Journal of Development Economics

Escrow Mechanisms for Group-based Reporting: Evidence from Bangladesh's Garments Sector

--Manuscript Draft--

Manuscript Number:	DEVEC-D-22-01717R1
Article Type:	Registered Report Stage 1: Proposal
Section/Category:	Experimental Papers, credit, insurance
Keywords:	Reporting escrow; retaliation; Coordination; Gender; garments; Bangladesh
Corresponding Author:	Laura Boudreau UNITED STATES
First Author:	Laura Boudreau
Order of Authors:	Laura Boudreau
	Sylvain Chassang
	Ada Gonzalez-Torres
Abstract:	In developing countries, misbehavior within organizations often goes unpunished due to weak governance. Employees whose livelihoods are precarious are especially vulnerable. Governance tools that safely provide voice and remedy may dramatically improve workers' welfare. Legal scholars have proposed reporting escrows to facilitate coordination among multiple victims of harassment (Ayres and Unkovic, 2012), but little is known about how they perform in practice. We collaborate with a worker helpline in Bangladesh's apparel sector to experimentally test how the availability and design of a reporting escrow affects reporting of harassment and other workplace misconduct. The project's first phase, in which callers from factories randomly assigned to have access to the escrow were informed about it, provides suggestive evidence that this option increases workers' willingness to report. This pre-registration focuses on the second phase, in which training will be provided in 89 factories to build common knowledge about the escrow among treatment workers.
Response to Reviewers:	

Laura Boudreau
Columbia Business School, Columbia University
New York, NY 10027

Re: Submission of MS DEVEC-D-22-01717

Dear Andrew,

Thank you for giving us the opportunity to revise our Phase 1 registered report for the *Journal of Development Economics (JDE)*. We hereby resubmit the attached Phase 1 registered report, "Escrow Mechanisms for Group-based Reporting: Evidence from Bangladesh's Garments Sector."

We thoroughly appreciate your comments and those from the referees. Your decision letter indicated that you have one key question that is necessary for us to answer for the JDE to make a final determination on Phase 1 acceptance; we respond to this question below and have addressed it in the revised report. The referees raised other helpful suggestions that, in the interest of resubmitting as expediently as possible, we have not incorporated in the revised Phase 1 report. We will consider these comments, though, when preparing the manuscript reporting our study's findings.

Your letter also indicated that you requested a revision of our Phase 1 report on the understanding that the endline data have not been collected. We confirm that at the time of resubmission, the endline data have not been collected. The project's implementation status is that the second Phase, which is the onsite training phase for factories that supply to cooperating buyers, is underway. The onsite training launched during the second half of November 2022 and gradually ramped up.[1] The onsite training is underway with 28 out of 89 factories that will receive the training. At the time of resubmission, because factories are large and face production pressure, the training is gradually being rolled out to factories' workers over several months, and the training remains underway at all 28 factories.[2] We are monitoring calls to the Helpline from these factories solely for the purpose of the escrow's implementation.

We next turn to the key question that you raise in your letter. Your and Referee 1's comments are reported in italics, and our response immediately follows.

> *I think the one key question that goes to the willingness of the JDE to commit to publishing a Phase 2 report is the issue raised by the first referee–the imbalance between*

[1] 3 factories started the training in the second half of November, 8 in December, 13 in January, and 3 so far in February.

[2] Employment information is currently available for 81% of factories participating in the onsite training; among these, the median factory in the onsite training group employs 2,185 workers. As the Helpline rolls out the onsite training, it updates its factory size information for this group.

training and treatment status. The obvious concern would be that the training is selective with respect to treatment. Of course, it is not an issue per se that different fractions of treatment and control receive the training...but it is an issue if the experience of the Phase 1 treatment is causing some firms to back out of the project. As the referee notes even if the training is selective it may be possible to show parallel trends or other evidence that the allocation of treatment and control given training is almost as good as random.

In any case, my primary request is you clear this issue up as much as possible so that we can make a final determination on Phase 1 (our Phase 1) acceptance.

The comment raised by Referee 1 is:

3a. My main concern is that most of the analysis proposed is based on non-experimental comparisons involving selection into trainings. This is particularly a concern because the selection into trainings appears imbalanced across treatment and control – 37/226 (16.4%) of control factories are participating in trainings, while just 52/448 (11.6%) of treatment factories are participating in trainings. Some additional discussion of this selection is therefore important to justify subsequent analysis. A balance table comparing characteristics of control and treatment factories that selected into trainings, for example, would be valuable (see below).

3b. Building on (3a), I would expect to see any registered analysis based on this selection to have appropriate balance and parallel trends tests already implemented, as the authors have the data to implement these tests...

First, we would like to clarify that most of our proposed analysis is based on experimental comparisons between treatment and control factories that *either* both receive the onsite training *or* both do not receive the onsite training. Within these groups, access to the reporting escrow is randomly assigned, so we should expect the experimental comparisons to be internally valid, i.e., they will provide an unbiased estimate of the average treatment effect (ATE) for the relevant group. As we discuss in the Phase 1 report (pg. 3, footnote 1 and pg. 19), the onsite training itself is not randomly assigned. Assignment to the onsite training group is conditional on supplying to at least 1 of 2 buyers that are cooperating with the onsite training program. This means that we should not necessarily expect the ATE of the escrow to be the same for the group of factories that receives onsite training and the group that does not.

We have no reason to suspect that there is differential selection of treatment and control factories into the onsite training group. Referee 1 raised the concern that the fraction of control and treatment factories being selected into the onsite training group is different, with a larger share of control factories being selected. First, we note that there are a small number of factories that have entered the experiment since the launch of Phase 1 due to new entrants into buyers' supplier bases (n=10).[3] Of these, 6 supply to the buyers that are cooperating with the onsite training. As a result, the true proportions of original treatment and control factories participating in the training are slightly different. Among factories that were originally randomized to treatment or to control prior to the launch of Phase 1, 49 of 442 (11.1%) of treatment factories and 34 of 226 (15.0%) of control factories are participating in the onsite training. A simple t-test of whether these two proportions are significantly different fails to reject ($p = 0.143$). While we acknowledge that the proportions are slightly different, they are no more different than we would expect by chance. Further, based on our knowledge of the Phase 1 roll-out, we think that it is implausible that treatment factories are "opting out." First, the buyers that participate in the Helpline have not been informed of their suppliers' treatment status. Second, the availability of the escrow is not salient to factories' management teams. Finally, the escrow has not triggered during Phase 1, so no issues have been raised to management through this mechanism.

Consistent with the claim that treatment and control status is randomly assigned within the onsite training group, we find that these groups are balanced and that they are on similar trends prior to the launch of Phase 1. We now include both a balance table (Table B.3) and two figures (Figures B.2 and B.3) showing the pre-trends for the onsite training group in the revised Phase 1 report. For ease of reference, we also included them in the Appendix of this letter. Table A.I shows that across 28 balance tests, one is statistically significant at the 10% level, which is no more than we would expect by chance. Further, the statistically significant test is actually for the share of onsite training factories for which data on factory size is available. Because the Helpline is collecting this information as it rolls out the onsite training, this variable will no longer be relevant once factory size has been collected for the full onsite training group.[4] Importantly, there are no statistically significant differences in calls to the Helpline for severe labor issues or for less severe labor issues, which are two of the study's three primary outcomes. The final primary outcome, the number of labor issues escalated for investigation by the Helpline, is not available yet.

Turning to pre-trends, Figures A.I and A.II show that factories in the three treatment arms were on similar trends prior to the launch of Phase 1. Factories' call volumes are very variable, which is why one group sometimes has more extreme peaks or troughs in the time series, but the groups' slopes and overall trends are very similar.

We would like to thank you and the referees again for providing us with your suggestions for improving the Phase 1 report and the project. We hope that you will find this revision

[3] We discuss how we accommodate churn in buyers' supplier bases in sections 3.2.1 and 4.2 in the Phase 1 pre-results report.

[4] We acknowledge that factory size is being collected after the launch of Phase 1, when factories assigned to the escrow are being treated. Given that the number of monthly callers to the Helpline is very small relative to the size of the factories, though, workers at treatment factories are only very slowly learning about the escrow's availability during Phase 1. We estimate the upper bound for the share of workers who have been informed about the escrow by the Helpline using the total number of calls from each escrow group between the launch of Phase 1 and January 25, 2023 divided by the total number of employees among factories with employment information available. This share is 0.62-0.82% of workers. Given this, we argue that it is implausible that factories' employment has been affected by the intervention.

to be responsive to your comments and suitable for Phase 1 acceptance at the JDE.

Best regards,

Laura Boudreau, Sylvain Chassang, and Ada González-Torres

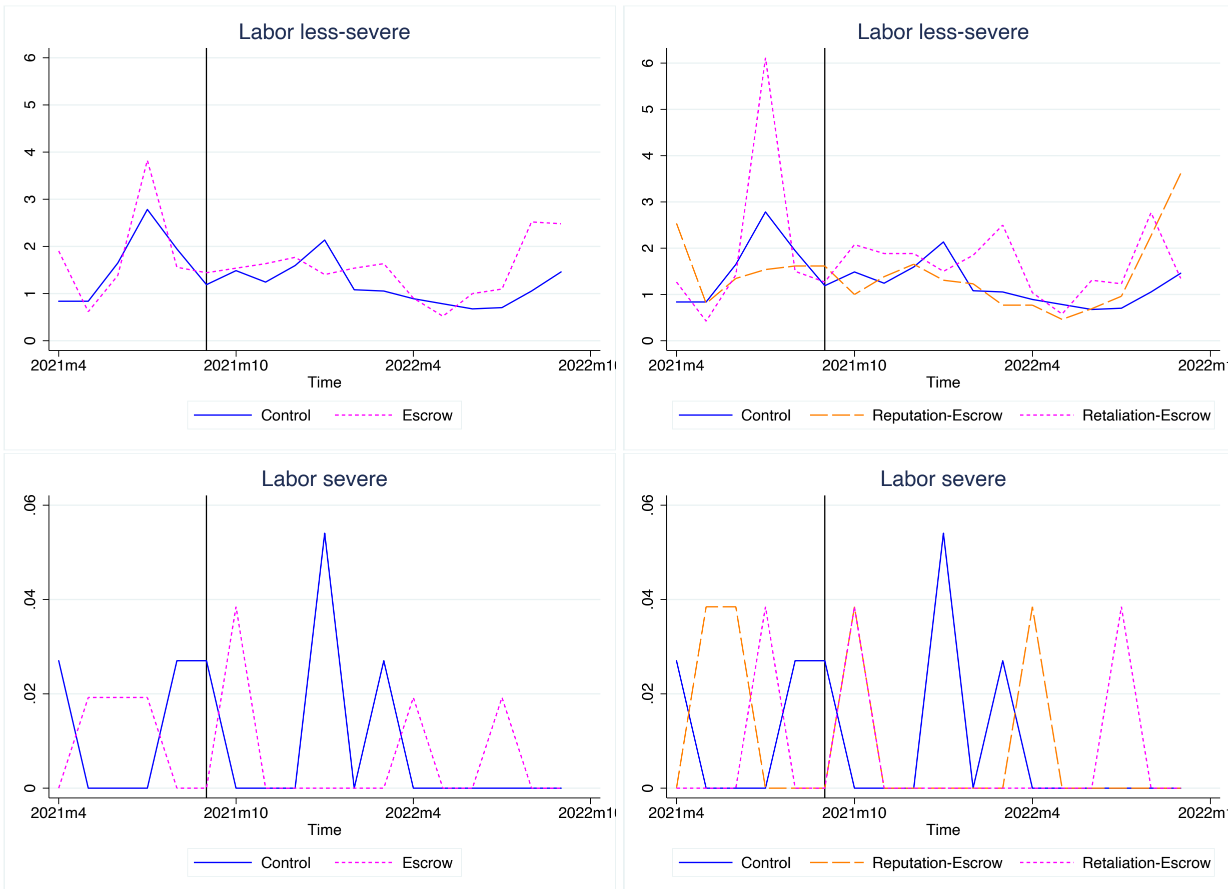
References

Freeman, Richard, and Edward Lazear. 1995. "An Economic Analysis of Works Councils." In *Works Councils: Consultation, Representation, and Cooperation in Industrial Relations*. , ed. Joel Rogers and Wolfgang Streeck, 27–52. Chicago:University of Chicago Press.

Appendix

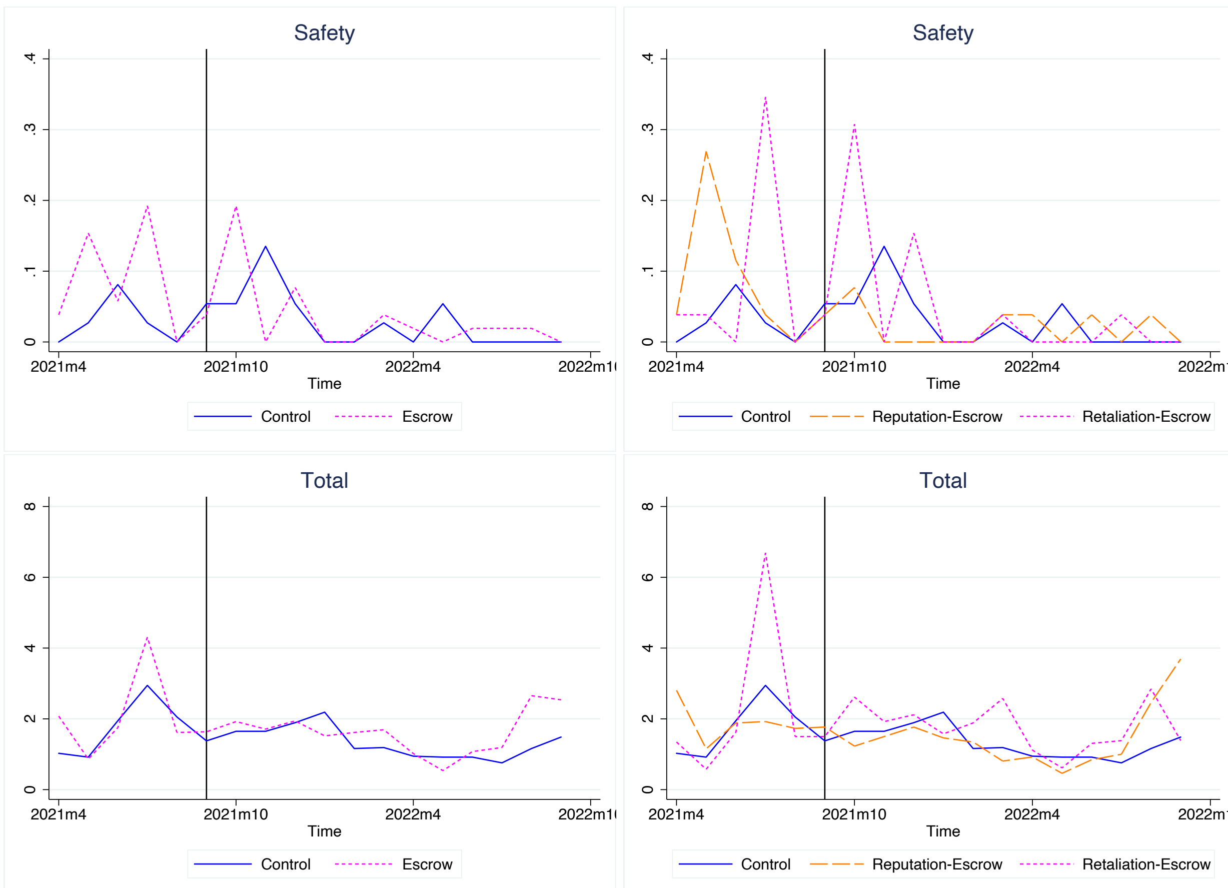
A: Figures and Tables

Figure A.I: Average monthly labor calls to the Helpline per factory by treatment arm for **training factories**



Notes: Average number of substantive calls to the Helpline, per factory, per month, by type of issue, for factories with onsite training. The black vertical lines reflect the month before the launch of Phase I of the experiment (October 19, 2021). Data is prior to the launch of Phase II of the experiment.

Figure A.II: Average monthly calls to the Helpline per factory by treatment arm for **training factories**



Notes: Average number of substantive calls to the Helpline, per factory, per month, by type of issue, for factories with onsite training. The black vertical lines reflect the month before the launch of Phase I of the experiment (October 19, 2021). Data is prior to the launch of Phase II of the experiment.

Table A.I: Balancing Table for **training factories**

	Mean / (SE)				Difference in means / [p-value]			
Variable	No Escrow	Escrow	Reputation	Retaliation	Escrow − No	Rep. − No	Ret. − No	Ret. − Rep.
Labor less severe	1.54	1.79	1.58	2.00	0.17	-0.06	0.54	0.54
	(3.80)	(7.39)	(3.40)	(9.90)	[0.76]	[0.91]	[0.53]	[0.56]
Labor severe	0.01	0.01	0.01	0.01	-0.00	0.00	-0.01	-0.01
	(0.12)	(0.10)	(0.11)	(0.08)	[0.80]	[0.84]	[0.49]	[0.55]
Safety	0.03	0.08	0.08	0.08	0.06	0.05	0.05	0.01
	(0.20)	(0.53)	(0.59)	(0.48)	[0.12]	[0.29]	[0.19]	[0.92]
All calls	1.71	2.04	1.88	2.21	0.26	0.09	0.57	0.44
	(3.94)	(7.50)	(3.70)	(9.95)	[0.65]	[0.88]	[0.51]	[0.65]
Observations	222	312	156	156	534	378	378	312
Factories	37	52	26	26	89	63	63	52
Strata FE					Yes	Yes	Yes	Yes
Factory Size[1]	2652.93	2291.24	2503.68	2115.74	-339.81	-13.09	-532.79	-564.19
	(1897.12)	(1501.23)	(1905.19)	(1076.60)	[0.46]	[0.99]	[0.18]	[0.35]
Data on Factory Size	0.81	0.81	0.73	0.88	0.04	-0.03	0.11	0.16*
	(0.40)	(0.40)	(0.45)	(0.33)	[0.57]	[0.64]	[0.20]	[0.06]
Strata FE					Yes	Yes	Yes	Yes
Fraction of new factories	0.08	0.06	0.08	0.04	-0.02	-0.00	-0.04	-0.04
	(0.28)	(0.24)	(0.27)	(0.20)	[0.68]	[0.95]	[0.48]	[0.56]
Observations	37	52	26	26	89	63	63	52
Strata FE					No	No	No	No

Notes: The table reports the mean values of each variable for each treatment arm and the differences in means between treatment arms for factories with onsite training. The top panel reports calls to the Helpline by issue type for April - September 2021, which are the 6 months preceeding the launch of Phase 1 of the experiment. The bottom panel reports factory size and the fraction of new factories entering the experiment, for which there is 1 observation per factory. In the top panel, columns (1)-(4) report group means and standard errors clustered by factory in parantheses, and columns (5)-(8) report the difference in means estimated using a regression with strata fixed effects and clustering standard errors by factory. In the bottom panel, columns (1)-(4) report group means and robust standard errors in parantheses, and columns (5)-(8) report the difference in means estimated using a regression with strata fixed effects and using robust standard errors. In both panels, [p-values] are reported in brackets. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.
Footnotes: [1] As the Helpline is preparing for the onsite training, it has been collecting updated factory size information for the sample of factories with onsite training. In this analysis, we combine our previously-available data on factory size with the updated data to improve coverage. We will continue to collect factory size as the training rolls out to the onsite training factories.

Journal of Development Economics
Registered Report Stage 1: Proposal

Escrow Mechanisms for Group-based Reporting: Evidence from Bangladesh's Garments Sector

Laura Boudreau
Columbia Business School
l.boudreau@columbia.edu

Sylvain Chassang
Princeton University
chassang@princeton.edu

Ada González-Torres
Ben Gurion University of the Negev
adagt@bgu.ac.il

February 6, 2023

Abstract

In developing countries, misbehavior within organizations often goes unpunished due to weak governance. Employees whose livelihoods are precarious are especially vulnerable. Governance tools that safely provide voice and remedy may dramatically improve workers' welfare. Legal scholars have proposed reporting escrows to facilitate coordination among multiple victims of harassment (**?**), but little is known about how they perform in practice. We collaborate with a worker helpline in Bangladesh's apparel sector to experimentally test how the availability and design of a reporting escrow affects reporting of harassment and other workplace misconduct. The project's first phase, in which callers from factories randomly assigned to have access to the escrow were informed about it, provides suggestive evidence that this option increases workers' willingness to report. This pre-registration focuses on the second phase, in which training will be provided in 89 factories to build common knowledge about the escrow among treatment workers.

Keywords: Reporting escrow, retaliation, coordination, gender, garments, Bangladesh
JEL codes: J28, J81, J83, K40, J16

Study pre-registration: OSF Registry; Registration DOI: 10.17605/OSF.IO/P8QFK; Link: https://osf.io/p8qfk; Registry date: November 6, 2021.

1 Introduction

In many developing countries, misbehavior within firms and public administrations often goes unpunished due to weak governance (**???**). Employees whose livelihoods are precarious - who have few alternative job opportunities and little recourse to legal institutions - are especially vulnerable to retaliation if they seek justice. Consequently, these workers are more susceptible to abuse by their superiors. In this context, governance tools that safely provide voice and remedy to workers in developing countries may dramatically improve their welfare.

Recent theoretical work on principal-agent-monitor problems has proposed new whistle-blowing mechanisms that help protect victims and improve information flows. Legal scholars **?** propose the use of reporting escrows to facilitate coordination among multiple victims of harassment. A reporting escrow allows individuals to make a report of misbehavior with a trusted intermediary, who holds that information until a pre-determined number of reports of the same misbehavior have been lodged. Once the threshold number of reports have been lodged, the intermediary acts on the information. Information escrows are increasingly being adopted by organizations to facilitate reporting of certain behaviors; for example, Project Callisto helps universities in the United States set up these systems with the goal of increasing reporting of sexual harassment.

To our knowledge, however, there is no well-identified evidence on reporting escrows' effectiveness. Building on **?** and on economic models of reporting misconduct (**??**), we identify three leading mechanisms through which reporting escrows may increase reporting. The first is reducing coordination problems, as they make it irrelevant who is the first to report, which reduces workers' uncertainty over whether others will report the misconduct. The second is reducing the retaliation risk of reporting, as the information only becomes public when there are multiple victims. Third is reducing the reputational cost of reporting, as they may make public that abuse is a widespread misbehavior as opposed to a consequence of an individual's conduct.

That said, there are also reasons why reporting escrows may not increase reporting, including that workers may not expect other workers to report, may not trust the party maintaining the escrow, or they may have a strong preference for prompt investigations, which are not guaranteed under the escrow system, as the intermediary only acts on the information once the threshold number of reports is reached.

Guided by these economic theories, we conduct a randomized controlled trial (RCT) to study how the availability and design of a reporting escrow affects the reporting and the subsequent investigation of harassment and other forms of workplace abuse. We collaborate with an innovative worker helpline that serves Bangladesh's apparel sector, the Amader Kotha (AK) Helpline, which means "Our Voice" in Bangla. The Helpline's operating model entails apparel buyers' subscribing to the Helpline's services for their supplier factories; supplier factories have strong incentives to cooperate in order to maintain buyers' business. We note that based on evidence from our analysis of Helpline data, discussions with the Helpline, and from focus group discussions and interviews with workers, we identify the mechanisms detailed above – coordination problems, risk of retaliation, and reputational concerns – as key barriers to reporting in our setting.

To distinguish between different barriers to reporting, we randomly assign factories to distinct treatment conditions and study how this affects the reporting and subsequent investigation of labor issues (e.g., physical and sexual harassment). We study two types of reporting escrows: A “reputation” escrow, which counts issues at the factory floor-issue type level, and a “burden-of-proof” or “retaliation” escrow, which counts issues at the manager-issue type level. The former tests the possibility that the key reason why workers do not report issues such as sexual harassment is because of concerns about their reputation. If this is the case, then the fact that their issue is only raised if multiple workers on their production floor report the same issue, which makes clear that that the issue is not specific to the worker’s own behavior, should increase workers’ willingness to report. The latter tests the possibility that the key reason why workers do not report issues such as sexual harassment is because of concerns about retaliation against them by the perpetrator. If this is the case, then the fact that multiple workers at have experienced the same issue by the same manager should increase workers’ willingness to report because it provides safety in numbers and increases the likelihood that the reports are actionable by the Helpline or by managers. Both types of escrows should facilitate coordination among workers, compared to the status quo of no reporting escrow option, which is the control condition. Importantly, workers at both groups of factories continue to have the option of reporting on their own.

We are conducing this RCT in two phases, which are displayed in Figure 1. In Phase 1, which launched in late October of 2021, we randomized the complete list of factories participating in the AK Helpline to an escrow treatment arm; we implemented a stratified randomized assignment of 668 factories to our three treatment arms: 221 to the reputation escrow, 221 to the retaliation escrow, and 226 to control. We then conducted a soft launch of the intervention. By soft launch, we mean that we trained the Helpline operators on the escrow protocols, started informing callers from treatment factories about the escrow reporting option, and began accepting reports via the escrow. We also preregistered the study’s design on the OSF registry at this time.

In parallel, we worked to establish collaborations with participating buyers to conduct onsite, factory-wide worker training sessions on the AK Helpline, which is the experiment’s second phase. In the treatment factories, the training will include information on the reporting escrow; the information about the escrow will be specific to a factory’s assigned escrow type, either the retaliation or the reputation escrow. In the control factories, the training will not include information on the reporting escrow. The onsite training is crucial to understand the potential effects of a reporting escrow because it will help ensure that workers have common knowledge about the escrow’s availability; we argue that common knowledge of the escrow is key to overcome the coordination problem.

We have agreements with two participating buyers to conduct onsite training with a total of 89 factories (37 control, 26 reputation escrow, and 26 retaliation escrow). Consequently, this is our proposed sample size for Phase 2 of the experiment with onsite training.

The two-phase approach to the experiment conveyed two benefits: (1) It allowed us to launch the escrow’s protocols and back-end management system while experiencing a low volume of calls to the escrow. (2) It allows us to compare the effects of the escrow on treatment factories that receive the onsite training with the effects on treatment factories that do not.[1]

[1]As the onsite training is not randomly assigned, we will identify the local average treatment effect of the onsite training on the factories that participate, which could be different from the effect of the onsite training on the factories that do not participate. That said, we have panel data on our primary outcomes, so we will be able to examine trends in the pre-experiment period and in Phase 1, which may lend support to a causal interpretation of any differences that we identify in Phase 2.

As of November 2022, we are finding somewhat encouraging preliminary results from Phase 1 of the experiment, which we report in Section 4.3 on pilot data. During the first 11 months, for our two primary outcomes, we find a 18.5% increase in calls for the less severe types of labor issues and a 80% increase in calls for the severe types of labor issues, although these preliminary effects are not statistically significant.[2] We do not, however, have our full sample size for the experiment's first phase, which will increase the precision of our estimates. Interestingly, these effects are not driven by an increase in reports deposited in the escrow. As we discuss in Section 4.3, we tentatively interpret the results as consistent with increases workers' beliefs that other workers will come forward. We are encouraged by these preliminary findings, but they leave an important question unanswered, which is to what extent workers value the escrow as a reporting option. The experiment's second phase, and its associated surveys of workers, are crucial to answer this question.

In this pre-results report, we focus on Phase 2, the onsite training intervention, which is the crux of our research design. The onsite training is crucial for understanding the escrow's effects because it allows to us to create common knowledge of the escrow among workers. In the final paper, we will report findings on both phases together, with a focus on our main results from Phase 2. If we were to publish the results in the *Journal of Development Economics* (JDE), we would clearly disclose that the Phase 2 results were the focus of our registered report.

2 Research Context

Our empirical setting is Bangladesh's apparel sector, which is critical to the country's economy. 60-80% of workers in the sector are women, while more than 90% of managers are men. The sector has long struggled with an international reputation for poor working conditions and limited labor rights. Workers who experience harassment or other forms of mistreatment by management often have few to no options to obtain resolution: Internal reporting systems are often corrupt or ineffective and legal institutions are weak.

The AK Helpline receives an average of 5,000 calls per month from a total of over 1,000 textile and apparel factories, with over 500 distinct substantive issues per month. Note that this means that many workers at factories that do not participate in the Helpline learn about the Helpline and call, which reflects the Helpline's positive reputation among workers (and likely the lack of alternative reporting options). Since its inception, the Helpline has addressed more than 34,450 substantive issues. We analyzed workers' usage of the Helpline using call data from their staggered expansion from 2014-2018. Using an event study design, we find that substantive calls increase 50-fold following the first training and more than double following subsequent "refresher" trainings.

Despite its successes, the Helpline has faced challenges to empower women in particular to report workplace misconduct, especially sexual harassment. Women are substantially

[2] From our OSF preregistration, less severe labor issue category includes verbal harassment, compensation issues (overtime, wages, benefits), termination, demotion, transfers, discrimination, working hours, leave policies, mobile phone use/access, factory facilities, transportation, and food served at the factory. Severe labor issues include physical and sexual harassment, forced and child labor, extreme compensation issues (e.g., delays in payments), violations of freedom of association and other personal liberties, bribery or corruption, and threat of retaliation for Helpline use. Other types of labor issues may eventually appear in these categories, as we take the Helpline's severity coding of labor issue types as given.

underrepresented in calls to the Helpline: despite comprising 60-80% of workers in the sector, they are responsible for only 23% of formal complaints to the Helpline. Over the past seven years, the Helpline received only 70 sexual harassment cases from among 34,000 issues received during this period. Further, from 2014-18, nearly 30% of callers reporting verbal, physical, or sexual harassment were unwilling to provide permission for the Helpline to pursue an investigation. The Helpline, and our research team, believe that these figures do not reflect that Bangladeshi women do not experience sexual harassment or other issues, but instead that it reflects systemic gender inequalities, norms, and direct threats.

3 Research Design

3.1 Research Questions, Hypotheses, & Outcomes

In this research, we ask: How does the availability and design of a reporting escrow affect reporting and subsequent investigation of harassment and other forms of workplace abuse? We hypothesize that the reporting escrow may increase an individuals' willingness to report, as it reduces coordination problems, the risk of retaliation, and the reputational costs of reporting.

We further hypothesize that the reputation escrow will lead to quicker investigations than retaliation escrow because we look for matching reports in a bigger population, which is a benefit of this approach. But compared to the retaliation escrow, the reputation escrow offers workers less protection against retaliation from the individual perpetrating the misbehavior. This is because the reputation escrow can be triggered by reports about a different individual, so there is less safety in numbers against any one perpetrator. Consequently, if the primary deterrent to reporting is the retaliation cost, it is possible that the retaliation escrow may increase reporting more despite likely aggregating over smaller numbers of workers.

The Helpline categorizes calls into several different groups. Substantive calls about issues at workers' factories can be one of four types: Less severe labor issues, severe labor issues, less severe safety issues, and severe safety issues. The reporting escrow is only available for labor issues due to the Helpline's inability to commit to *not* act upon safety issues such as active fires, which the research team fully agrees with. Further, we hypothesize that the reporting escrow is most relevant for harassment and other forms of workplace misconduct that fall under the labor issue categories, such as bribery or corruption, violation of personal liberties, and so on.

Following the Helpline's issue categorization system, we will examine three primary outcomes:

1. The number of less severe labor issues reported to the Helpline. Less severe labor issues include verbal harassment, compensation issues (overtime, wages, benefits), termination, demotion, transfers, discrimination, working hours, leave policies, mobile phone use/access, factory facilities, transportation, and food served at the factory.

2. The number of severe labor issues reported to the Helpline. We use the term "severe labor issues" to correspond to the Helpline's issue categorization system. The Helpline's categories for severe labor issues include physical and sexual harassment, forced and

child labor, extreme compensation issues (e.g., delays in payments), violations of freedom of association and other personal liberties, bribery or corruption, and threat of retaliation for Helpline use.

3. The number of labor issues escalated for investigation/action by AK Helpline. We include this outcome for the following reason: While we expect the reporting escrow to either increase or not change investigations, since workers always have the option to skip the escrow and raise the issue immediately at any time, if the reporting escrow never triggers, then it will not increase workers' access to remedy. This may be the case if, say, perpetrators only ever abuse one victim.

In our original OSF preregistration, we pre-specified these outcome variables normalizing them to be per 1,000 workers in a factory. The rationale for this choice was that call volume is highly correlated with factory size, so we were concerned that small, chance imbalances in factory size between treatment and control group could bias our estimates of the treatment effects. Subsequent to our preregistration, we determined that it would not be possible to measure employment for all factories participating in the AK Helpline and that the employment information available might be outdated. Consequently, we decided not to normalize the outcome variable to be per 1,000 workers. We will separately report results for the sample with factory size available, controlling for the number of employees in these factories.[3]

For the factories that will participate in the onsite training experiment in Phase 2, because the Helpline collects this information when scheduling the training, we anticipate being able to measure their numbers of employees. Further, because of the smaller sample size, there is an increased possibility that chance baseline imbalances in factory size could contribute to differences in the average number of calls across groups. As a result, if we determine that we are sufficiently confident in the measurement of employment for this group, if it improves baseline balance for our primary outcome variables, we will normalize the outcome variables to be per 1,000 workers. If not, we will take the same approach as above when reporting regression results.

We will also report results separately for each type of labor issue reported to the Helpline. We consider the specific types of labor issues as secondary outcomes that will help us to understand the mechanisms through which the availability of the reporting escrow affects reporting. We hypothesize that the reporting escrow is primarily relevant for issues that affect multiple workers for which workers' face uncertainty about others' willingness to report and for which the social stigma associated with reporting is higher. We think that this is more likely to be the case for harassment, in particular sexual and physical harassment.

We will further explore the mechanisms underlying the reporting escrow's effects in our secondary research questions:

(a) How does the impact of the reporting escrow when there is common-knowledge about it at the factory-level (through training) compare to its impact when only callers to the Helpline are informed?

(b) What is the impact of the reporting escrow on workers' mental health?

[3] We are updating our OSF preregistration concurrently with submitting this pre-results report, including uploading this pre-registration for Phase 2 of the experiment.

(c) How does workers' propensity to join an escrow depend on the information they have about the number of other workers who have already joined?

For (a), we will examine this question by testing for differences in the reporting escrow's effects on our primary outcomes among the group of factories that participate in the onsite training compared to the group that does not participate in the onsite training. As the onsite training is not randomly assigned, we acknowledge that differences in these effects are not causal. That said, we will have panel data on our primary outcomes, so we will be able to examine trends in the pre-period and in Phases 1 and 2, which may lend support to a causal interpretation of any differences that we identify in Phase 2.

For (b), we will test for effects on an index of mental health measures, which we enumerate in Appendix section B.1. Mental health is our only non-reporting-related secondary outcome. We focus on mental health because in the short- to medium- term, we expect that reporting escrows improve workers' welfare by making them more comfortable to report, seeing more of their coworkers' issues being resolved, and possibly, reducing workplace misconduct. If these effects are present, we expect workers to feel less stress, anxiety, and depression.

For (c), if we find that workers begin to report via the escrow, we will investigate how workers' propensity to join an escrow depending on the information they have about the number of people who have already joined. This will inform possible messaging policies and will clarify the importance of frictions limiting participation. To estimate this, we plan to use randomized callbacks to workers who report to the Helpline whose issues are eligible to be deposited in an escrow account that has at least one report, in which we provide information about the status of reports via the escrow.[4]

3.2 Intervention

Assignment to treatment. We randomly assign factories whose workers have access to the AK Helpline to one of three groups:

- Treatment group 1: Reputation reporting escrow (escrow at factory floor-issue level). In this version of the escrow, if 2-5 workers (depending on issue type) on the same production floor report a similar issue, the reporting escrow will trigger, even if the manager perpetrating the misconduct is different. We denote this as the "reputation reporting escrow," as it tests the possibility that a key reason why workers do not report issues such as sexual harassment is because of concerns about their reputation. If this is the case, then the fact that multiple workers on their production floor have experienced the same issue should increase workers' willingness to report.

- Treatment group 2: Retaliation reporting escrow (escrow at manager-issue level). In this version of the escrow, if 2-5 workers (depending on issue type) report a similar issue with a given manager, the reporting escrow will trigger. We denote it as the "retaliation reporting escrow," as it tests the potential that a key reason why workers do not report issues such as sexual harassment is because of concerns about retaliation by

[4] If we find that workers take-up the escrow, we are also considering developing a structural model to evaluate the value of possible escrow and non-escrow reporting mechanisms to workers. In this case, we will also use our measures of workers' propensity to join the escrow following callbacks in our estimation.

the perpetrator. If this is the case, then the fact that multiple workers have experienced the same issue by the same manager should increase workers' willingness to report.

- Control group (no reporting escrow).

When assigning treatment, we stratified factories by whether or not they supplied to a coalition of multinational corporations that source from Bangladesh (Nirapon), whether or not they received a safety training being provided by Nirapon,[5] above or below median factory size or missingness of factory size information, and above or below median number of calls to the Helpline in the previous six months.

Experimental Phases. As detailed in Figure 1 and in Table 2, we are conducting this experiment in two phases; in this pre-results report, we focus on the experiment's second and main phase, the onsite training phase.

- **Phase 1** launched in October 2021. This phase entailed randomization of the complete list of factories participating in the AK Helpline to an escrow treatment arm and implementation of a soft launch of the reporting escrow. By soft launch, we mean that we trained the Helpline operators on the escrow protocols, started informing callers from treatment factories about the escrow reporting option, and began accepting reports via the escrow. Helpline operators inform all callers from treatment factories about the new reporting option. They do so at the end of each call and follow distinct scripts developed for each treatment group.

- **Phase 2** launched on November 16, 2022.[6] In this phase, we conduct factory-wide worker training with 89 factories supplying to two cooperating buyers. The training sessions entail 30-40 minute training sessions that include a powerpoint presentation about the Helpline and a role play in which workers from the audience act out a problematic workplace scenario and conduct a live call to the Helpline for its assistance. Both treatment and control factories will participate in the onsite training program. In the treatment factories, the training curriculum includes information on the reporting escrow that is specific to a factory's assigned escrow type. Factories also receive cards with the Helpline's phone number for workers to keep with their employee ID badges, 20 posters, 20 large stickers, and 20 danglers, all with the Helpline's information, per 1000 workers.[7] Treatment factories receive 10 additional danglers advertising the reporting escrow per 1000 workers. For control factories, to hold the overall number of Helpline danglers fixed, they receive 10 additional standard danglers per 1000 workers.

[5] This variable was included because it would impact the timing of when suppliers to Nirapon members would be able to participate in training, if they were to be included in the training group. To prevent production disruptions, buyers aim to space out the implementation of different training interventions, so for factories that had recently participated in the safety training program, there would be a cool-off period until they were eligible to participate in the Helpline training.

[6] At the time of submission on November 29, 3 factories have started participation in the onsite training, and the research team does not yet have access to data on outcome variables subsequent to the launch of Phase 2.

[7] Danglers are posters that hang from the ceiling.

In factories that participate in the onsite training, the Helpline will conduct the training either using a train-the-trainer (TtT) model or by directly providing the training to workers. In the TtT model, the Helpline's training team will conduct 2-3 days' worth of training sessions directly to workers and will train the factory's HR personnel to conduct the remaining training sessions. The Helpline then monitors the implementation of the training sessions through the role play training phone calls to the Helpline and a one-day verification visit to the factory. The number of days of training provided by the Helpline team will depend on the buyer and on the factory's size; our analysis will use extensive margin variation in training participation and will not use intensive margin variation in training implementation by the Helpline.[8]

We emphasize that assignment to onsite factory training is not random but is based on whether factories supply to buyers that are participating in the research project. Within this group, we will verify that treatment and control factories are balanced on their characteristics.

The duration of Phase 2 will depend on how long it takes to roll onsite training out to the 89 factories. We will collect the Helpline's call data until 12 months after the final factory participates in the onsite training; we expect to complete onsite training with all factories by June 2023 at the latest, which implies that we will complete Phase 2 data collection by June 2024 at the latest.

3.2.1 Participation in the experiment

Due to the dynamic nature of buyers' supplier bases, there is churn in the factories that participate in the Helpline. This implies that there will be attrition from the sample, as workers at factories that stop supplying to participating buyers lose access to the Helpline. Between March 2021 and July 2022 (18 months), a total of 63 out of 668 factories (9.4% of factories originally randomized) became ineligible for the Helpline. Even though the Helpline no longer services calls from workers employed in ineligible factories, however, we can still observe calls from these factories. So, we will present the robustness of our main results to including attrited factories' observations in our estimation sample.

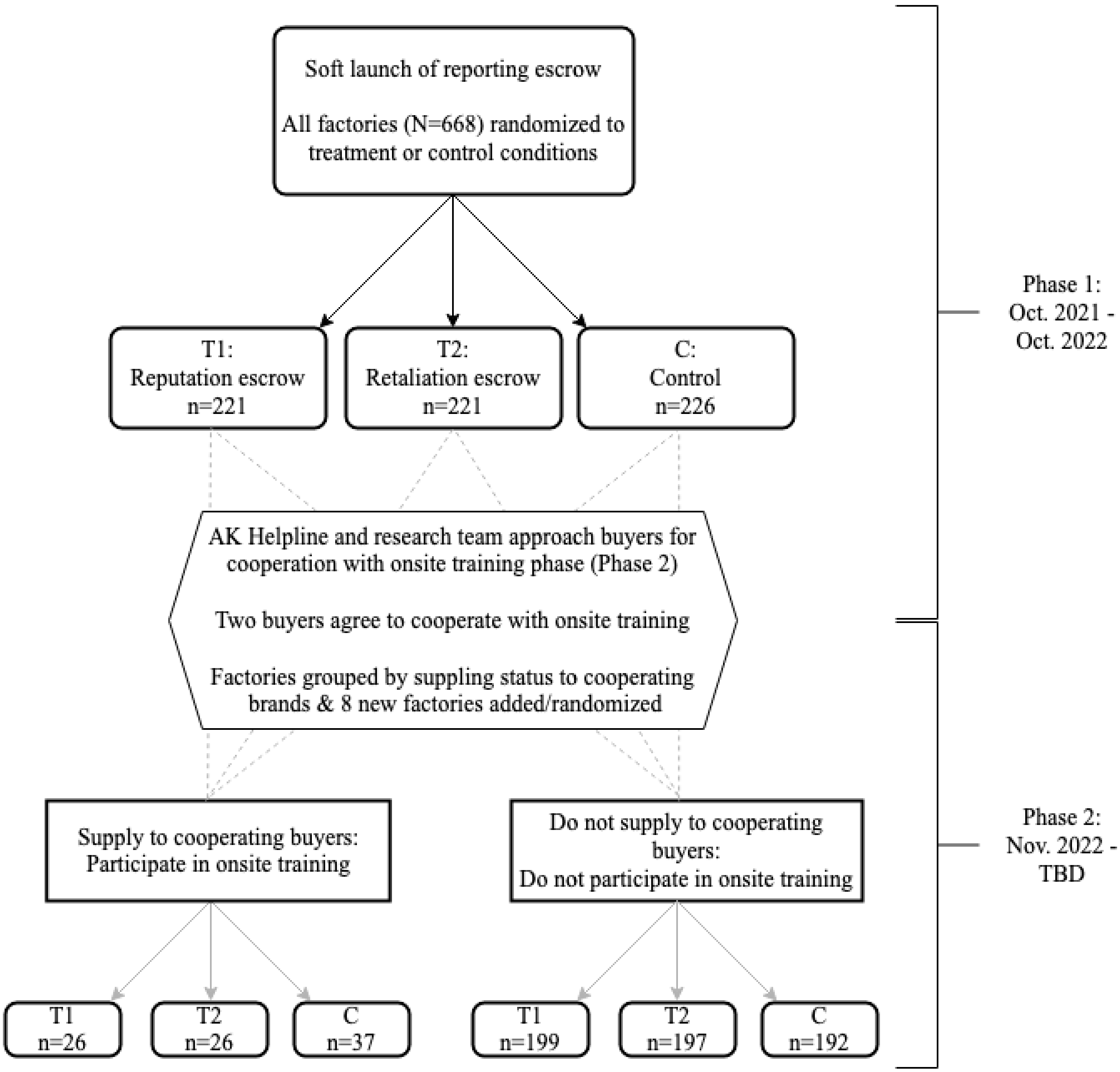
For buyers that are cooperating with the training, we are incorporating new factories that they have added by November 2022. This is a total of 6 factories.[9] For these factories, we stratified them using as many of the variables above as possible, subject to the constraint that there are at least three factories per stratum.

In Phase 2, among factories that supply to buyers that are cooperating with the training, we have obtained the buyers' lists of factories from the Helpline. Among these factories, we will consider participation in the training to be completion of at least one day of onsite training by the Helpline team. We anticipate very high rates of compliance with the onsite training. We will verify that the participation rates are similar for treatment and control factories.

[8] We collaborate with two participating buyers. One buyer has opted for an approach in which the Helpline's training team provides fewer days of training per factory and completes the training program more quickly. The other buyer has opted for more days of training per factory,which means that it will take longer for the Helpline to complete the training with all of this buyer's supplier factories.

[9] Three new factories supplying to a third buyer that ultimately decided not to participate in the training were also randomized and are in the non-onsite training group.

Figure 1: Reporting escrow experiment design & phases



3.2.2 Information provided to factories and buyers about the research

The Helpline's operating model depends on its being trusted as a neutral third party by buyers, apparel factories, and workers. Consequently, we and the Helpline are being transparent with buyers and factories about the fact that the Helpline is evaluating the reporting escrow using a RCT. Our approach to transparency was for our research team to publish a short article about the reporting escrow in the Helpline's quarterly newsletter in Q3 of 2021. The newsletter is distributed to its partners and is publicly posted on the Helpline's website.

While information about the RCT is publicly available, we do not anticipate that factories' managerial teams are especially likely to be aware of it or that it is salient to them. If it is, we expect that if factories respond, it will be by reducing misbehavior, which would push against our finding treatment effects.

3.3 Sample size & statistical power

Our sample size per treatment arm is based on power calculations by simulation conducted using the "pcpanel" package in Stata. We used existing Helpline call data for the simulations. Using call data, we constructed the number of less severe labor issues and the number of severe labor issues at the monthly-level (without normalizing to be per 1000 workers). Our unit of analysis is a factory. We make the following assumptions:

1. Significance level: We set $\alpha = 0.05$ and apply the conservative Bonferroni adjustment with 3 primary outcomes $0.05/3 = 0.016$.
2. We assume 12 post-treatment observations per factory and an ANCOVA specification that controls for the mean of 6 months of pre-treatment calls.
3. We conduct a stratified randomization and use a sample size of 87 factories due to the strata sizes.
4. We use the actual share of factories assigned to treatment in the onsite training group: 59%.

With these assumptions, for the onsite training experiment in Phase 2, we are powered to detect a 0.2 standard deviation (sd) increase in less severe labor calls with 81.4% power. We are powered to detect a 0.25 sd increase in severe labor calls with 80.6% power.

For the Phase 1 experiment with the AK Helpline's full factory base, our sample size at the time of planning was 218-220 factories per treatment arm. With the same assumptions for (1) and (2) above, assuming a stratified randomization with 656 factories, and assuming 67% of the sample allocated to treatment, we are powered to detect a 0.1 sd increase in less severe labor calls with 82.2% power. We are powered to detect a 0.1 sd increase in severe labor calls with 79.2% power.

4 Data

4.1 Data collection and processing

We use four sources of data:

1. *Helpline's call record dataset*: This dataset provides a comprehensive set of characteristics about every call received by the Helpline. It includes information about the nature of the report, the reporting actions taken by the caller prior to calling, the actions taken by the Helpline and by other parties to resolve the issue being reported, and the issue's resolution status. It also includes a variable that indicates whether a call is reported via the reporting escrow. Our three primary outcomes will be constructed using this dataset.

2. *Helpline's other administrative data*: The Helpline will also provide information on factories' participation in the Helpline, on factories' status as suppliers to buyers that are participating in the research, and for factories that will participate in onsite training, the implementation data for these visits. We will use these data sources to determine our sample, to conduct randomization, to track entry and exit from participation in the Helpline, to determine the timing of factories' participation in onsite training, etc.

3. *Surveys of Helpline callers from onsite training factories*: We will conduct a survey of callers to the Helpline from factories that will participate in onsite training. The phone-based survey will include questions about the caller's satisfaction with their call, economic well-being and empowerment, mental health, employment and job satisfaction, reporting preferences, and among treatment group callers, their awareness of the reporting escrow. We will match these surveys to callers' demographic and employment information, which will enable us to test for effects on the composition of callers to the Helpline. We will conduct 3 rounds of surveys, once before the onsite training is launched, the second 2-3 months after a factory is trained, and the third 5-6 months after a factory is trained. We will primarily use these survey data for three types of analysis: First, to test whether the composition of callers to the Helpline changes due to the escrow in terms of demographics, economic well-being, and/or empowerment. Second, to test how aware callers from treatment factories are of the reporting escrow's availability. Third, to examine whether callers from treatment factories report greater satisfaction with their experience using the Helpline.

4. *Endline survey of workers from onsite training factories*: We will conduct an endline survey with workers at onsite training factories that will take place 5-6 months after the factory's participation in onsite training. We will recruit participants through the community. The endline survey will be representative at the factory-level and will be more in-depth. It will include questions about mental health, experience of harassment and other workplace issues, social acceptability around reporting and related behaviors, reporting through internal factory channels, and awareness of the Helpline (and at treatment factories, of the reporting escrow).

The main rationale for the endline survey is to provide the data for our secondary research question on workers' mental health, as this will be our main source of data for this analysis. We will also use both types of survey data to help disentangle the mechanisms through which the escrow may affect reporting.

4.2 Variations from the intended sample size

There will be variations from the intended factory sample size due to churn in buyers' supplier bases, which results in churn in the set of factories served by the Helpline. This will result in exit from the sample. Between March 2021 and July 2022 (18 months), a total of 63 out of 668 factories (9.4% of factories originally randomized) became ineligible for the Helpline. Assuming that the exit rate is constant, this implies that 3.5 factories exit per month, so we can expect 6% of the supplier base to exit per year. For factories that are *not* participating in the onsite training program, we will not replace factories that exit. For factories that are participating in onsite training, we have the opportunity to replace those that exit. We will report the counts of factories that exit and enter and the implications for our total sample size in the paper. We will also address attrition by presenting our main results including attrited factories' observations in our estimation sample.

4.3 Phase 1 data

We use the data from Phase 1 of the experiment, the soft launch of the reporting escrow, to analyze the preliminary treatment effects of informing Helpline callers about the escrow. These results are preliminary because we do not yet have the call data for October 2022, which is the final month prior to the launch of Phase 2. In the appendix, we use Helpline data prior to the launch of Phase 1 to verify that baseline variables are balanced across treatment arms for all experimental factories (Table B.2), as well as for the subset of factories with onsite training (Table B.3, Figures B.2-B.3).

We run the following analysis of covariance (hereafter, ANCOVA) specification:

$$Y_{jt} = \beta\, Escrow_j + \gamma Y_{j,t=0} + \mu_s + \lambda_t + \theta X_j + \epsilon_{jt} \tag{1}$$

Y_{jt} is the number of calls of a particular type to the Helpline from factory j in month t, where t refers to any month in the Phase 1 period for which we have data (November 2021-September 2022). $Escrow_j$ denotes whether a given factory j is assigned to the escrow-treatment arm. $Y_{j,t=0}$ is the average number of monthly calls from factory j over the six months before the experiment's launch (April - September 2021). We exclude the month of the launch (October 2021), as the experiment was launched on October 19, 2021, and we needed to work with the Helpline to adapt to the experiment's protocols. We also control for stratum fixed-effects μ_s, and for time fixed effects λ_t.

We also report results separately for each type of escrow; we use $RepEscrow_j$ and $RetEscrow_j$ to denote whether the factory is assigned to the reputation or the retaliation escrow, respectively, and we estimate:

$$Y_{jt} = \beta_1\, RepEscrow_j + \beta_2\, RetEscrow_j + \gamma Y_{j,t=0} + \mu_s + \lambda_t + \theta X_j + \epsilon_{jt} \tag{2}$$

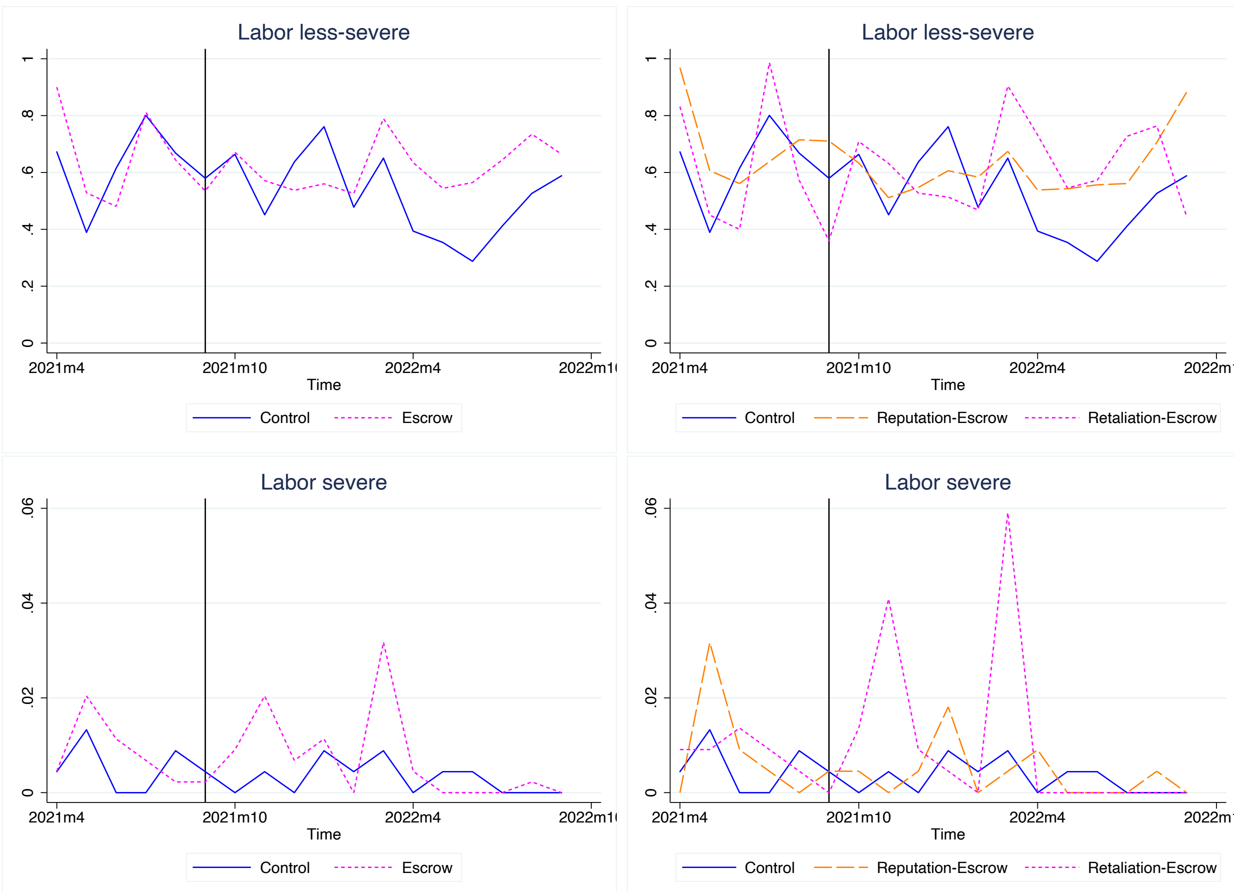
In equation 1, the coefficient of interest is β, which is the average treatment effect (ATE) of the reporting escrow. In equation 2, the coefficients of interest are β_1, the ATE of the reputation reporting escrow, and β_2, the ATE of the retaliation reporting escrow. As discussed in Section 3.1, if reputational concerns are the primary deterrent to reporting, we expect $\beta_1 > \beta_2$. In contrast, if retaliation concerns are the primary deterrent to reporting, we expect $\beta_1 < \beta_2$.

Phase I Results. In Figures 2, we plot the average number of less severe and severe labor calls to the Helpline per factory-month, by treatment arm, before and after the launch of Phase I. For both variables, the panels show that the treatment and control groups had similar levels and trends in the six months prior to the experiments' launch. The top panel, which reports results for less severe labor issues, shows that it takes several months, but calls from escrow factories begin to diverge from those from control factories in March 2022 and subsequently remain elevated compared to calls from control factories. The gap between the experiment's launch and the emergence of a treatment effect is consistent with our expectations; during Phase 1, we have only been informing callers to the Helpline, and relative to factories' sizes, only a very small share of workers call the Helpline each month. Consequently, we expected that if there were a treatment effect, it would not materialize until after enough time had passed for information about the reporting escrow to spread in treatment factories. The second panel presents less clear evidence that severe labor calls may be increasing; there is a somewhat higher call volume from treatment factories between October 2021 and April 2022, but it is not sustained.

Regression results in Table 1 confirm the patterns in the raw data. They indicate that escrow factories saw an increase in average less severe labor calls by about 18.5% in the 11 months after the launch of the escrow, although the difference is not statistically significant. When subsetting to factories for which employment data is available, and controlling for factory size, the effect's magnitude is very similar, but is now statistically significant at the 10% level. In both cases, the effects for the reputation and retaliation escrow are similar. We also observe an increase in severe labor calls, both in Figure 2 and Table B.2, driven by the retaliation escrow, but the effect is not statistically significant.

Interpretation of the Results Interestingly, the increase in less severe labor calls is not driven by an increase in reports deposited in the escrow. What, then, is the mechanism underlying the effects? One possibility is increased trust in the Helpline, as Helpline operators inform all callers from treatment factories about the escrow, telling them that the Helpline is offering this new reporting option to help workers who would otherwise not feel comfortable reporting alone to feel safe. If increased trust is the mechanism, we might also expect to see an increase in occupational safety and health (OSH)-related calls. As we report in appendix Figure B.1 and Table B.1, though, there is no effect on OSH-related reporting. Given this, our preferred interpretation is that the availability of the escrow for labor issues increases workers' beliefs that other workers will come forward to report labor issues. Our worker survey will provide more conclusive insights, however.

Figure 2: Average monthly labor calls to the Helpline per factory by treatment arm



Notes: Average number of substantive calls to the Helpline, per factory, per month, by type of issue, for the original set of factories that entered the experiment in Phase I. The black vertical lines reflect the month before the launch of Phase I of the experiment (October 19, 2021). We exclude 1 outlier-factory (more than 200 calls in one month in the pre-treatment period). Data is prior to the launch of Phase II of the experiment.

Table 1: Phase 1 treatment effects on number of labor issues reported to the Helpline per month

	Labor less-severe				Labor severe			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Escrow	0.110	0.160**			0.004	0.006		
	(0.078)	(0.081)			(0.004)	(0.005)		
Reputation-Escrow			0.105	0.196*			0.001	0.002
			(0.098)	(0.104)			(0.003)	(0.003)
Retaliation-Escrow			0.115	0.125			0.007	0.009
			(0.101)	(0.107)			(0.008)	(0.009)
Observations	7337	6325	7337	6325	7337	6325	7337	6325
Factories	667	575	667	575	667	575	667	575
Baseline control mean	0.621	0.621	0.621	0.621	0.005	0.005	0.005	0.005
Coef. comparison p-val.			0.93	0.59			0.41	0.42
Strata FE	Y	Y	Y	Y	Y	Y	Y	Y
Month FE	Y	Y	Y	Y	Y	Y	Y	Y
Factory Size[1]		Y		Y		Y		Y

Notes: The table reports preliminary impacts (prior to the launch of Phase II of the experiment) of the reporting escrow, as well as separately of the reputation and the retaliation escrow, for the original set of factories that entered the experiment in Phase I. We exclude 1 outlier-factory (more than 200 calls in one month in the pre-treatment period); note that including it does not affect the nature of the results. ANCOVA regression, conditioning on post-experimental period (November 2021 - September 2022). We control for calls at t=0, i.e. the average calls in the 6 months before the launch. The dependent variables reported are the number of less-severe labor calls to the Helpline, and the number of severe labor calls to the Helpline. (Robust standard errors) clustered by factory are reported. $^{*}p < 0.10, ^{**}p < 0.05, ^{***}p < 0.01$.

Footnotes: [1] As the Helpline is preparing for the onsite training, it has been collecting updated factory size information for the sample of factories with onsite training. In this analysis, we combine our previously-available data on factory size with the updated data to improve coverage. We will continue to collect factory size as the training rolls out to the onsite training factories.

4.4 Timeline

We expect the study to be completed by September 2023. Details of our realized and planned timelines are shown in Table 2.

Table 2: Timeline

I. Realized timeline:	Dates (start-end)	Details
IRB approval	Sept 30, 2020	Ben-Gurion U. (Sept, 2020; updated in Aug 2022), Columbia U. (July 2021)
Training of reporting escrow manager	Aug - Oct 2021	Training on escrow protocols and maintenance
Training of Helpline operators	Sept 15-16, 2021	Training on screening factories, providing information, and handling escrow reports
Randomization for Phase I	Oct 2021	Randomized assignment of full population of Helpline factories into treatment arms
Pre-registration	Nov 6, 2021	OSF Registration
Phase I experiment: Soft launch of escrow	Oct 19, 2021	Soft launch of the escrow by informing callers to the Helpline & accepting reports
Refresher training of Helpline operators	Aug 25, 2022	Phase 1 debrief and refresher training with Helpline operators for Phase II
Pilot survey	Oct 2022	Pilot survey with Helpline callers in non-training factories

II. Planned timeline:	Dates (start-end)	Details
Phase II pre-analysis plan & update OSF pre-registration	Nov 2022	Update pre-registration in OSF & publish pre-analysis plan
Baseline survey of Helpline callers	Nov 2022 - March 2023	Baseline survey with Helpline callers in training factories
Phase II experiment: Roll-out of onsite training	Nov 16, 2022 - June 2023	Onsite training is rolled out to supplier factories of participating buyers
Midline survey of Helpline callers	Feb 2023 - Sept 2023	Midline survey with Helpline callers in training factories
Endline survey of Helpline callers	May 2023 - Jan 2024	Midline survey with Helpline callers in training factories
Endline survey	May 2023 - Jan 2024	Endline survey with a representative sample of workers recruited through the community

5 Statistical methods

Analogous to the approach presented in 4.3, we will begin our analysis with visual evidence from figures that plot the means of each of our primary outcomes by group at the monthly frequency. For groups, we will group factories by their membership in the onsite training group and by their treatment arm. We will present the data first pooling and then separating the reputation and retaliation escrow treatment arms.[10] In the appendix, we will also show the results using the Phase 1 randomization groups (i.e., not separating the onsite training and non-onsite training groups). The figures will clearly indicate the study's key milestones: The beginning of Phase 1 (reporting escrow's soft launch), the beginning of Phase 2 (launch of onsite training), and the completion of onsite training. We will also show one figure with a grouping by treatment status for the pre-experiment launch and Phase 1, for all factories (i.e., as in Figure 2 in this pre-results report).

The figures will show whether the treatment and control groups in the onsite training and no onsite training experiments, respectively, are balanced on means and on trends prior to the reporting escrow's soft launch.[11] The figures will also show the pre-Phase 2 trends in the (non-randomly-assigned) onsite training and offsite training groups, although there is no requirement that they be on similar trends, as we will estimate the treatment effects within each group. Importantly, they will allow us to check baseline balance *within* the non-onsite training group and the onsite training group, respectively, across treatment and control groups.

We will then proceed to estimate the treatment effects using regression. To begin, we will use the following ANCOVA specification:

$$\begin{aligned} Y_{jt} = \beta_1\, Escrow_j + \beta_2\, Train_j + \beta_3\, Phase2_t + \beta_4\, Escrow_j * Train_j \\ + \beta_5\, Escrow_j * \ Phase2_t + \beta_6\, Train_j * \ Phase2_t \\ + \beta_7\, Escrow_j * Train_j * \ Phase2_t + \gamma Y_{j,t=0} + \mu_s + \lambda_t + \theta X_j + \epsilon_{jt} \end{aligned} \tag{3}$$

where Y_{jt} is the outcome of interest for factory j at time t. $Escrow_j$ is an indicator for randomized assignment to the escrow treatment group. $Phase2_t$ is an indicator for the time period after the onsite training is launched with the onsite training group.[12] The Phase 1 time period encompasses the period between when the reporting escrow was softly launched (October 19, 2021) and when the onsite training was launched (November 16, 2022). $Train_j$ is an indicator for non-randomized assignment to the onsite training group. We control for the baseline value of the dependent variable, for stratum fixed-effects μ_s, and for time fixed effects λ_t. We will also present results including factories' pre-treatment characteristics X_j that are selected using the post double selection (PDS) lasso (**?**). We will cluster standard errors by factory.[13]

[10] If we do not find statistical evidence that the reputation and retaliation escrows have different effects, we may present the separate results in the appendix to reduce the main paper's length.

[11] We will also present standard balance tests using difference-in-means tests implemented using regression.

[12] We will code this from the first date of training at the first factory to participate in training, which is November 16, 2022.

[13] In the appendix, we will also report the results using the Phase 1 randomization groups: $Y_{jt} = \beta_1\, Escrow_j + \beta_2\, Phase2_t + \beta_3\, Escrow_j * \ Phase2_t + \gamma Y_{j,t=0} + \mu_s + \lambda_t + \theta X_j + \epsilon_{jt}$. In this equation, the

Equation 3 is the main regression specification of interest. In this specification, we are interested in several local average treatment effects (LATEs). First, we are interested in the LATEs for the reporting escrow for the non-onsite training group and the onsite training groups during Phase 1, which are β_1 and $(\beta_1 + \beta_4)$, respectively. Second, we are interested in the LATE for the reporting escrow for the non-onsite training group during Phase 2, which is $(\beta_1 + \beta_5)$. Third, our primary focus in this pre-results report is the LATE for the reporting escrow with onsite training for the onsite training group during Phase 2, which is $(\beta_1 + \beta_4 + \beta_5 + \beta_7)$.

We will separately test the effects of the reputation and retaliation reporting escrows. We continue to denote the former as "RepEscrow" and the latter as "RetEscrow."

$$\begin{aligned} Y_{jt} = \beta_1\, RepEscrow_j + \beta_2\, RetEscrow_j + \beta_3\, Train_j + \beta_4\, Phase2_t \\ + \beta_5\, RepEscrow_j * Train_j + \beta_6\, RetEscrow_j * Train_j + \beta_7\, RepEscrow_j * Phase2_t \\ + \beta_8\, RetEscrow_j * Phase2_t + \beta_9\, Train_j * Phase2_t \\ + \beta_{10}\, RepEscrow_j * Train_j * Phase2_t + \beta_{11}\, RetEscrow_j * Train_j * Phase2_t \\ + \gamma Y_{j,t=0} + \mu_s + \lambda_t + \theta X_j + \epsilon_{jt} \quad (4) \end{aligned}$$

In equation 4, we are interested in the reputation and retaliation escrow-specific LATEs for each subgroup and phase referenced when discussing equation 3. We are also interested in the difference in treatment effects between the two types of escrows. As discussed in Section 3.1, if reputational concerns are the primary deterrent to reporting, in equation 4, we expect $\beta_1 > \beta_2$, $(\beta_1+\beta_5) > (\beta_2+\beta_6)$, $(\beta_1+\beta_7) > (\beta_2+\beta_8)$, and $(\beta_1+\beta_5+\beta_7+\beta_{10}) > (\beta_2+\beta_6+\beta_8+\beta_{11})$. In contrast, if retaliation concerns are the primary deterrent to reporting, we expect the reverse.

To examine the dynamics in the treatment effects, we will estimate the effects separately for each post-treatment month and will plot them.

For the onsite training group, the Helpline is rolling out the training as quickly as possible, but it will likely take several months. Consequently, we will also use a difference-in-difference (DiD) event study design to estimate the effects of the escrow with onsite training. We will estimate:

$$Y_{jt} = \alpha_j + \delta_t + \sum_{k=-6}^{12} \lambda_k\, Train^k_{jt} + \sum_{k=-6}^{12} \beta_k\, Train^k_{jt} * Escrow_j + \epsilon_{jt} \quad (5)$$

where t indicates month, and we will estimate results using $k = -6$ months before through $k = 12$ months after, respectively, a factory participates in onsite training, α_j are factory fixed effects, and δ_t are month fixed effects. The omitted term is the interaction of the treatment with the month prior to participation in the Helpline's training.

The set-up in equation 5, with more than two time periods and variation in the timing of when factories participate in the onsite training, diverges from the canonical DiD design. Consequently, it is important to clarify the implications of the recent econometrics literature

coefficients of interest are β_1 and $(\beta_1 + \beta_3)$, which are the ATEs of the reporting escrow on the full population of AK-covered factories during Phases 1 and 2, respectively. We will also present results separating the reputation and retaliation-escrow arms.

on this topic (summarized in **?**) for our analysis. The key issue that the literature identifies is one of "forbidden" comparisons between already-treated units in a staggered DiD setting when treatment effects are allowed to be heterogeneous across units or time (**?**). Our main coefficients of interest, the β_k, do not face this problem, as their estimation makes comparisons between treated units and pure control units. If, however, we want to interpret the coefficients on the λ_k, then we do need to consider the implications of the recent literature. At the time of writing, the most relevant papers to our setting are **?**, **?**, and **?**, which focus on estimation of dynamic treatment effects. These papers propose estimators and approaches to inference; we plan to report the estimates using these approaches in the paper or appendix.[14] We will have access to control factories in the non-onsite training group that we anticipate will be able to serve as pure controls to estimate the effects of standard Helpline training for the onsite training group.

We note that it is plausible that the treatment and control factories will be different in levels and/or in trends, in the six months leading up to their participation in the onsite training. This is because treatment factories have access to the reporting escrow during these months. If we find that the treatment and control groups are different in levels or in trends in the months leading up to the onsite training, in the paper, we will discuss the implications for the interpretation of these results.

Finally, we will estimate the effect of the escrow on workers' mental health using equation 6:

$$Y_{ij} = \beta_1 \, Escrow_j + \mu_s + \theta X_{ij} + \epsilon_{ij} \tag{6}$$

We measure the mental health of worker i from factory j using data from the endline survey of workers from onsite training factories. The vector X_{ij} includes worker and factory-level controls that are selected using the post double selection (PDS) lasso (**?**). If we find that the intervention affects the composition of callers to the Helpline, we will not include worker-level controls in these regressions, as they will be measured at endline and may be affected by the treatment.

Outliers: We will follow standard practices for identifying and addressing outliers (e.g., presenting results including the outlier and showing pre-treatment balance and the main the results after dropping any outliers).

Multiple hypothesis testing: For our three primary outcome variables, we will show p-values adjusted to control the False Discovery Rate. For the mental health variables, which are listed in Appendix B.1, we will aggregate them into an index. We will construct the mental health index following the procedure from **?**.

Heterogeneous effects (HTEs): We hypothesize that the treatment effects may be heterogeneous along two dimensions. First, we hypothesize that the effects will be smaller if there is a trade union or other form of collective worker voice organization present at the factory. While there are multiple reasons why this might be the case, we are interested in the possibility that the worker voice institution already exists to help coordinate multiple

[14] The econometrics literature on DiD is rapidly evolving; if needed, we will update our estimation approach based on the latest research.

victims to come forward. Alternative mechanisms that could generate HTEs by collective worker voice organizations include improved information flows between workers and managers and workers' having greater bargaining power, both of which may result in lower incidence of labor issues in these factories (**?**).

The second dimension that we are interested in testing is gender composition. As discussed in section 2, women are dramatically underrepresented in calls to the Helpline, as are issues that may differentially affect women, such as sexual harassment. In principle, the reporting escrow is most relevant for workers who would otherwise not come forward under the status quo system, so it's plausible that women are more likely to be in this group. Further, there is substantial interest in using reporting escrows to increase detection of serial sexual harassment cases, as illustrated by the example of Project Callisto. Consequently, we will test for HTEs by factories' gender composition.

We note that we will only be able to measure covariates for the onsite training group, and that there is uncertainty about whether we will be able to measure these variables even for this group. The reason for this uncertainty is that the Helpline has limited information about participating factories, so we need to collect this information ourselves. During the process of scheduling the training, we are collecting information about factories' number of employees and number of employees who are women. For the worker voice measures, we will collect these measures during the first round Helpline caller survey, but we will not be able to survey workers from all factories in the first round, as some factories are new to the Helpline and do not have callers. Thus, the decision of whether to test for HTEs will depend on the number of factories that we can collect data for and whether we have baseline balance between treatment and control factories within subgroups with this information available.

6 Interpreting the Results

As we have described in this pre-results report, our research design and data sources make us well-positioned to disentangle the mechanisms through which the reporting escrow may increase workers' willingness to report labor issues at their factories. Our design of testing two variants of the reporting escrow, the reputation and retaliation escrows, will allow us to disentangle the importance of these two barriers to reporting. Our approach of conducting surveys of Helpline callers will allow us to examine whether the escrow changes the types of workers who report and how satisfied workers are with their experience reporting through the Helpline. Finally, our approach of using randomized callbacks to Helpline callers whose issues are eligible to reported via an active escrow account will inform possible messaging policies and will clarify the importance of frictions limiting participation.

We believe that the evidence from this research will be highly policy relevant as, to our knowledge, there are no rigorous, randomized evaluations of reporting escrows. At the same time, reporting escrows are increasingly being advanced as policies to alleviate sexual harassment in a variety of organizational settings (e.g., Project Callisto helps universities in the United States set up these systems). We know very little about how a reporting escrow's availability and design affects reporting of issues such as workplace harassment. Further, as discussed above, there are multiple mechanisms through which reporting escrows may affect reporting; we lack evidence of which among these mechanisms are empirically most important. This evidence could improve not only the design of reporting escrows, but of "standard" or status quo reporting systems.

Finally, our preliminary results from Phase 1 of the experiment suggest an increase in labor-related reporting, but not through the reporting escrow. As we discussed in the Introduction, we think it is very plausible that the reporting escrow's effects depend on workers' having common knowledge of it. Hence the importance of the Phase 2 onsite training experiment. That said, it is possible that we will continue to find null results on reporting via the escrow among onsite training factories in Phase 2. Given the substantial interest in reporting escrows among practitioners and academics, we argue that an informative zero would be an important finding. In this scenario, we would provide several pieces of evidence on possible mechanisms underlying the null results. Given the data that we will have access to through the Helpline and through the worker surveys, we would be equipped to provide suggestive evidence on the following mechanisms:

1. Lack of awareness of escrow
2. Lack of understanding of escrow
3. Strong preference for speedy, possibly less rigorous, resolution compared to slower, possibly more rigorous, resolution
4. Low perceived additional protection of group-based reporting
5. Beliefs that other workers are not experiencing similar issues
6. Beliefs that other workers are experiencing similar issues but will not report
7. Beliefs that even if the worker reports via the escrow, the Helpline will not refrain from raising worker's issue immediately
8. Actual nature of labor issues tends to entail single worker being affected

A Administrative information

Funding

This work was supported by G2LM|LIC [5-510, 2020]; the Israeli Science Foundation (ISF) [Personal Research Grant 860/20, 2020]; the J-PAL Gender and Economic Agency Initiative [AABR6319, 2021]; and the Robert Wood Johnson Foundation [98575, 2021].

Institutional Review Boards / Ethics approvals

This research has received ethics approval from the Human Subjects Research Committee at Ben-Gurion University on September 30, 2020 [request number 1927-1] for Phase I of the experiment, and in August 11, 2022 [request number 238-2] for Phase II of the experiment. It has also received approval from the Institutional Review Board at Columbia University on July 23, 2021 [protocol number IRB-AAAS7822].

Declaration of interest

The authors declare no competing interests.

Acknowledgments

We gratefully acknowledge excellent research assistance from Ms. Akriti Dureja, Mr. Feroz Golam Rosul, Mr. Krishna Kampelali, Ms. Ferdausi Sumana, and Dr. Sabina Yesmin. We are thankful to our implementing partners, the Amader Kotha Helpline, which comprises The Cahn Group, ELEVATE, and Phulki.

B Appendix

B.1 Variables in mental health index

1. In the past 7 days, how often...

   A ...Have you felt nervous, anxious, or on edge?

   B ...Have you felt depressed?

   C ...Have you felt lonely?

   D ...Have you felt hopeful about the future?

   E ...Being so restless that it is hard to sit still

   F ...Becoming easily annoyed or irritable

   G ...Feeling afraid, as if something awful might happen

   Select one: Not at all or less than 1 day; 1-2 days; 3-4 days; 5-7 days.

2. Please imagine a ladder with steps numbered from zero at the bottom to 10 at the top. The top of the ladder represents the best possible life for you and the bottom of the ladder represents the worst possible life for you.

   A On which step of the ladder would you say you personally feel you stand at this time?

   B On which step do you think you will stand about five years from now?

   Select one: 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10.

B.2 Additional Tables and Figures

Figure B.1: Average monthly calls to the Helpline per factory by treatment arm

Notes: Average number of substantive calls to the Helpline, per factory, per month, by type of issue, for the original set of factories that entered the experiment in Phase I. We exclude 1 outlier-factory (more than 200 calls in one month in the pre-treatment period). The black vertical lines reflect the month before the launch of Phase I of the experiment (October 19, 2021). Data is prior to the launch of Phase II of the experiment.

Table B.1: Phase 1 treatment effects on number of total issues and safety issues reported to the Helpline per month

	All Calls				Safety			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Escrow	0.091	0.145*			-0.001	0.000		
	(0.083)	(0.086)			(0.007)	(0.008)		
Reputation-Escrow			0.075	0.171			-0.005	-0.003
			(0.103)	(0.108)			(0.007)	(0.008)
Retaliation-Escrow			0.108	0.120			0.003	0.003
			(0.111)	(0.116)			(0.009)	(0.011)
Observations	7337	6325	7337	6325	7337	6325	7337	6325
Factories	667	575	667	575	667	575	667	575
Baseline control mean	0.711	0.711	0.711	0.711	0.027	0.027	0.027	0.027
Coef. comparison p-val.			0.80	0.72			0.40	0.52
Strata FE	Y	Y	Y	Y	Y	Y	Y	Y
Month FE	Y	Y	Y	Y	Y	Y	Y	Y
Factory Size[1]		Y		Y		Y		Y

Notes: The table reports preliminary impacts (prior to the launch of Phase II of the experiment) of the reporting escrow, as well as separately of the reputation and the retaliation escrow, for the original set of factories that entered the experiment in Phase I. We exclude 1 outlier-factory (more than 200 calls in one month in the pre-treatment period); note that including it does not affect the nature of the results. ANCOVA regression, conditioning on post-experimental period (November 2021 - September 2022). We control for calls at t=0, i.e. the average calls in the 6 months before the launch. The dependent variables reported are the total number of substantive calls to the Helpline, and the number of safety calls to the Helpline. (Robust standard errors) clustered by factory are reported. $^{*}p < 0.10, ^{**}p < 0.05, ^{***}p < 0.01$.

Footnotes: [1] As the Helpline is preparing for the onsite training, it has been collecting updated factory size information for the sample of factories with onsite training. In this analysis, we combine our previously-available data on factory size with the updated data to improve coverage. We will continue to collect factory size as the training rolls out to the onsite training factories.

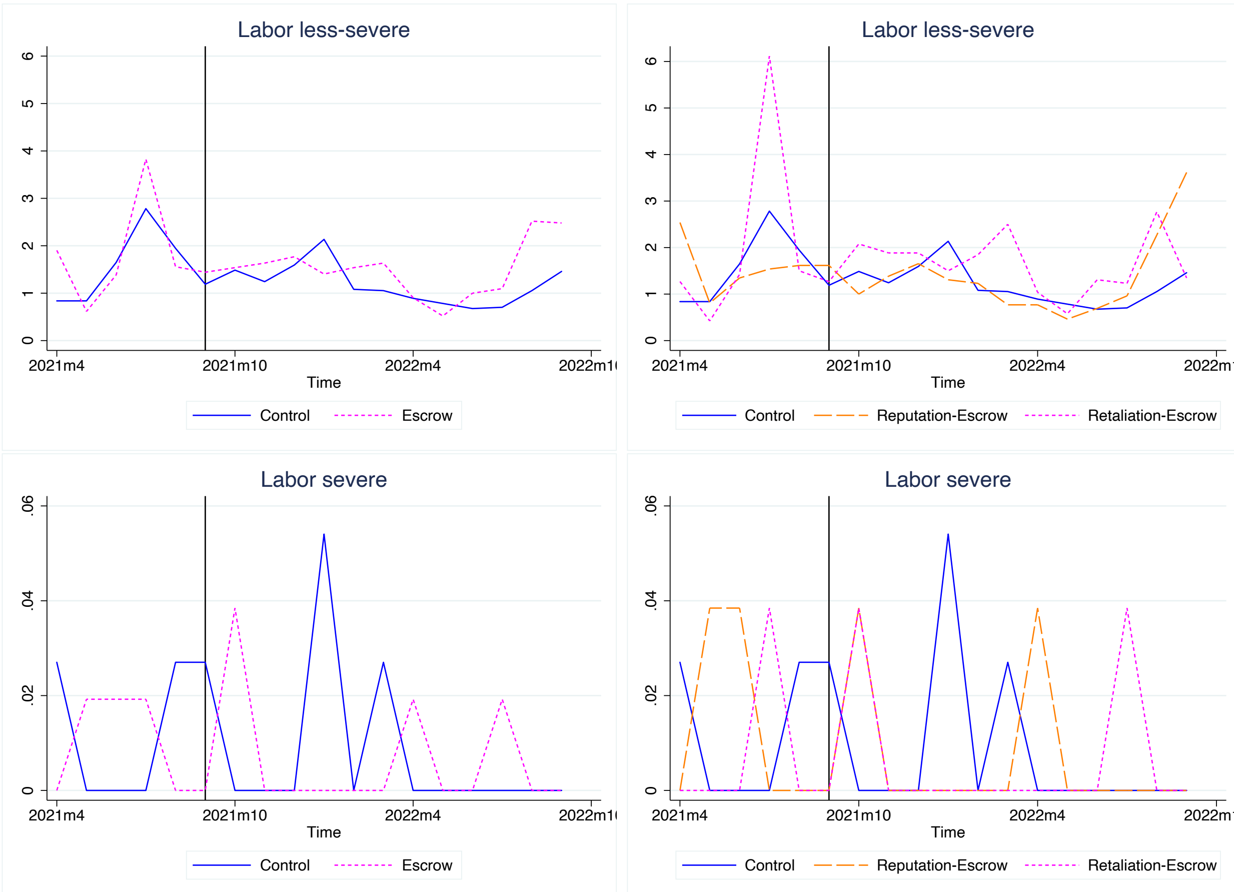
Table B.2: Balancing Table

	Mean / (SE)				Difference in means / [p-value]			
Variable	No Escrow	Escrow	Reputation	Retaliation	Escrow − No	Rep. − No	Ret. − No	Ret. − Rep.
Labor less severe	0.62	0.65	0.70	0.60	0.04	0.09	-0.01	-0.09
	(2.13)	(3.31)	(2.40)	(4.02)	[0.70]	[0.43]	[0.93]	[0.50]
Labor severe	0.01	0.01	0.01	0.01	0.00	0.00	0.00	-0.00
	(0.10)	(0.11)	(0.12)	(0.10)	[0.42]	[0.47]	[0.56]	[0.94]
Safety	0.03	0.05	0.05	0.06	0.03*	0.02	0.03	0.01
	(0.20)	(0.56)	(0.61)	(0.50)	[0.07]	[0.26]	[0.11]	[0.67]
All calls	0.71	0.76	0.82	0.71	0.06	0.11	0.01	-0.10
	(2.23)	(3.45)	(2.60)	(4.13)	[0.56]	[0.33]	[0.97]	[0.49]
Observations	1,356	2,646	1,326	1,320	4,002	2,682	2,676	2,646
Factories	226	441	221	220	667	447	446	441
Strata FE					Yes	Yes	Yes	Yes
Factory Size[1]	2075.03	2015.15	1939.34	2088.99	-71.71	-137.40	-0.64	147.64
	(1733.53)	(1674.42)	(1502.84)	(1827.06)	[0.51]	[0.24]	[1.00]	[0.22]
Data on Factory Size	0.87	0.86	0.85	0.87	-0.00	-0.02*	0.01	0.03***
	(0.34)	(0.35)	(0.36)	(0.33)	[0.77]	[0.07]	[0.28]	[0.01]
Observations	226	441	221	220	667	447	446	441
Strata FE					Yes	Yes	Yes	Yes

Notes: The table reports the mean values of each variable for each treatment arm and the differences in means between treatment arms for the set of factories that entered the experiment in Phase I. We exclude 1 outlier-factory (more than 200 calls in one month in the pre-treatment period). The top panel reports calls to the Helpline by issue type for April - September 2021, which are the 6 months preceeding the launch of Phase 1. The bottom panel reports factory size. In the top panel, columns (1)-(4) report group means and standard errors clustered by factory in parantheses, and columns (5)-(8) report the difference in means estimated using a regression with strata fixed effects and clustering standard errors by factory. In the bottom panel, columns (1)-(4) report group means and robust standard errors in parantheses, and columns (5)-(8) report the difference in means estimated using a regression with strata fixed effects and using robust standard errors. In both panels, [p-values] are reported in brackets. $^{*}p < 0.10,^{**}p < 0.05,^{***}p < 0.01$.
Footnotes: [1] As the Helpline is preparing for the onsite training, it has been collecting updated factory size information for the sample of factories with onsite training. In this analysis, we combine our previously-available data on factory size with the updated data to improve coverage. We will continue to collect factory size as the training rolls out to the onsite training factories.

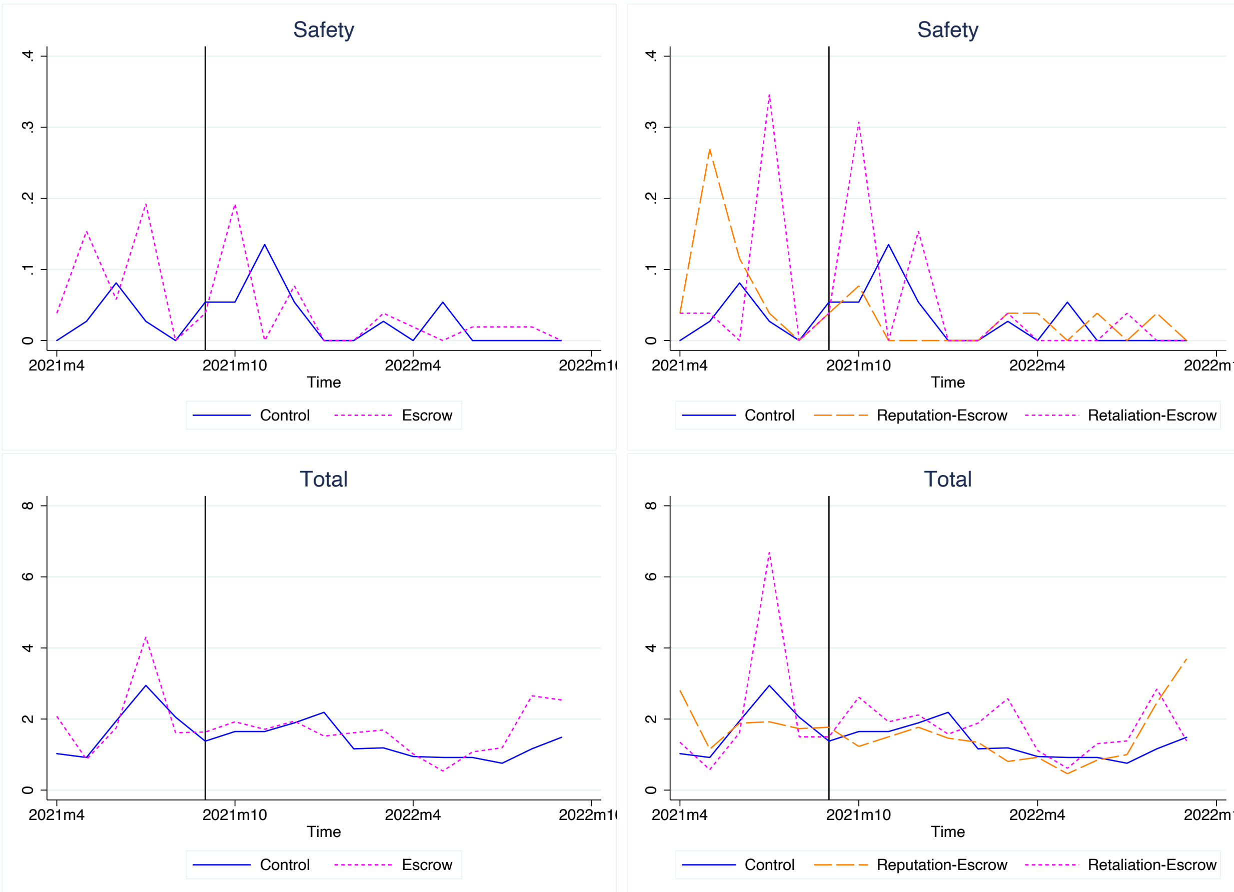
B.2.1 Additional Tables and Figures for onsite training factories

Figure B.2: Average monthly labor calls to the Helpline per factory by treatment arm for **training factories**



Notes: Average number of substantive calls to the Helpline, per factory, per month, by type of issue, for factories with onsite training. The black vertical lines reflect the month before the launch of Phase I of the experiment (October 19, 2021). Data is prior to the launch of Phase II of the experiment.

Figure B.3: Average monthly calls to the Helpline per factory by treatment arm for **training factories**



Notes: Average number of substantive calls to the Helpline, per factory, per month, by type of issue, for factories with onsite training. The black vertical lines reflect the month before the launch of Phase I of the experiment (October 19, 2021). Data is prior to the launch of Phase II of the experiment.

Table B.3: Balancing Table for **training factories**

	Mean / (SE)				Difference in means / [p-value]			
Variable	No Escrow	Escrow	Reputation	Retaliation	Escrow − No	Rep. − No	Ret. − No	Ret. − Rep.
Labor less severe	1.54	1.79	1.58	2.00	0.17	-0.06	0.54	0.54
	(3.80)	(7.39)	(3.40)	(9.90)	[0.76]	[0.91]	[0.53]	[0.56]
Labor severe	0.01	0.01	0.01	0.01	-0.00	0.00	-0.01	-0.01
	(0.12)	(0.10)	(0.11)	(0.08)	[0.80]	[0.84]	[0.49]	[0.55]
Safety	0.03	0.08	0.08	0.08	0.06	0.05	0.05	0.01
	(0.20)	(0.53)	(0.59)	(0.48)	[0.12]	[0.29]	[0.19]	[0.92]
All calls	1.71	2.04	1.88	2.21	0.26	0.09	0.57	0.44
	(3.94)	(7.50)	(3.70)	(9.95)	[0.65]	[0.88]	[0.51]	[0.65]
Observations	222	312	156	156	534	378	378	312
Factories	37	52	26	26	89	63	63	52
Strata FE					Yes	Yes	Yes	Yes
Factory Size[1]	2652.93	2291.24	2503.68	2115.74	-339.81	-13.09	-532.79	-564.19
	(1897.12)	(1501.23)	(1905.19)	(1076.60)	[0.46]	[0.99]	[0.18]	[0.35]
Data on Factory Size	0.81	0.81	0.73	0.88	0.04	-0.03	0.11	0.16*
	(0.40)	(0.40)	(0.45)	(0.33)	[0.57]	[0.64]	[0.20]	[0.06]
Strata FE					Yes	Yes	Yes	Yes
Fraction of new factories	0.08	0.06	0.08	0.04	-0.02	-0.00	-0.04	-0.04
	(0.28)	(0.24)	(0.27)	(0.20)	[0.68]	[0.95]	[0.48]	[0.56]
Observations	37	52	26	26	89	63	63	52
Strata FE					No	No	No	No

Notes: The table reports the mean values of each variable for each treatment arm and the differences in means between treatment arms for factories with onsite training. The top panel reports calls to the Helpline by issue type for April - September 2021, which are the 6 months preceeding the launch of Phase 1 of the experiment. The bottom panel reports factory size and the fraction of new factories entering the experiment, for which there is 1 observation per factory. In the top panel, columns (1)-(4) report group means and standard errors clustered by factory in parantheses, and columns (5)-(8) report the difference in means estimated using a regression with strata fixed effects and clustering standard errors by factory. In the bottom panel, columns (1)-(4) report group means and robust standard errors in parantheses, and columns (5)-(8) report the difference in means estimated using a regression with strata fixed effects and using robust standard errors. In both panels, [p-values] are reported in brackets. $^{*}p < 0.10, ^{**}p < 0.05, ^{***}p < 0.01$.
Footnotes: [1] As the Helpline is preparing for the onsite training, it has been collecting updated factory size information for the sample of factories with onsite training. In this analysis, we combine our previously-available data on factory size with the updated data to improve coverage. We will continue to collect factory size as the training rolls out to the onsite training factories.

February 6, 2023

Author statement:

Laura Boudreau: Conceptualization; Data curation; Formal analysis; Funding acquisition; Investigation; Methodology; Project administration; Resources; Software; Supervision; Validation; Visualization; Roles/Writing - original draft; Writing - review & editing.

Sylvain Chassang: Conceptualization; Data curation; Formal analysis; Funding acquisition; Investigation; Methodology; Project administration; Resources; Software; Supervision; Validation; Visualization; Roles/Writing - original draft; Writing - review & editing.

Ada González-Torres: Conceptualization; Data curation; Formal analysis; Funding acquisition; Investigation; Methodology; Project administration; Resources; Software; Supervision; Validation; Visualization; Roles/Writing - original draft; Writing - review & editing.