover CL_i^*, the ground truth for child labor in country i. This is thanks to the fact that HC_i is free of bias (since = 1 for 5-11 year olds, and elicited directly from 12-14 year-old children), and to the multiplicative nature of its definition:

$$CL_i^* = CE_i^* \text{ x } HC_i = \hat{\alpha}_i \text{ x } CE_i \text{ x } HC_i = \hat{\alpha}_i \text{ x } CL_i \tag{4}$$

Concretely, we estimate a quadratic polynomial, allowing the extent of under-reporting to vary with both the local level of children in employment reported by parents and with the local urbanization rate:

$$CE_j^c = \sum_{k=0}^{2} \beta_k \left(CE_j^p\right)^k + \sum_{k=0}^{2} \gamma_k \left[\left(CE_j^p\right)^k \text{ x } urban_j\right] + \varepsilon_j, \tag{5}$$

where CE_j^c is % child labor according to children in classroom j, CE_j^p, that according to parents in classroom j, and $urban_j$ is the % of students from urban areas in classroom j.

We only estimate heterogeneity with respect to children in employment reported by adults and urbanization rates because these are available both in our data and in the World Development Indicators. Other variables, like annual per capita income, are not available in our data.

Last, we compute $\hat{\alpha}_i = \max\left\{1, \frac{\hat{CE}_i^c}{CE_i}\right\}$, where $\hat{CE^c}_i$ is the predicted value of CE^c for country i using equation (5). As the formula indicates, we constrain the estimated bias-adjustment factor (as well as its confidence interval) to be greater or equal to 1. In practice, this constraint does not affect Côte d'Ivoire or Ghana.

World Development Indicators track child labor for 97 countries (those where the issue is considered to be relevant by the data collection organizations, such as UNICEF country offices), focusing on 7-14 year-old children. With the bias-adjustment factor that we predict for Côte d'Ivoire and Ghana using the procedure described above, we compute country-level bias-adjusted prevalence of child labor and number of child workers (and accompanying 95% CIs) using data from the most recent year available for each country in the World Development Indicators. To arrive at the bias-adjusted prevalence figures, we have to rely, in addition, on World Development Indicators data on the number of 0-14 year old children by country, and subtract these from the UNICEF figures for under-5 children by country. For both population counts, we use 2020 data – assuming that child labor indicators remained constant since their most recent measurement.

Table E1: Relationship between children's and parents' reports at the classroom-level

	(1)
Child labor according to parents	0.463
	(0.294)
Urban	-0.361***
	(0.056)
(Child labor according to parents)^2	-0.069
	(0.439)
(Child labor according to parents)*Urban	0.850
	(0.633)
(Child labor according to parents)^2*Urban	-1.491
	(1.270)
Constant	0.477***
	(0.041)
Observations	198
Adjusted R^2	0.570

*Notes: * p<0.10, **p<0.05, ***p<0.01.*
Estimation based on the following equation: $CL_c = \alpha + \beta_1 CL_p + \beta_2 Urban + \beta_3 {CL_p}^2 + \beta_4 CL_p * Urban + \beta_5 {CL_p}^2 * Urban + \varepsilon$, in which: CL_c is child labor according to children, CL_p is child labor according to parents, and $Urban$ the percentage of students living in urban areas. Results are for classroom-level averages. Rural and urban areas are defined according to parents' main occupation (agricultural or plantation activities are assigned to the former). Child labor according to parents corresponds to the answer on the following question "I will now ask you some questions about activities that your children might have recently performed. In the last month, has any of your children engaged in one or more of the following activities, for one hour or more? Work in a cocoa plantation", as described in Table B1. Child labor according to children stands for the baseline share of students who report to have worked in cocoa plantations in the last month for one hour or more, in response to the following question: "In the last month, have you engaged in one or more of the following activities, for one hour or more? Work in a cocoa plantation", as described in Table B1. All variables correspond to baseline answers.

Table E2: Bias-adjusted prevalence of child labor by country

Country	**Prevalence**			**Children in employment**		
	WDI*	**Bias-adjusted**	**95% CI**	**WDI****	**Bias-adjusted**	**95% CI**
Côte D'Ivoire	36.50%	50.75%	[44.80%; 56.71%]	2,488,650	3,460,540	[3,054,542; 3,866,538]
Ghana	28.80%	46.67%	[40.74%; 52.61%]	2,122,198	3,439,275	[3,001,906; 3,876,644]

Notes: * using the latest available WDI figures, following Edmonds and Theoharides (2021); ** using WDI and UNICEF population figures for 0-14 and under-5 children for 2020. Bias-adjusted prevalence of child labor based on Table E1's estimates for country-specific adjustment factors.