

Broken Promises

Evaluating an Incomplete Startup Business Program

Angelika J. Budjan, Utz Pape and Laura Ralston*

June 27, 2025

Abstract

This study examines the socio-economic, behavioral, and psychological consequences of a terminated business loan program in South Sudan on intended beneficiaries. While all participants received business training, only some were able to obtain the promised loan before the program was canceled due to renewed conflict. We combine data from face-to-face interviews and data from lab experiments to examine outcomes one year after the program's cancellation. Results from LATE estimations show that those participants who failed to receive the loan display significant declines in consumption. Moreover, this group exhibits a significant reduction in trust, particularly trust in institutions. Our results highlight that more attention should be given to the detrimental effects of implementation failures.

Keywords: business loans, startup business program, trust attitudes, impact evaluation, violent conflict, unconditional cash transfers

JEL classification: C93, D13, D81, O2A

*Corresponding author: Budjan (University of Stuttgart), email: angelika@budjan.de; Pape (World Bank), email: upape@worldbank.org; Ralston (World Bank), email: lauraralston@gmail.com.

Funding: This work was supported by the by the Korean Trust Fund [TF015917]; and the i2e Trust Fund [TF018691].

Acknowledgements: We are grateful for contributions from Mollie Foust, Luca Parisotto, Nadia Selim, Jeremy Shapiro, and James Walsh, as well as Nico a Pontara. We also thank the editor and three anonymous referees, as well as Richard Bluhm, Bledi Celiku, Axel Dreher, Arevik Gnutzmann-Mkrtchyan, Markus Goldstein, Seema Jayachandran, Lennart Kaplan, and seminar participants at Heidelberg University, the University of Göttingen, the UC Davis, the Development Economics and Policy Conference, the German Development Economics Conference, and the Tinbergen European Peace Science Conference for their valuable feedback. The design of this study underwent an ethical review by an ad hoc "review board" of experts from the World Bank.

1 Introduction

This study investigates the effects of the unplanned and unanticipated cancellation of a startup business program in South Sudan on its intended beneficiaries. The South Sudan Youth Startup Business Grant Program offered participants a combination of business and life skills training, along with a low-risk concessional loan worth approximately US\$ 1,000, whose default was fully covered by the government. While the training component was implemented as planned, the loan disbursement was prematurely terminated due to escalating conflict. This resulted in two treatment subgroups: While a subset of participants received the training and loan as initially intended (hereafter referred to as the "training and loan" group), a considerable proportion of the intended beneficiaries underwent the training but did not receive the promised loan (henceforth referred to as the "training, no loan" group). Since the intervention was originally designed as a randomized controlled trial with a control group, data were also collected on unsuccessful applicants for the program who knew that they had not been selected. Taking advantage of these data, we analyze whether the program cancellation left those who did not receive the full intervention worse off than if the program had never happened.

The identification of this effect is complicated by the fact that selection into receiving the

loan was partly endogenous. Specifically, participants needed to initiate the loan disbursement via a formal application at their local bank branch of Kenia Credit Bank (KCB), which acted as a partner in the program. This characteristic of the program was included to increase financial literacy among participants. Consequently, participants who initiated this process early had a higher likelihood of receiving the money before the program was halted.

To address endogenous selection into the "training, no loan" subgroup, we construct an instrumental variable by interacting the selection for the original treatment group with the distance to the closest KCB bank branch.¹ The distance to the closest KCB bank branch determined the transaction costs of initiating the loan disbursement, with higher proximity resulting in lower transaction costs and a positive shock to credit access. Random selection for the treatment group mediated exposure to this shock, as treatment participants experienced an increase in credit access, while control participants did not.²

Our results reveal substantial and statistically significant adverse effects for participants who received the training but failed to receive the promised loan. Specifically, we observe a reduction in consumption of between 0.4 and 0.7 standard deviations and a decline in general trust levels of approximately 0.6 – 0.9 standard deviations. Importantly, the decline in trust is driven by a reduction in institutional trust. These effect sizes are large and economically meaningful, particularly given that the endline survey was conducted more than a year after the program was halted. The findings indicate persistent negative consequences, both in terms of material welfare and in psychological terms. The results are robust to alternative estimation strategies, including a two-instrument specification that accounts for non-compliance and the use of sensitivity bounds to test robustness against attrition bias.

We contribute to the literature that highlights the particular challenges that donors face in conflict-affected regions. An increasing share of the world's poor are living in fragile states

¹Interacted instruments consisting of a common shock and an exogenous exposure to the shock have gained popularity in panel data studies, but can also be applied to spatial frameworks. Notable examples in panel data include [Nunn and Qian \(2014\)](#), [Hanna and Oliva \(2015\)](#) and [Dreher et al. \(2021\)](#).

²Our instrumental variable differs from common examples of interacted instruments in using a continuous shock and a binary exposure variable. However, the validity of the instrument follows the same logic as in cases where the shock is discrete and the exposure variable continuous.

(Kim, 2019), making it harder for development projects to achieve sustainable outcomes. Fragile states present a unique set of complexities and risks that challenge the established modes of operation of international donors (Corral et al., 2020, Campbell and Spilker, 2022). Even under well-functioning institutions, conflict events can cause program interruptions or program failures beyond the control of the donors and their partners in the recipient government. According to an estimate by Caselli et al. (2021), aid projects in fragile states are 8 percent less likely to be successful than comparable projects in non-fragile states. The risk of program failures due to conflict that donors face in fragile states poses a major obstacle to eradicating global poverty. Our findings highlight that apart from direct monetary costs for the donors, conflict-induced program failures can cause large indirect costs to the intended beneficiaries.

In addition, our study contributes to the wider literature on the effects of operational problems of development projects. A large literature analyzes the causes of operational problems besides violent conflict. Weak institutional capacity, limitations in human capital, and low quality of public infrastructure are well-documented obstacles to achieving intended outcomes (Campos et al., 2014, Brinkerhoff et al., 2018, Abate et al., 2020, Briggs, 2018). However, less attention is given to their downstream effects. An important exception is Ghosh and Kochar (2018), who show how benefits can appear despite implementation failures. In contrast, our study suggests that operational problems in public service delivery could produce self-reinforcing dynamics by eroding institutional trust, which in turn could undermine the uptake of future inventions and compromise the long-term effectiveness of development programs.

Finally, we contribute to the emerging field of implementation science, which originated in public health as a primarily qualitative approach for understanding not only whether interventions are cost-effective under ideal conditions but also whether they are effective in real-world settings (Bauer et al., 2015). Economists are increasingly concerned with similar questions in the context of randomized controlled trials, shifting attention from simply identifying *what* works to understanding *how* and *why* certain interventions succeed or fail in practice (Duflo, 2017). For instance, Schaffner et al. (2024) use a mixed-methods approach to explain why a teacher training program in Nepal had little effect on students' outcomes. In addition, Angrist and Meager (2023) provides a compelling example of how implementation science can be

approached rigorously and quantitatively. We contribute to this literature by demonstrating that implementation failures do not merely attenuate program effects toward zero but can actively generate *negative* impacts, undermining both trust and economic outcomes among intended beneficiaries. This highlights the importance of anticipating and mitigating unintended consequences in the design and ethical evaluation of public programs.

2 Context

The Youth Startup Business Grant Program was developed by the World Bank and the Government of South Sudan and aimed to address high youth unemployment by increasing self-employment opportunities with low implementation complexity. In addition to individual benefits, it sought to generate local economic spillovers and strengthen the private sector.

The program consisted of a one-time grant of US\$ 1,000 paired with a one-week business and life skills training. Participants accessed funds by opening a bank account and submitting a loan request at their local bank branch of Kenia Credit Bank (KCB), which acted as a partner in the program. Although successful repayment enabled eligibility for a second US\$ 1,000 loan, defaults were covered by a government guarantee, effectively making the funds a grant.

Historically, the region that is today South Sudan had been marred by Africa's longest-running civil war. While most of the violence ceased with independence from the Republic of Sudan in the North in 2011, much of the fighting had consisted of clashes between the 63 distinct ethnic and language groups within the territory of South Sudan. This legacy of violence led to renewed conflict in December 2013. In the years 2013 to 2015, violence was concentrated in the North of the country. Thus, the business loan program selected six Southern states for its implementation: Eastern Equatoria, Central Equatoria, Western Equatoria, Lakes State, Northern Bahr el Ghazal, and Western Bahr el Ghazal.

Eligible applicants were South Sudanese youth aged 18–35 in the six program states, with 60% of slots reserved for young women. Applicants had to submit a one-page business idea, but funding was not conditional on implementing the proposal, allowing for flexibility in use. A broad outreach campaign between July and November 2014 yielded 8,240 applications, of

which 4,699 were eligible. Using an oversubscription design stratified by gender and state, 1,200 applicants were randomly selected into the treatment group and 1,200 into the control group. Training sessions were held in March and April 2015. Afterward, participants could open their bank accounts and could initiate the loan disbursement.

However, after program initiation, the conflict moved further south, and particularly the states Lakes and Central Equatoria, as well as the urban centers in Western Bahr el Ghazal and Eastern Equatoria became major conflict sites. With the conflict's resurgence in early 2016, loan disbursement was halted. Due to inconsistent responses across branches and poor documentation, it remains unclear when disbursement ceased in each location.

3 Study Design

In the following, we outline the study design. Figure 1 illustrates the timeline of the program implementation, cancellation, and data collection, as well as the original treatment groups and the treatment subgroups caused by the program cancellation.

Baseline randomization. The baseline survey was conducted between January to March 2015, and data were collected from 1,144 treatment participants and 1,148 control participants. Approximately 4.5 percent of initially selected study participants could not be tracked and did not participate in either the baseline survey or the program. The baseline survey was concluded before beneficiaries were informed whether they were selected for the loan and before the one-week training (see Figure 1). The original treatment and control group were balanced across a wide set of covariates (see Appendix Table A.1 for covariate definitions and Appendix Table A.2 for the balance test).

Phone survey and endline. After the program's cancellation, we conducted a phone survey in May 2017 to assess the feasibility of an endline survey and inform participants about the program halt. We reached 55% of baseline participants (642 control, 622 treatment), with 99% agreeing to a follow-up.

Encouraged by this response, we designed a face-to-face endline survey. We expanded

the baseline questionnaire to include new measures on psychological wellbeing, trust, risk preferences, crime, and migration. Trust and risk preferences were captured using both self-reported variables and incentivized tasks (see Appendix B for full methodological details).

Due to the difficulty of conducting fieldwork during an ongoing conflict, we targeted a subsample of 1,800 participants (78% of baseline), prioritizing those reached by phone and filling the remaining endline sample with randomly selected participants from the baseline. Intensive tracking efforts from September to December 2017 resulted in 1,264 completed interviews ($\approx 82\%$).³

Attrition. Given the limited target sample size at endline, attrition rates compared to the baseline sample are relatively high but balanced across the treatment and the control group (33.54% in the control group and 33.57% in the treatment group). Moreover, attritor characteristics did not differ significantly between the treatment and control group (Appendix Table A.3). However, we do find systematic differences between attritors and endline participants (Appendix Table A.4). To assess the robustness of our results to this systematic attrition, we derive Kling and Liebman sensitivity bounds (Kling et al., 2007, ; see Section 7).

Data handling and multiple hypothesis testing. Before aggregation, all non-negative continuous variables were winsorized at the 99th percentile. All monetary variables were adjusted to constant prices using monthly deflators. Indicators with limited variation (defined as 95% of responses showing the same value) were excluded, resulting in the omission of six items.⁴

To mitigate bias from multiple hypothesis testing, we employ two strategies. First, we group related outcomes into standardized indices to reduce dimensionality and increase statistical power following Anderson (2008). Each index is standardized using the control group (mean

³To increase sample size, enumerators also surveyed 179 participants from the replacement pool of participants not originally selected for the endline. In the main analysis, we exclude female replacement observations due to some evidence of convenience sampling, though results are robust to their inclusion (Appendix Table A.11).

⁴Excluded indicators include engagement in cattle raids, frequency of cattle raids, number of times beaten in the past month, in-kind wage payments, and remaining loan amounts (formal and informal).

= 0, SD = 1) to facilitate interpretation. Second, we control the false discovery rate using the two-step procedure by [Benjamini and Hochberg \(1995\)](#), applied separately to two outcome families: socio-economic and behavioral/psychological.

Minimum detectable effects. To assess the risk of a type II error, we compute the *ex post* minimum detectable effect (MDE) size of the LATE estimates (Appendix Table A.7).

4 Estimation strategy

4.1 Selection into “training, no loan”

Since the cancellation of the program was not planned, the selection process into "training, no loan" was not systematically controlled. As a result, individual characteristics may have influenced whether a participant received a loan. To assess the degree of endogenous selection, we test the balance on covariates between "training, no loan" and "training and loan" (Table 1). The balance test reveals significant differences between groups (joint orthogonality p -value < 0.01). Specifically, participants who did not receive the loan were less likely to be married, had higher education, higher numeracy skills, and lived in smaller households with a lower number of children. Importantly, we find no difference in conflict exposure before or after the program onset.

A major determinant of the treatment subgroup was the distance to the closest KCB branch. This distance was measured using the coordinates of the baseline interview, which typically took place at participants' home addresses, and the coordinates of the KCB bank branches. [Figure 2](#) displays a map of participants' baseline locations and the locations of KCB bank branches. Participants who received the loan lived, on average, 1.7 kilometers closer to a KCB bank branch. This suggests that transaction costs to access the loan increased in bank distance and affected the selection into the "training, no loan" subgroup. We exploit this variation in our instrumental variable strategy.

4.2 Identification strategy

To estimate the effects of "training, no loan," we instrument for selection into this treatment subgroup with the interaction between the treatment group dummy and the distance to the closest KCB bank branch. We exploit the fact that receiving the loan was conditional upon holding a formal bank account at KCB Bank. During the study period, KCB operated only 15 bank branches and not in every large city. However, KCB was only one of at least 8 commercial banks active in South Sudan since independence ([Bank of South Sudan, 2010](#)). This led to some variation in the transaction costs of accessing the loan. This difference in transaction costs meant that participants living close to a KCB bank branch experienced a large positive shock to their access to credit. In turn, participants who lived far from a KCB bank branch experienced only a limited improvement in their access to credit due to high transaction costs, making them more likely to end up in the "training, no loan" subgroup.

Algebraically, our estimation strategy reads as follows.

Second stage equation:

$$y_i = \alpha + \beta_1 D_i^T + \beta_2 D_i^{T-L} + \lambda KCBDist_i + X_i' \gamma + s_i + \epsilon_i, \quad (1)$$

First stage equations:

$$D_i^{T-L} = \tau + \delta D_i^T + \sigma D_i^T \times KCBDist_i + \mu KCBDist_i + X_i' \eta + s_i + \epsilon_i, \quad (2)$$

where y_i is an outcome for individual i and α is a constant. D_i^T is a dummy variable for *all* individuals that were originally selected for the treatment, and its corresponding coefficient tells us how participants of this group compare to the control group. Since selection into the treatment group was randomized, this coefficient has a causal interpretation. D_i^{T-L} is a binary indicator for the treatment subgroup that received only "training, no loan." As outlined above, selection into this subgroup was partly endogenous. Therefore, we instrument for this endogenous regressor in the first-stage displayed in Equation (2) with the interaction term $D_i^T \times KCBDist_i$.

All regressions control for $KCBDist_i$ as the logarithmic distance to the closest KCB

bank branch. This addresses the risk from omitted variables that correlate with KCB bank distance, effectively conditioning on any observable outcome differences among control group participants across locations. X_i is a set of control variables. In particular, we include a set of geographic controls to address the potential correlation of KCB distance with remoteness. These include distance to the closest city center, distance to the closest road, average land gradient, conflict exposure during the program, as well as the interactions of all geographic covariates with the treatment group dummy. In addition, some specifications control for baseline covariates. s_i are strata (state-gender) fixed effects and ϵ_i is the error term clustered at the boma level to address spatial correlation.

To understand how participants of "training, no loan" compared to the control group — *i.e.* how they compare to a counterfactual scenario where the program had never taken place — we need to consider the additive effect of being randomly selected into the treatment *and* ending up in the subgroup "training, no loan." This effect is given by the sum of parameters β_1 and β_2 . Given these considerations parameter β_1 can be interpreted as the effect of being randomly selected into the program for all participants that did not end up in the subgroup "training, no loan," *i.e.* the effect of the originally planned program.

Relevance of the instrument. In Table 2, we report first-stage results of our estimation strategy. Column 1 reports results without any control variables, column 2 with our set of geographic controls, including state-gender fixed effects, and column 3 with geographic controls and additional baseline covariates. Conceptually, we prefer the last two specifications, which condition on geographic covariates to control for remoteness effects. In all specifications, our interacted instrument increases the likelihood of ending up in the treatment subgroup "training, no loan" (significant at 1 percent). In contrast, the non-interacted distance to any KCB bank branch is statistically insignificant.

We report the effective F -statistic of the first stage by [Montiel and Pflueger \(2013\)](#) to judge the power of the instrument, following the recommendation by [Andrews et al. \(2019\)](#). The tests show that conditioning on the set of geographic controls is not only conceptually preferable but also improves the statistical power of the instrument.⁵ Therefore, we focus our main analysis

⁵While the first specification is weak with an effective F -statistic < 10 , the last two

on specifications with geographic controls.

In the case of IV estimations, the usual t-test confidence intervals can be biased if the degree of endogeneity is large and the sample is finite (Lee et al., 2022, Angrist and Kolesár, 2024). We address this in two ways: First, we derive ρ test statistics following Angrist and Kolesár (2024) for every main regression showing that the degree of endogeneity ρ is low and t-ratio inference therefore reliable (Appendix Table A.4). Second, we derive tF -confidence intervals using worst-case assumptions about the degree of endogeneity following Lee et al. (2022) and show that the main results remain robust even under these extreme assumptions (Appendix Table A.7).

Validity of the instrument. The identifying assumption of an interacted instrument follows a similar logic as a parallel trends assumption in a difference-in-differences design (Christian and Barrett, 2024, Goldsmith-Pinkham et al., 2020). In our study, the identifying assumption requires that in the absence of the program, any (randomly occurring) outcome difference between treatment and control group participants was uncorrelated with their distance to the closest KCB bank branch. We test this assumption by running baseline characteristics and selection into the endline sample on the reduced form of our main estimation as given by:

$$y_i = \tau + \delta D_i^T + \sigma D_i^T \times KCBDist_i + \mu KCBDist_i + X_i'\eta + s_i + \epsilon_i, \quad (3)$$

where τ is a constant, D_i^T is the treatment group dummy, $D_i^T \times KCBDist_i$ is our interacted instrument, $KCBDist_i$ is the logarithmic bank distance, X_i are our control variables, s_i are strata (state-gender) fixed effects, and ϵ_i is the error term.

Table 3 reports the results on the coefficients for the treatment group dummy in column (1), (log) bank distance in column (2), and, importantly, the coefficient on our interacted instrument in column (3). With most of our baseline covariates, the interacted instrument is not significantly correlated. Moreover, the instrument does not predict selection into the endline sample, as shown by a small and insignificant coefficient on attrition status.⁶ A joint test of specifications are not weak with an effective F -statistic of 25.59 or 30.83, respectively.

⁶Note also that non-interacted (log) bank distance (column 2) is also not significantly correlated with attrition, implying even in the control group, distance to the closest bank branch

orthogonality of all baseline covariates and the attrition dummy is statistically insignificant with a p -value > 0.1 . However, the instrument shows some correlation with education. While this might have occurred by chance, we address any concern from this correlation by including regressions that control for education level, literacy skills, numeracy skills, and their interaction with the treatment group dummy as additional control variables in our main estimations.

5 Main Results

Table 4 reports results for LATE estimations from specifications with geographic control variables and with a combination of geographic control and baseline covariates. For every specification, we report the coefficient on D_i^T , D_i^{T-L} , and on their sum. As outlined above, the latter provides us with an estimate for the effect of "training, no loan" compared to a counterfactual of never participating in the program.

Socio-economic outcomes. We find a negative and statistically significant effect on the consumption index for participants in the "training, no loan" subgroup. Relative to the control group, these individuals experienced a reduction in consumption of between 0.4 and 0.7 standard deviations. This effect is particularly strong in the specification with full controls, where it remains significant with an adjusted p -value of 0.055. This effect size is remarkable given that the endline was collected more than one year after the program was canceled. The effect cannot be explained simply by short-term overspending, but suggests a more lasting mechanism. This result provides clear evidence that participants were worse off than they would have been had the program never been implemented.

By contrast, we find no effect on business skills. The absence of an impact on business skills is somewhat unexpected, as even participants in the subgroup "training, no loan" did complete a one-week business training. The coefficient on the treatment group dummy — which estimates the effect of the remaining treatment group participants — is also small and statistically insignificant. Given that most participants of this group also received the training, this result suggests that the training was largely ineffective.

does not predict selection into the endline sample.

Moreover, we find no significant effects of assignment to the "training, no loan" subgroup on employment or savings. The point estimates for both outcomes are close to zero, suggesting that these null results are not simply due to limited statistical power.⁷ However, particularly the absence of any negative impact on the savings index makes the substantial reduction in consumption even more striking. The aggregate savings index might conceal underlying nuances in different aspects of savings. To explore this further, in Section 6, we disaggregate the consumption index and analyze its components individually, complemented by a similar breakdown of the savings index.

Psychological and behavioral outcomes. We find a substantial and statistically significant reduction in trust among participants who failed to receive their loans. Specifically, trust index scores for this group are approximately 0.6 to 0.9 standard deviations lower than those of the control group, with adjusted p -values below 0.05 in both specifications. This effect is large, especially given that the endline survey was conducted a full year after the program's cancellation, suggesting that the program raised strong expectations. The observed drop in trust is consistent with rational belief updating in response to a salient negative experience. However, because the aggregate trust index includes both institutional and interpersonal dimensions, this result does not reveal whether the loss in trust is mainly driven by a loss of trust in the institutions involved with the program or whether it had spillover effects on other dimensions of trust. We therefore examine the subcomponents of the trust index separately in Section 6.

We also observe a weakly significant decline of 0.4 standard deviations in the psychological wellbeing index for the "training, no loan" group when controlling only for geographic characteristics. Although the coefficient remains similar in magnitude after adding baseline covariates, the effect loses statistical significance, making it less conclusive if general psychological wellbeing decreased. For all other psychological and behavioral outcomes, we find no significant effects for the "training, no loan" subgroup. However, given that the analysis was powered only to detect large effect sizes (Appendix Table A.7), we cannot rule out the possibility of medium or small effects on the remaining indicators.

⁷Note, however, that the LATE estimates are powered to detect only medium to large effects (see Appendix Table A.7).

Heterogeneity by gender. Both the reduction in consumption and the reduction in trust are consistent across genders, with no statistically significant gender differences in the effects of "training, no loan." (see Appendix Tables A.8 and A.9).

6 Analysis of index components

To better understand the mechanisms underlying the negative effects on consumption and trust, we disaggregate our LATE estimates by examining the individual components of each index. To ease interpretation, we use the raw data here without relying on the standardization of the variables.

Consumption and savings. Panel A of Table 5 presents the results for components of the consumption index. The decline in consumption among participants who failed to receive the loan is primarily driven by reductions in both food consumption and asset acquisition. In particular, participants of the "training, no loan" subgroup consumed on average 1.3 to 1.5 fewer food items than the control group. While only significant in the specification with geographic controls, we also observe a decrease in food expenditures of approximately 20 – 35 percent. Moreover, in both specifications, we observe a weakly significant reduction in asset acquisition. The coefficients on non-food daily consumption are negative but not statistically significant. Taken together, these findings suggest that the program's cancellation led to reduced consumption across several categories, with the most pronounced effect on food consumption.

As a complementary analysis, we examine the components of the savings index in Panel B of Table 5. We observe a highly significant reduction in formal lending of approximately 23 percent in the specification with only geographic control variables. In the specification with full controls, the estimate is even larger, though less precisely estimated (p -value = 0.17). Crucially, this reduction is not offset by an increased uptake of informal loans. These findings provide suggestive evidence that the failure to receive the loan could have lowered participants' lending behavior and thereby contributed to the observed decline in consumption.

Trust. In Table 6, we report LATE estimates for all trust index components. The trust index consists of sixteen components, out of which 13 are self-reported measures of trust, and three variables are the results of a trust game. Results on the self-reported variables show that the decline in the aggregate trust index is driven by a reduction in trust in institutions. Participants who failed to access the loan show a significant reduction in trust in the local government, trust in NGOs, trust in elders, and a weakly significant reduction in trust in the police. The reduction of trust in NGOs and the local government suggests that participants had difficulty understanding which institution was actually in charge of the program, *i.e.* the central government rather than the local government or an NGO. We also see negative medium-sized to large-sized, but statistically insignificant, coefficients on components including trust in family, trust in friends, and trust in neighbours.

The experimental measures of trust were collected as follows: In the first trust game, all participants played as player A, while we framed player B as the World Bank, giving us an experimental measure of trust in the World Bank. In the second trust game, the sample was split so that half of the participants played as player A and the other half played as player B. Therefore, we elicited the experimental measure of trust in another player only for half the sample and an experimental measure of trustworthiness for the other half (details of the experiment are described in Appendix B). Although the estimated effect of the “training, no loan” subgroup on trust in the World Bank is not statistically significant, the coefficient is negative and substantial in magnitude. Participants in this subgroup sent 13 to 16 South Sudanese Pounds (SSD) less to the “World Bank” out of a 100 SSD endowment⁸, corresponding to a reduction of 0.5 to 0.6 standard deviations. In the second trust game, the estimated treatment effect is smaller (around 2 SSD) in the specification with geographic controls. When full controls are included, the effect increases to a magnitude similar to that in the first game and becomes weakly statistically significant. Notably, this effect appears to be driven by the overall treatment group rather than specifically the “training, no loan” subgroup, as shown in Column (5). Thus, it should not be interpreted as a reduction in trust driven by the program cancellation, but could be related to a priming effect given the context of the loan program.

⁸Approximately 0.83 USD

Taken together, the results suggest that the program's cancellation led to an erosion of institutional trust, with spillover effects on uninvolved actors such as local elders and the police. The erosion of trust has potential long-term consequences. If credibility in development programs deteriorates, future initiatives may face lower uptake, especially in fragile contexts where trust in the government is already low.

7 Robustness

Two-IV-estimation. A caveat for our main estimation strategy is the presence of non-compliers. Among all the participants randomly selected for the treatment group, there is a subgroup of non-compliers who never went to the training and had no account opened at KCB. Since this group did not comply with the conditions of the program and had no expectation to receive a loan, we do not count these participants into the "training, no loan" subgroup. Their presence in the data implies that they might suppress the effect of the treatment group dummy D^T , given that non-compliers did not receive any of the benefits of the program. Thus, the coefficient on the treatment group dummy D^T should be interpreted similarly to an intention-to-treat effect of receiving the full program. However, their presence should not substantially affect our estimated effect for "training, no loan." To test this directly, we run an alternative specification featuring two endogenous regressors for the two subgroups: "training, no loan" and "training and loan." Non-compliers are now coded as not being part of either of these groups. We run the following LATE estimation:

Second stage equation:

$$y_i = \alpha + \beta_1 D_i^{T+L} + \beta_2 D_i^{T-L} + \lambda KCBDist_i + X_i' \gamma + s_i + \epsilon_i, \quad (4)$$

First stage equations:

$$D_i^{T+L} = \delta + \sigma_1 D_i^T + \sigma_2 D_i^T \times KCBDist_i + \eta KCBDist_i + X_i' \theta + s_i + \epsilon_i, \quad (5)$$

$$D_i^{T-L} = \iota + \kappa_1 D_i^T + \kappa_2 D_i^T \times KCBDist_i + \phi KCBDist_i + X_i' \tau + s_i + \epsilon_i, \quad (6)$$

where y_i is an outcome for individual i , α is a constant, D_i^{T+L} and D_i^{T-L} are dummy variables indicating treatment subgroup “training and loan” and “training, no loan”, respectively, X_i are control variables, s_i are strata fixed effects, and ϵ_i is the error-term clustered at the boma level. Equations (5) and (6) display the first-stage equations, which instrument D_i^{T+L} and D_i^{T-L} with the original assignment to treatment D_i^T as well as the interaction between D_i^T and the logarithmic distance to the closest KCB branch $KCBDist_i$. The LATE of D_i^{T+L} and D_i^{T-L} is estimated by parameters β_1 and β_2 respectively.

Table 7 reports the results of this estimation. Results in columns 2 and 4 report the effect of "training, no loan" relative to the control group, equivalent to columns 3 and 6 of Table 4. Encouragingly, the estimates for the negative effects on consumption and trust remain qualitatively unaffected by using the two-IV specification. While the effect sizes are slightly smaller, both effects remain statistically significant and of economically meaningful magnitude across specifications.

Results on "training and loan" are reported in columns 1 and 3. Comparing these results to the effects of the treatment group dummy in our main estimation shows again relatively similar results. This provides further evidence that our handling of non-compliers in the main estimation strategy does not introduce substantial bias.

Sensitivity bounds. Next, we address the issue of our high attrition rates by deriving sensitivity bounds. A bounding approach increasingly used in economics when dealing with high attrition rates is the method proposed Kling et al. (2007). It is useful when there is a risk that potential outcomes might have differed between attritors of the control and the treatment group.⁹ This approach bounds treatment effects by making reasonable assumptions about unobserved outcomes of attritors based on the distribution of outcomes among observed participants. Since we have no knowledge of whether attritors from the treatment group were selected into the "training, no loan" subgroup, the calculation of the sensitivity bounds also requires an assumption about their treatment subgroup. The most extreme assumption we can make is that all treatment group attritors failed to receive the loan. Assigning all treatment group attritors to "training, no loan," we derive sensitivity bounds for the LATE estimations, assuming that

⁹Notable applications include, for instance, Blattman et al. (2020) and Özler et al. (2021).

attritors performed 0.1 or 0.25 standard deviations better (or worse) than the rest of their group. This gives us a better notion of the effect size for the original program target population.

We rely on the specification with geographic controls as our more conservative estimate. Results of the bounding exercise are reported in Table 8. The negative effect of "training, no loan" on consumption is robust when assuming that attritors differed by 0.1 standard deviations. When relying on the more extreme assumption that attritors differed by 0.25 standard deviations, the lower bound of the effect has a p-value of 0.103, but remains of a relevant effect size (0.34 standard deviations). The negative effect of "training, no loan" on trust is robust across specifications. Participants of this subgroup show, on average, at least a 0.36 standard deviation lower trust level than the control group. Taken together, these results suggest that our findings are not driven by selective attrition. Additional robustness tests are presented in Online Appendix A.

8 Discussion and conclusion

Our study analyses the conflict-induced unplanned cancellation of the South Sudan Youth Business Start-Up Program on intended beneficiaries. Overall, our results suggest that participants who were selected for the program but received only the training and not the loan experienced a welfare loss compared to participants who were not selected for the program. In particular, participants who failed to receive their loans show a large and statistically significant reduction in their consumption and their trust level one year after the program cancellation.

The decline in consumption spans both daily expenditures and asset acquisition, with suggestive evidence pointing to a behavioral shift related to lending practices. Importantly, the observed erosion of trust is concentrated in institutional trust. While part of this effect pertains to trust in the World Bank and the government, it also extends to unrelated institutions such as NGOs, the police, and local elders. This pattern indicates that program failures can have broader reputational spillovers, diminishing trust in a wider range of actors, even those not directly accountable for the program's implementation. Such a decline in institutional trust carries long-term risks for development effectiveness, as it may reduce participation in future

initiatives and hinder program uptake.

Our study is innovative in using random assignment to a treatment group as an exogenous exposure variable in an interacted instrument. Since RCT randomization is, in most cases, a credible source of exogenous variation, this approach has the potential for numerous applications outside our study. In particular, future studies that investigate the effect of incomplete or interrupted program implementations could employ this approach to expand our knowledge on this under-researched topic.

More broadly, our findings underscore the importance of considering the full range of consequences that may arise from interrupted or incomplete program delivery in fragile and conflict-affected settings. While such failures are often framed as logistical setbacks in efforts to improve the lives of beneficiaries, our results suggest they can cause meaningful economic and psychological harm. In our case, the disruption was driven not by implementation error but by renewed conflict, highlighting the limited control that international donors often have over program delivery in fragile environments. Given that interruptions and cancellations are not always avoidable in conflict-prone regions, we argue that development programs must be designed with contingency planning in mind, accounting not only for delivery risks but also for the potential harm caused by partial implementation. Our study provides rare empirical evidence of these risks and calls for greater integration of such considerations into both the design and ethical evaluation of development interventions in fragile contexts.

Data availability

The data used in this article are available online in the Harvard Dataverse and can be accessed at <https://doi.org/10.7910/DVN/ACMSUV>.

References

- Abate, G. T., Dereje, M., Hirvonen, K. and Minten, B. (2020), ‘Geography of public service delivery in rural Ethiopia’, *World Development* **136**, 105–133.
- Anderson, M. L. (2008), ‘Multiple inference and gender differences in the effects of early intervention: A reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects’, *Journal of the American Statistical Association* **103**(484), 1481–1495.
- Andrews, I., Stock, J. H. and Sun, L. (2019), ‘Weak instruments in instrumental variables regression: Theory and practice’, *Annual Review of Economics* **11**, 727–753.
- Angrist, J. and Kolesár, M. (2024), ‘One instrument to rule them all: The bias and coverage of just-ID IV’, *Journal of Econometrics* **240**(2), 105398.
- Angrist, N. and Meager, R. (2023), ‘Implementation matters: Generalizing treatment effects in education’, *mimeo* .
- Bank of South Sudan (2010), ‘List of banks authorized to operate in South Sudan’. Archived 2010-09-29 at the Wayback Machine.
- Bauer, M. S., Damschroder, L., Hagedorn, H., Smith, J. and Kilbourne, A. M. (2015), ‘An introduction to implementation science for the non-specialist’, *BMC psychology* **3**, 1–12.
- Benjamini, Y. and Hochberg, Y. (1995), ‘Controlling the false discovery rate: A practical and powerful approach to multiple testing’, *Journal of the Royal Statistical Society* **57**(1), 289–300.
- Blattman, C., Fiala, N. and Martinez, S. (2020), ‘The long-term impacts of grants on poverty: Nine-year evidence from Uganda’s youth opportunities program’, *American Economic Review: Insights* **2**(3), 287–304.
- Briggs, R. C. (2018), ‘Poor targeting: A gridded spatial analysis of the degree to which aid

- reaches the poor in Africa', *World Development* **103**, 133–148.
- Brinkerhoff, D. W., Wetterberg, A. and Wibbels, E. (2018), 'Distance, services, and citizen perceptions of the state in rural Africa', *Governance* **31**(1), 103–124.
- Campbell, S. P. and Spilker, G. (2022), 'Aiding war or peace? The insiders' view on aid to postconflict transitions', *The Journal of Politics* **84**(3), 1370–1383.
- Campos, F., Coville, A., Fernandes, A. M., Goldstein, M. and McKenzie, D. (2014), 'Learning from the experiments that never happened: Lessons from trying to conduct randomized evaluations of matching grant programs in Africa', *Journal of the Japanese and International Economies* **33**, 4–24.
- Caselli, F. G., Presbitero, A. F., Chami, R., Espinoza, R. and Montiel, P. J. (2021), Aid effectiveness in fragile states, in 'Macroeconomic Policy in Fragile States', Oxford University Press, Oxford, pp. 493–520.
- Christian, P. and Barrett, C. B. (2024), 'Spurious regressions and panel iv estimation: revisiting the causes of conflict', *The Economic Journal* **134**(659), 1069–1099.
- Corral, P., Irwin, A., Krishnan, N. and Mahler, D. G. (2020), *Fragility and conflict: On the front lines of the fight against poverty*, World Bank Publications.
- Dreher, A., Fuchs, A., Parks, B., Strange, A. and Tierney, M. J. (2021), 'Aid, China, and growth: Evidence from a new global development finance dataset', *American Economic Journal: Economic Policy* **13**(2), 135–74.
- Duflo, E. (2017), 'The economist as plumber', *American Economic Review* **107**(5), 1–26.
- Ghosh, P. and Kochar, A. (2018), 'Do welfare programs work in weak states? Why? Evidence from a maternity support program in India', *Journal of Development Economics* **134**, 191–208.
- Goldsmith-Pinkham, P., Sorkin, I. and Swift, H. (2020), 'Bartik instruments: What, when, why, and how', *American Economic Review* **110**(8), 2586–2624.
- Hanna, R. and Oliva, P. (2015), 'The effect of pollution on labor supply: Evidence from a natural experiment in Mexico City', *Journal of Public Economics* **122**, 68–79.
- Kim, J. Y. (2019), Fixing fragility: A new approach to state fragility, in B. S. Coulibaly, ed., 'Foresight Africa: Top priorities for the continent in 2019', Africa Growth Initiative at

Brookings, pp. 59–75.

Kling, J. R., Liebman, J. B. and Katz, L. F. (2007), ‘Experimental analysis of neighborhood effects’, *Econometrica* **75**(1), 83–119.

Lee, D. S., McCrary, J., Moreira, M. J. and Porter, J. (2022), ‘Valid t-ratio inference for IV’, *American Economic Review* **112**(10), 3260–3290.

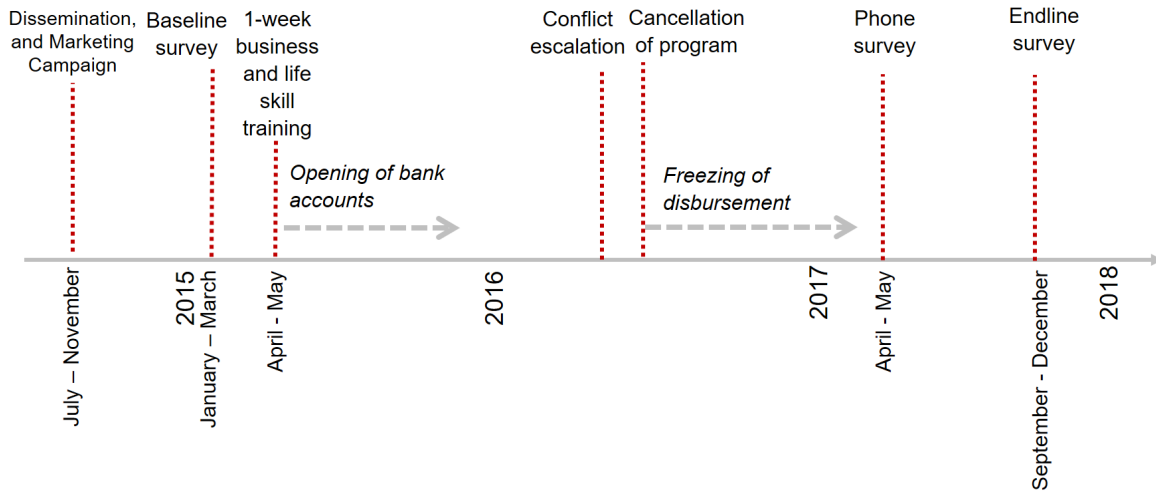
Montiel, O. J. L. and Pflueger, C. (2013), ‘A robust test for weak instruments’, *Journal of Business & Economic Statistics* **31**(3), 358–369.

Nunn, B. N. and Qian, N. (2014), ‘US food aid and civil conflict’, *American Economic Review* **104**(6), 1630–1666.

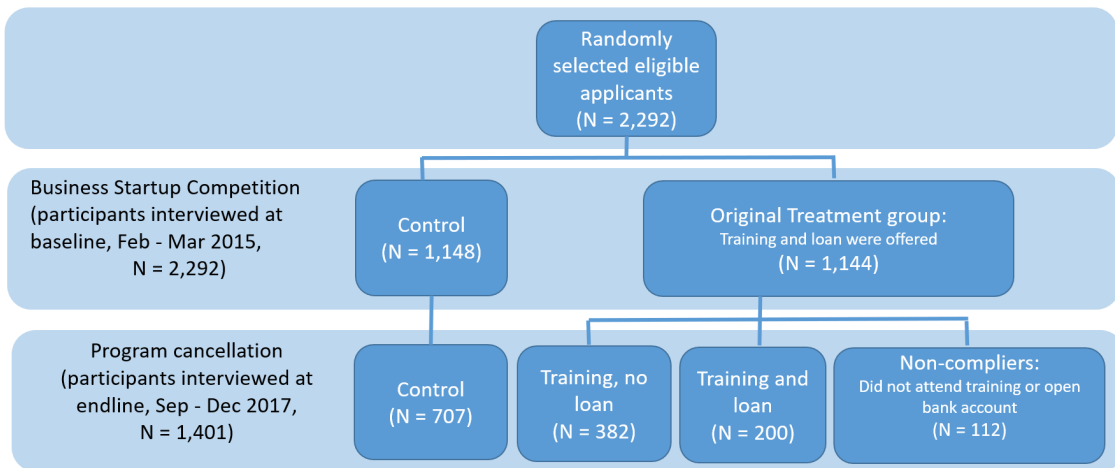
Özler, B., Çelik, Ç., Cunningham, S., Cuevas, P. F. and Parisotto, L. (2021), ‘Children on the move: Progressive redistribution of humanitarian cash transfers among refugees’, *Journal of Development Economics* **153**, 102733.

Schaffner, J., Glewwe, P. and Sharma, U. (2024), ‘Why programs fail: Lessons for improving public service quality from a mixed-methods evaluation of an unsuccessful teacher training program in nepal’, *The World Bank Economic Review* p. 473–496.

Figures



(a) Timeline of program implementation, cancellation, and data collection



(b) Treatment streams of original randomization and *ex post* treatment subgroup

Figure 1 – Program implementation timeline and treatment status overview.

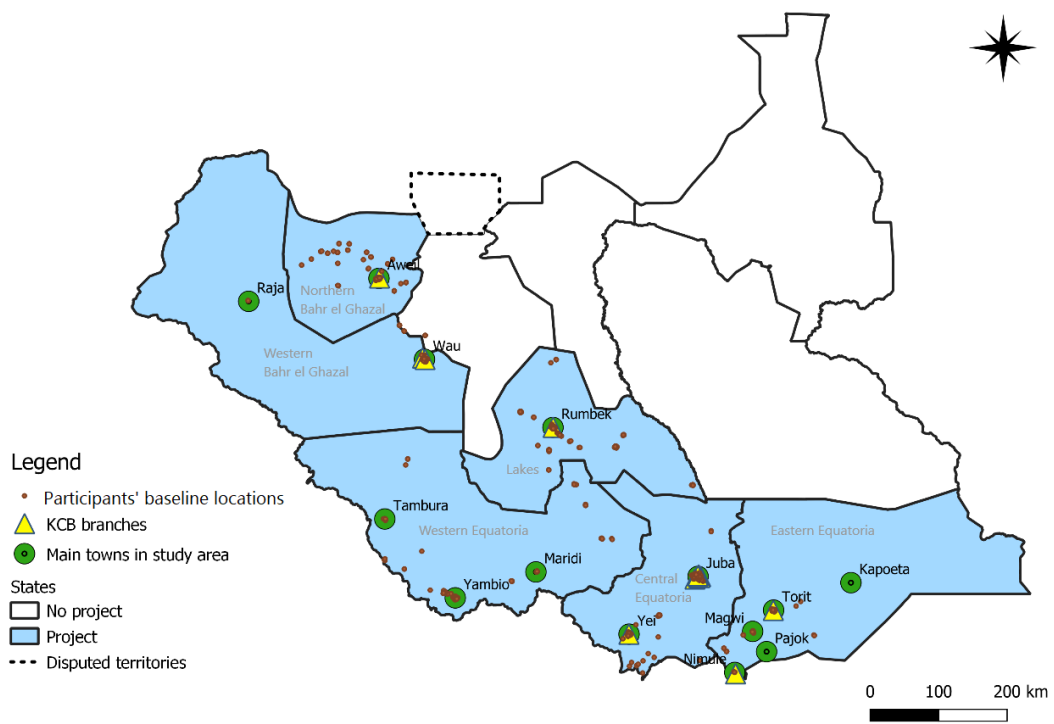


Figure 2 – Map of selected program states, participants’ baseline locations, and bank branch locations

Tables

Table 1 – Balancing between “training, no loan” vs. “training and loan”

	(1)		(2)		(3)		(4)		(5)	(6)
	Training, no loan		Training and loan						Diff.	SE
	Mean	SD	Mean	SD						
Age	27.463	4.696	28.220	4.763					0.757*	0.412
Female	0.644	0.479	0.565	0.497					-0.079*	0.042
Married	0.607	0.489	0.735	0.442					0.128***	0.041
Business ownership	0.641	0.480	0.720	0.450					0.079*	0.041
(Log) food consumption food	5.397	1.161	5.489	1.236					0.092	0.104
Formal bank account	0.429	0.496	0.425	0.496					-0.004	0.043
Formal loans in past 5 years	0.039	0.194	0.030	0.171					-0.009	0.016
Informal loans in past 5 years	0.131	0.338	0.175	0.381					0.044	0.031
Education level	No education	0.160	0.367	0.230	0.422				0.070**	0.034
	Some primary	0.296	0.457	0.300	0.459				0.004	0.040
	Some secondary	0.419	0.494	0.365	0.483				-0.054	0.043
	Some university	0.126	0.332	0.105	0.307				-0.021	0.028
Literacy	No English	0.209	0.407	0.260	0.440				0.051	0.037
	Some English	0.267	0.443	0.300	0.459				0.033	0.039
	Good English	0.524	0.500	0.440	0.498				-0.084*	0.044
Numeracy	Low	0.175	0.381	0.295	0.457				0.120***	0.036
	Counting 0-100	0.141	0.349	0.100	0.301				-0.041	0.029
	Addition	0.045	0.206	0.050	0.218				0.005	0.018
	Multiplication	0.639	0.481	0.555	0.498				-0.084**	0.043
Household size		7.003	3.238	8.010	3.653				1.007***	0.296
Number of children		3.141	2.239	3.635	2.505				0.494**	0.204
Number of elderly		0.073	0.340	0.115	0.391				0.042	0.031
Number of rooms		3.202	1.699	3.055	1.589				-0.147	0.145
Number of buildings		3.581	2.002	3.485	2.105				-0.096	0.178
(Log) distance to KCB branch		2.715	2.095	1.730	1.496				-0.985***	0.169
Conflict exposure (2011-2014)		0.236	4.583	0.026	0.548				-0.210	0.325
Conflict exposure (2015-2017)		0.166	3.888	0.196	1.034				0.029	0.280
Observations		382		200						
Joint orthogonality									<i>F</i> -stat	<i>p</i> -value
									4.024***	0.000

Notes: This table shows a balance test on baseline characteristics between participants that received “training, no loan” and participants that received “training and loan”. Columns (1) and (2) show the control group mean and SD, columns (3) and (4) show the treatment group mean and SD, column (5) shows the difference in means, and column (6) the respective standard errors of the difference in means tests. * (**, ***) indicates statistical significance at the ten-percent (five-percent, one-percent) level.

Table 2 – First stage results from LATE estimation

	(1) D_i^{T-L}	(2) D_i^{T-L}	(3) D_i^{T-L}
D_i^T	0.433*** (0.000)	0.403*** (0.000)	0.411*** (0.000)
Log distance to closest KCB branch	0.000 (1.000)	-0.012 (0.295)	-0.010 (0.367)
D_i^T x (log) distance to KCB branch	0.051*** (0.004)	0.072*** (0.000)	0.074*** (0.000)
Observations	1,378	1,378	1,378
Partial R_ρ^2	0.021	0.033	0.034
Effective F -stat	8.349	25.592	26.500
Geography controls	No	Yes	Yes
Baseline controls	No	No	Yes

Notes: This table displays the first stage results for LATE estimates. Column (1) is without controls only conditioned on a constant. Column (2) conditions on geographic controls, including state-gender fixed effects, as well as (log) distance to the closest road, (log) distance to the closest city center, land gradient, conflict exposure, and the interactions of the geographic controls with the treatment group dummy. Column (3) conditions on geographic controls and baseline controls, including baseline education level, baseline numeracy skills, baseline literacy level and their interaction with the treatment group dummy. Standard errors are clustered at the baseline boma level. P -values are in parentheses displayed below the estimated coefficients. * (**, ***) indicates statistical significance at the ten-percent (five-percent, one-percent) level.

Table 3 – Test of instrument correlation with baseline covariates

		(1)	(2)	(3)	(4)	(5)	(6)
		D_i^T		(Log) KCB distance		$D_i^T \times (\log) \text{ KCB dist.}$	
		coeff.	<i>p</i> -value	coeff.	<i>p</i> -value	coeff.	<i>p</i> -value
Age		0.141	0.727	-0.207	0.243	0.051	0.668
Married		-0.032	0.552	0.009	0.455	-0.004	0.766
Business ownership		0.082**	0.039	0.012	0.361	0.008	0.405
(Log) food consumption		0.227**	0.016	-0.053*	0.079	-0.007	0.762
Formal bank account		0.003	0.942	-0.012	0.330	-0.003	0.838
Formal loans (past 5 yrs.)		-0.011	0.548	-0.012**	0.015	0.010	0.104
Informal loans (past 5 yrs.)		-0.009	0.782	-0.009	0.246	0.009	0.243
Education	none	0.026	0.381	0.012	0.336	-0.001	0.868
	primary	-0.031	0.400	-0.022*	0.099	0.024***	0.006
	secondary	0.002	0.966	0.022*	0.063	-0.011	0.291
	tertiary	0.003	0.921	-0.013**	0.028	-0.012**	0.027
Literacy	No Engl.	-0.033	0.383	0.014	0.223	0.020*	0.070
	Some Engl.	0.019	0.604	0.004	0.729	-0.001	0.965
	Good Engl.	0.014	0.698	-0.018	0.110	-0.019*	0.050
Numeracy	Low	-0.019	0.565	-0.004	0.704	0.017*	0.051
	Count.	0.033	0.165	0.019*	0.098	-0.014*	0.057
	Add.	-0.005	0.788	0.011*	0.087	0.001	0.891
	Multipl.	-0.009	0.799	-0.025**	0.044	-0.004	0.761
Household size		-0.263	0.366	-0.187**	0.013	-0.007	0.913
Number of children		-0.225	0.219	-0.067	0.306	0.059	0.240
Number of elderly		-0.043	0.103	0.005	0.532	-0.000	0.967
Number of rooms		-0.105	0.470	-0.079*	0.099	-0.005	0.895
Number of buildings		-0.077	0.612	-0.071	0.124	0.010	0.792
Attrition		0.066	0.183	0.036	0.133	-0.001	0.954
Observations							2,256
Joint	<i>F</i> -stat						1.289
orthog.	<i>p</i> -value						(0.179)

Notes: This table reports results on reduced form estimation on baseline covariates and attrition status following Equation (3). All regression control for strata-fixed effects and our set of geographic control variables. Standard errors are clustered at the baseline boma level. Columns (1), (3), and (5) report the coefficients of the treatment group dummy, the logarithmic distance, and the interaction of distance with treatment group dummy. Columns (2), (4), and (6) report the respective *p*-values. * (**, ***) indicates statistical significance at the ten-percent (five-percent, one-percent) level.

Table 4 – LATE estimations on main outcomes

	(1)	(2)	(3)	(4)	(5)	(6)
	geo controls			geo + baseline controls		
	D_i^{T-L}	D_i^T	(1) + (2)	D_i^{T-L}	D_i^T	(4) + (5)
<i>Panel A: Main socio-economic outcomes</i>						
Employment index	-0.024 (0.953) [0.954]	0.173 (0.440) [0.703]	0.149 (0.519) [0.773]	-0.017 (0.967) [0.968]	0.418 (0.170) [0.338]	0.401 (0.161) [0.264]
Consumption index	-0.879** (0.013) [0.050]	0.495** (0.039) [0.076]	-0.384** (0.030) [0.120]	-0.974*** (0.005) [0.038]	0.248 (0.456) [0.608]	-0.725** (0.014) [0.055]
Savings index	-0.946** (0.023) [0.059]	0.820*** (0.003) [0.022]	-0.126 (0.580) [0.773]	-0.900** (0.026) [0.068]	0.741** (0.014) [0.053]	-0.159 (0.600) [0.601]
Business skills index	-0.077 (0.811) [0.928]	0.059 (0.791) [0.928]	-0.018 (0.904) [0.905]	-0.102 (0.756) [0.864]	-0.267 (0.346) [0.553]	-0.368 (0.198) [0.264]
<i>Panel B: Main psychological and behavioral outcomes</i>						
Psychological index	-0.728** (0.044) [0.371]	0.322 (0.158) [0.392]	-0.406** (0.050) [0.124]	-0.638* (0.068) [0.222]	0.267 (0.349) [0.437]	-0.371 (0.202) [0.505]
Risk index	-0.773 (0.293) [0.402]	0.504 (0.156) [0.392]	-0.269 (0.531) [0.531]	-0.690 (0.315) [0.437]	0.618 (0.143) [0.356]	-0.072 (0.861) [0.861]
Trust index	-0.612* (0.075) [0.371]	0.012 (0.953) [0.953]	-0.600*** (0.002) [0.009]	-0.613* (0.066) [0.222]	-0.257 (0.446) [0.495]	-0.869*** (0.003) [0.018]
Crime index	-0.452 (0.287) [0.402]	0.260 (0.297) [0.402]	-0.192 (0.407) [0.509]	-0.398 (0.350) [0.437]	0.118 (0.719) [0.720]	-0.281 (0.416) [0.559]
Migration index	-0.296 (0.322) [0.402]	0.149 (0.502) [0.558]	-0.148 (0.292) [0.487]	-0.355 (0.249) [0.437]	0.559* (0.051) [0.222]	0.204 (0.447) [0.559]
Observations			1,378			1,378
Effective F -stat			25.592			26.500

Notes: This table reports results from 2 different specifications of our main estimation. Columns (1) - (3) report results conditional on geographic controls, including state-gender fixed effects, as well as (log) distance to the closest road, (log) distance to the closest city center, land gradient, conflict exposure, and the interactions of the geographic controls with the treatment group dummy. Columns (4) - (6) report results conditional on geographic controls and baseline individual-level controls including, education level, baseline numeracy skills, baseline literacy level, and their interactions with the treatment group dummy. Columns (2) and (4) report coefficients on the treatment group dummy, columns (1) and (3) on the treatment subgroup “training, no loan”. Columns (3) and (6) report the main coefficients of interest, *i.e.* the combined effect of the treatment group dummy and the subgroup “training, no loan”. Standard errors are clustered at the baseline boma level. P -values are in parentheses displayed below the estimated coefficients. * (**, ***) indicates statistical significance at the ten-percent (five-percent, one-percent) level. Adjusted Benjamini-Hochberg p -values are reported in square brackets.

Table 5 – LATE estimations of consumption and savings index components

	(1)	(2)	(3)	(4)	(5)	(6)
	geo controls			geo + base controls		
	D_i^{T-L}	D_i^T	(1) + (2)	D_i^{T-L}	D_i^T	(4) + (5)
<i>Panel A: Consumption index components</i>						
No. of food items purchased	-2.726*** (0.001)	1.394** (0.016)	-1.332*** (0.001)	-2.609*** (0.000)	1.142* (0.093)	-1.467*** (0.004)
(Log) food purchased	-0.488** (0.028)	0.284 (0.103)	-0.204** (0.012)	-0.482** (0.038)	0.127 (0.561)	-0.354 (0.203)
(Log) food home production	0.147 (0.617)	-0.034 (0.832)	0.113 (0.493)	-0.008 (0.978)	-0.067 (0.790)	-0.075 (0.728)
(Log) non-food consumption	-0.949* (0.084)	0.662** (0.050)	-0.287 (0.338)	-1.051* (0.068)	0.519 (0.229)	-0.532 (0.133)
(Log) asset acquisition	-0.212 (0.423)	0.005 (0.978)	-0.207* (0.098)	-0.201 (0.482)	-0.278 (0.311)	-0.478* (0.054)
<i>Panel B: Savings index components</i>						
Bank account ownership	-0.348** (0.018)	0.271*** (0.003)	-0.078 (0.323)	-0.338** (0.020)	0.177 (0.226)	-0.162 (0.178)
Any savings	-0.197* (0.081)	0.146* (0.092)	-0.051 (0.290)	-0.194* (0.075)	0.166 (0.226)	-0.027 (0.825)
(Log) amount in account	-0.822 (0.127)	0.502* (0.093)	-0.319 (0.263)	-0.729 (0.144)	0.122 (0.757)	-0.607* (0.071)
No. of formal loans in past month	0.032 (0.692)	-0.056 (0.326)	-0.024 (0.448)	0.004 (0.962)	-0.029 (0.718)	-0.025 (0.757)
Any other debt	0.312 (0.263)	-0.183 (0.131)	0.129 (0.448)	0.325 (0.220)	-0.140 (0.330)	0.184 (0.395)
No. of informal loans in past month	0.105 (0.642)	-0.274* (0.059)	-0.169 (0.111)	0.137 (0.541)	-0.222 (0.248)	-0.085 (0.610)
(Log) amount formal loans	-0.238 (0.314)	0.007 (0.971)	-0.231*** (0.001)	-0.301 (0.239)	-0.005 (0.983)	-0.306 (0.166)
(Log) amount informal loans	-0.282 (0.556)	-0.127 (0.670)	-0.409 (0.105)	-0.294 (0.526)	-0.069 (0.854)	-0.363 (0.289)
Business ownership	-0.048 (0.717)	0.113 (0.155)	0.065 (0.365)	-0.112 (0.420)	0.122 (0.295)	0.011 (0.916)
Any training in past year	-0.315 (0.213)	0.159 (0.214)	-0.156 (0.310)	-0.284 (0.244)	0.253* (0.060)	-0.031 (0.856)
No. of trainings in past year	-0.534** (0.036)	0.260* (0.087)	-0.274* (0.082)	-0.457* (0.070)	0.286 (0.163)	-0.171 (0.339)
Observations			1,378			1,378
Effective <i>F</i> -stat			25.592			26.500

Notes: This table shows the results of two different specifications of our main estimation on components of the consumption index. Columns (1) - (3) condition on geographic controls, including state-gender fixed effects, as well as (log) distance to the closest road, (log) distance to the closest city center, land gradient, conflict exposure and the interactions of the geographic controls with the treatment group dummy. Columns (4) - (6) condition on geographic controls and baseline individual level controls including, education level, baseline numeracy skills, literacy level, and their interactions with the treatment group dummy. Standard errors are clustered at baseline boma level. *P*-values are in parentheses displayed below the estimated coefficients. * (**, ***) indicates statistical significance at the ten-percent (five-percent, one-percent) level.

Table 6 – LATE estimations of trust index components

	(1)	(2)	(3)	(4)	(5)	(6)
	geo controls			geo + base controls		
	D_i^{T-L}	D_i^T	(1) + (2)	D_i^{T-L}	D_i^T	(4) + (5)
<i>Panel A: Self reported measures</i>						
Most people can be trusted	-0.019 (0.954)	-0.142 (0.425)	-0.161 (0.395)	-0.225 (0.496)	0.050 (0.841)	-0.175 (0.549)
People try to be helpful	-0.343 (0.251)	-0.038 (0.823)	-0.381** (0.024)	-0.431 (0.164)	0.212 (0.403)	-0.218 (0.389)
Most people would try to take advantage if given the chance	-0.284 (0.387)	0.037 (0.850)	-0.247 (0.166)	-0.236 (0.488)	0.395 (0.160)	0.159 (0.631)
How often do you lend money	0.161 (0.832)	-0.177 (0.642)	-0.016 (0.971)	0.230 (0.765)	-0.267 (0.646)	-0.037 (0.947)
How often do you lend possessions	0.159 (0.817)	0.194 (0.601)	0.353 (0.374)	0.150 (0.832)	0.266 (0.611)	0.416 (0.360)
Trust in family	-0.014 (0.976)	-0.128 (0.577)	-0.143 (0.627)	-0.053 (0.914)	-0.354 (0.190)	-0.408 (0.281)
Trust in friends	-0.243 (0.580)	-0.010 (0.961)	-0.252 (0.347)	-0.256 (0.558)	-0.314 (0.297)	-0.570 (0.121)
Trust in neighbours	-0.356 (0.335)	0.032 (0.865)	-0.324 (0.142)	-0.305 (0.399)	-0.513* (0.088)	-0.819** (0.020)
Trust in police	-0.266 (0.437)	-0.093 (0.624)	-0.359* (0.056)	-0.325 (0.342)	-0.080 (0.790)	-0.405* (0.082)
Trust in NGO	-1.010*** (0.008)	0.432* (0.061)	-0.578*** (0.003)	-0.920*** (0.007)	0.137 (0.662)	-0.783*** (0.008)
Trust in elders	-0.777 (0.103)	0.271 (0.256)	-0.505* (0.071)	-0.837* (0.079)	-0.181 (0.484)	-1.018** (0.025)
Trust in local government	-0.263 (0.440)	-0.135 (0.479)	-0.398** (0.033)	-0.362 (0.260)	-0.438 (0.107)	-0.800*** (0.002)
Trust in SSD government	0.632 (0.138)	-0.361 (0.136)	0.271 (0.251)	0.582 (0.155)	-0.440 (0.186)	0.141 (0.641)
<i>Panel B: Trust game</i>						
Trust in WB	-18.953 (0.265)	5.742 (0.455)	-13.211 (0.194)	-18.009 (0.309)	2.025 (0.821)	-15.984 (0.232)
Trust in other player	2.143 (0.849)	-3.865 (0.593)	-1.721 (0.779)	6.961 (0.559)	-26.154** (0.011)	-19.193* (0.057)
Trustworthiness	1.587 (0.945)	-1.856 (0.851)	-0.269 (0.985)	4.148 (0.875)	1.532 (0.897)	5.681 (0.789)
Observations			1,378			1,378
Effective <i>F</i> -stat			25.592			26.500

Notes: This table shows the results of two different specifications of our main estimation on components of the trust index. Columns (1) - (3) condition on geographic controls, including state-gender fixed effects, as well as (log) distance to the closest road, (log) distance to the closest city center, land gradient, conflict exposure, and the interactions of the geographic controls with the treatment group dummy. Columns (4) - (6) condition on geographic controls and baseline individual level controls including, education level, baseline numeracy skills, baseline literacy level, and their interaction with the treatment group dummy. Standard errors are clustered at the baseline boma level. *P*-values are in parentheses displayed below the estimated coefficients. * (**, ***) indicates statistical significance at the ten-percent (five-percent, one-percent) level.

Table 7 – Two-IV-estimation on main outcomes

	(1)	(2)	(3)	(4)
	geo controls		geo + baseline controls	
	D_i^{T+L}	D_i^{T-L}	D_i^{T+L}	D_i^{T-L}
<i>Panel A: Main socio-economic outcomes</i>				
Employment index	0.344 (0.200) [0.373]	0.118 (0.565) [0.826]	0.175 (0.623) [0.885]	-0.081 (0.776) [0.885]
Consumption index	0.601** (0.049) [0.099]	-0.289** (0.040) [0.099]	0.226 (0.540) [0.885]	-0.720** (0.033) [0.201]
Savings index	1.015*** (0.003) [0.006]	0.009 (0.958) [0.958]	0.694* (0.068) [0.230]	-0.340 (0.347) [0.885]
Business skills index	-0.085 (0.726) [0.826]	0.056 (0.687) [0.826]	-0.119 (0.717) [0.885]	0.018 (0.944) [0.944]
<i>Panel B: Main psychological and behavioral outcomes</i>				
Psychological wellbeing index	0.190 (0.444) [0.726]	-0.266 (0.139) [0.366]	0.063 (0.846) [0.907]	-0.333 (0.198) [0.499]
Risk index	0.657* (0.091) [0.366]	-0.200 (0.563) [0.796]	0.472 (0.210) [0.499]	-0.438 (0.379) [0.744]
Trust index	-0.099 (0.663) [0.824]	-0.544*** (0.003) [0.010]	-0.218 (0.500) [0.814]	-0.703** (0.028) [0.185]
Crime and violence index	0.042 (0.868) [0.867]	-0.042 (0.837) [0.867]	-0.050 (0.875) [0.907]	-0.176 (0.574) [0.814]
Migration index	0.360 (0.197) [0.366]	-0.198 (0.162) [0.366]	0.040 (0.907) [0.907]	-0.500** (0.046) [0.185]
Observations		1,378		1,378
F-stat		27.386		26.259

Notes: This table tests the robustness of our main results against the use of two endogenous regressors and two instrumental variables, addressing the presence of non-compliers. It shows results from two different specifications. Columns (1) - (2) report results conditional on geographic controls, including state-gender fixed effects, as well as (log) distance to the closest road, (log) distance to the closest city center, land gradient, conflict exposure and the interactions of the geographic controls with the treatment group dummy. Columns (3) - (4) report results conditional on geographic controls and baseline individual-level controls including, education level, baseline numeracy skills, baseline literacy level, and their interaction with the treatment group dummy. Standard errors are clustered at baseline boma level. *P*-values are in parentheses displayed below the estimated coefficients. * (**, ***) indicates statistical significance at the ten-percent (five-percent, one-percent) level. Adjusted Benjamini-Hochberg *p*-values are reported in square brackets.

Table 8 – Sensitivity bounds for the LATE on main outcomes

	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: +/- .1 SD						
	control overperforms			“training, no loan” overperforms		
	D_i^{T-L}	D_i^T	(1) + (2)	D_i^{T-L}	D_i^T	(4) + (5)
Employment index	0.110 (0.749)	0.107 (0.578)	0.218 (0.269)	0.073 (0.838)	0.072 (0.709)	0.146 (0.481)
Consumption index	-1.007** (0.013)	0.613** (0.015)	-0.394** (0.046)	-1.044*** (0.008)	0.578** (0.022)	-0.465** (0.013)
Savings index	-0.679* (0.078)	0.661*** (0.002)	-0.018 (0.937)	-0.716** (0.043)	0.626*** (0.003)	-0.090 (0.670)
Business skills index	-0.042 (0.874)	0.055 (0.769)	0.012 (0.919)	-0.079 (0.766)	0.020 (0.914)	-0.059 (0.630)
Psychological wellbeing index	-0.314 (0.315)	0.064 (0.724)	-0.251 (0.153)	-0.351 (0.249)	0.029 (0.874)	-0.322* (0.055)
Risk index	-0.621 (0.332)	0.344 (0.263)	-0.277 (0.456)	-0.658 (0.286)	0.309 (0.309)	-0.349 (0.325)
Trust index	-0.333 (0.311)	-0.078 (0.659)	-0.411** (0.025)	-0.370 (0.220)	-0.113 (0.508)	-0.482*** (0.003)
Crime and violence index	-0.540 (0.102)	0.299 (0.158)	-0.241 (0.134)	-0.577 (0.108)	0.264 (0.229)	-0.313* (0.082)
Migration index	-0.322 (0.266)	0.159 (0.417)	-0.163 (0.241)	-0.359 (0.217)	0.124 (0.534)	-0.235* (0.089)
Panel B: +/- .25 SD						
	control overperforms			“training, no loan” overperforms		
	D_i^{T-L}	D_i^T	(1) + (2)	D_i^{T-L}	D_i^T	(4) + (5)
Employment index	0.138 (0.684)	0.134 (0.492)	0.271 (0.159)	0.046 (0.902)	0.046 (0.815)	0.092 (0.673)
Consumption index	-0.980** (0.019)	0.640** (0.011)	-0.340 (0.103)	-1.071*** (0.006)	0.552** (0.030)	-0.519*** (0.004)
Savings index	-0.652 (0.114)	0.687*** (0.002)	0.035 (0.887)	-0.744** (0.026)	0.600*** (0.004)	-0.144 (0.467)
Business skills index	-0.015 (0.957)	0.081 (0.673)	0.066 (0.598)	-0.107 (0.695)	-0.007 (0.970)	-0.113 (0.389)
Psychological wellbeing index	-0.287 (0.378)	0.090 (0.621)	-0.197 (0.287)	-0.379 (0.215)	0.002 (0.989)	-0.376** (0.024)
Risk index	-0.594 (0.369)	0.371 (0.234)	-0.223 (0.563)	-0.686 (0.254)	0.283 (0.351)	-0.403 (0.241)
Trust index	-0.305 (0.388)	-0.051 (0.779)	-0.357* (0.076)	-0.397 (0.165)	-0.139 (0.408)	-0.536*** (0.000)
Crime and violence index	-0.512 (0.103)	0.325 (0.119)	-0.187 (0.212)	-0.605 (0.116)	0.237 (0.294)	-0.367* (0.063)
Migration index	-0.294 (0.320)	0.185 (0.343)	-0.109 (0.452)	-0.386 (0.194)	0.098 (0.631)	-0.288** (0.043)
Observations			1,879			1,879
Effective F-stat			28.157			28.157

Notes: This table derives sensitivity bounds for LATE results of columns 1 - 3 of Table 4, assuming that all treatment group attritors received “training, no loan”. Panel A column 1 - 3 assume that all treatment group attritors underperformed by - 0.1 SD of their group mean and control group attritors overperformed by +0.1 SD of their group mean, Panel A column 4 - 6 assumes that all treatment group attritors overperformed by + 0.1 SD of their group mean and control group attritors underperformed by -0.1 SD of their group mean, Panel B column 1-3 assumes that all treatment group attritors underperformed by - 0.25 SD of their group mean and control group attritors overperformed by +0.25 SD of their group mean, Panel B column 4-6 assumes that all treatment group attritors overperformed by + 0.25 SD of their group mean and control group attritors underperformed by -0.25 SD of their group mean. *P*-values are in parentheses displayed below the estimated coefficients. * (**, ***) indicates statistical significance at the ten-percent (five-percent, one-percent) level.

Online Appendix

Broken Promises: Evaluating an Incomplete Startup Business Program

Angelika J. Budjan, Utz Pape and Laura Ralston*

June 27, 2025

*Corresponding author: Budjan (University of Stuttgart), email: angelika@budjan.de; Pape (World Bank), email: upape@worldbank.org; Ralston (World Bank), email: lauraralston@gmail.com.

A Additional Tables

Table A.1 – Description of balancing variables

variable	description
Age	age of respondent [<i>discrete</i>]
Married	dummy variable marking married respondents with 1 [<i>binary</i>]
Business ownership	dummy variable marking respondents that own a business or are self-employed with 1 [<i>binary</i>]
(Log) consumption food	Logarithm of total food consumption of food items in the last 7 days (including home production or gifts) converted to monetary values in SSP; items include chicken, fish, beef goat, lamb, pork, rice, bananas, orange, other fruits, soft drinks, alcohol, sweets, cigarettes, honey, sugar, bottled water, restaurant meals [<i>continuous</i>]
Formal bank account	dummy variable marking respondents that hold a formal bank account with 1 [<i>binary</i>]
Formal loans in past 5 years	Dummy variable indicating any formal loans in the past 5 years [<i>binary</i>]
Informal loans in past 5 years	Dummy variable indicating any informal loans in the past 5 years [<i>binary</i>]
Education level	dummy variable marking respondents without formal education with 1 [<i>binary</i>]
Some Primary	dummy variable marking respondents with some primary school education with 1 [<i>binary</i>]
Some Secondary	dummy variable marking respondents with some secondary school education with 1 [<i>binary</i>]
University	dummy variable marking respondents with some tertiary education with 1 [<i>binary</i>]
Literacy	dummy variable marking respondents with no English literacy with 1 [<i>binary</i>]
No English	dummy variable marking respondents with no English literacy with 1 [<i>binary</i>]
Some English	dummy variable marking respondents with some English literacy with 1 [<i>binary</i>]
Good English	dummy variable marking respondents with good English literacy with 1 [<i>binary</i>]
Numeracy	dummy variable marking respondents without number recognition up to 100 [<i>binary</i>]
Low	dummy variable marking respondents without number recognition up to 100 [<i>binary</i>]
Counting 0-100	dummy variable marking respondents with number recognition up to 100 [<i>binary</i>]
Addition	dummy variable marking respondents that were able to perform simple addition tasks with 1 [<i>binary</i>]
Multiplication	dummy variable marking respondents that were able to perform simple multiplication tasks with 1 [<i>binary</i>]
Household size	number of persons living in the same household [<i>discrete</i>]
Number of children	number of children living in the same household [<i>discrete</i>]
Number of elderly	number of elderly living in the same household [<i>discrete</i>]
Number of rooms	number of rooms the household occupies [<i>discrete</i>]
Number of buildings	number of building the household occupies [<i>discrete</i>]
(Log) distance to KCB branch	Natural logarithm of distance to closest KCB branch in km
Conflict exposure (2011-2014)	Spatial lag of fatalities from conflict events reported in the UCDP data set within a 300km buffer from 2011-2014
Conflict exposure (2015-2017)	Spatial lag of fatalities from conflict events reported in the UCDP data set within a 300km buffer from 2015-2017

Table A.2 – Balancing original control and treatment group at baseline

	(1)	(2)	(3)	(4)	(5)	(6)
	Control group		Treatment group			
	Mean	SD	Mean	SD	Diff.	SE
<i>Panel A: Baseline covariates</i>						
Age	27.417	4.901	27.683	5.014	0.265	0.207
Female	0.602	0.490	0.611	0.488	0.009	0.020
Married	0.666	0.472	0.649	0.477	-0.016	0.020
Business ownership	0.642	0.480	0.659	0.474	0.017	0.020
(Log) food consumption food	5.330	1.212	5.400	1.251	0.070	0.051
Formal bank account	0.373	0.484	0.369	0.483	-0.004	0.020
Formal loans in past 5 years	0.044	0.204	0.041	0.199	-0.002	0.008
Informal loans in past 5 years	0.152	0.360	0.148	0.355	-0.005	0.015
Education level						
No education	0.191	0.393	0.206	0.405	0.016	0.017
Some primary	0.315	0.465	0.330	0.471	0.015	0.020
Some secondary	0.404	0.491	0.373	0.484	-0.031	0.020
Some university	0.090	0.286	0.090	0.286	0.000	0.012
Literacy						
No English	0.247	0.432	0.263	0.441	0.016	0.018
Some English	0.273	0.446	0.295	0.456	0.022	0.019
Good English	0.480	0.500	0.442	0.497	-0.038*	0.021
Numeracy						
Low	0.238	0.426	0.247	0.432	0.010	0.018
Counting 0-100	0.105	0.306	0.128	0.334	0.023*	0.013
Addition	0.047	0.212	0.049	0.216	0.002	0.009
Multiplication	0.611	0.488	0.576	0.494	-0.035*	0.021
Household size	7.310	3.377	7.260	3.392	-0.050	0.141
Number of children	3.107	2.282	3.241	2.325	0.134	0.096
Number of elderly	0.109	0.344	0.087	0.334	-0.021	0.014
Number of rooms	3.180	1.775	3.087	1.641	-0.093	0.071
Number of buildings	3.676	1.951	3.538	1.867	-0.138*	0.080
(Log) distance to KCB branch	6.578	2.349	6.589	2.337	0.011	0.099
Conflict exposure (2011-2014)	-0.000	1.000	0.084	2.720	0.084	0.086
Conflict exposure (2015-2017)	-0.000	1.000	0.128	2.405	0.128*	0.077
Observations	1,148		1,144			
Joint orthogonality					<i>F</i> -stat	<i>p</i> -value
					1.030	0.422
<i>Panel B: States</i>						
Central Equatoria	0.169	0.375	0.167	0.373	-0.002	0.016
Eastern Equatoria	0.160	0.367	0.152	0.359	-0.008	0.015
Lakes	0.158	0.365	0.159	0.366	0.001	0.015
Northern Bahr El Ghazal	0.170	0.376	0.176	0.381	0.006	0.016
Western Bahr El Ghazal	0.172	0.377	0.171	0.377	-0.000	0.016
Western Equatoria	0.172	0.377	0.175	0.380	0.003	0.016
Observations	1,148		1,144			
Joint orthogonality					<i>F</i> -stat	<i>p</i> -value
					0.083	0.995

Notes: Columns (1) and (2) show the control group mean and SD, columns (3) and (4) show the treatment group mean and SD, column (5) shows the difference in means, and column (6) the respective standard errors of the difference in means tests.* (**, ***) indicates statistical significance at the ten-percent (five-percent, one-percent) level.

Table A.3 – Balance test between attritors from original control vs. attritors from original treatment group

	(1)	(2)	(3)	(4)	(5)	(6)	
	Control group		Treatment group				
	Mean	SD	Mean	SD	Diff.	SE	
<i>Panel A: Baseline covariates</i>							
Age	27.161	5.196	27.682	5.210	0.522	0.348	
Female	0.636	0.482	0.587	0.493	-0.049	0.033	
Married	0.670	0.471	0.644	0.479	-0.026	0.032	
Business ownership	0.672	0.470	0.644	0.479	-0.028	0.032	
(Log) food consumption food	5.283	1.291	5.371	1.307	0.088	0.087	
Formal bank account	0.335	0.472	0.320	0.467	-0.015	0.031	
Formal loans in past 5 years	0.050	0.218	0.056	0.229	0.006	0.015	
Informal loans in past 5 years	0.127	0.333	0.140	0.347	0.013	0.023	
Education level	No education	0.210	0.408	0.207	0.405	-0.004	0.027
	Some primary	0.344	0.476	0.367	0.482	0.023	0.032
	Some secondary	0.380	0.486	0.367	0.482	-0.013	0.032
	Some university	0.066	0.248	0.060	0.238	-0.006	0.016
Literacy	No English	0.274	0.446	0.287	0.453	0.013	0.030
	Some English	0.258	0.438	0.304	0.461	0.047	0.030
	Good English	0.468	0.500	0.409	0.492	-0.059*	0.033
Numeracy	Low	0.258	0.438	0.262	0.440	0.004	0.029
	Counting 0-100	0.124	0.330	0.129	0.335	0.004	0.022
	Addition	0.052	0.222	0.049	0.216	-0.003	0.015
	Multiplication	0.566	0.496	0.560	0.497	-0.006	0.033
Household size		7.299	3.470	7.151	3.471	-0.148	0.232
Number of children		3.113	2.297	3.140	2.341	0.027	0.155
Number of elderly		0.111	0.349	0.082	0.298	-0.029	0.022
Number of rooms		3.147	1.807	3.042	1.670	-0.105	0.116
Number of buildings		3.683	1.872	3.593	1.714	-0.090	0.120
(Log) distance to KCB branch		2.523	1.937	2.452	1.924	-0.071	0.130
Conflict exposure (2011-2014)		0.002	1.107	-0.042	0.653	-0.044	0.061
Conflict exposure (2015-2017)		0.001	1.068	0.036	0.915	0.035	0.067
Observations		442		450			
Joint orthogonality					<i>F</i> -stat	<i>p</i> -value	
					0.889	0.619	
<i>Panel B: States</i>							
Central Equatoria		0.165	0.372	0.173	0.379	0.008	0.025
Eastern Equatoria		0.156	0.363	0.191	0.394	0.035	0.025
Lakes		0.174	0.380	0.187	0.390	0.012	0.026
Northern Bahr El Ghazal		0.161	0.368	0.144	0.352	-0.016	0.024
Western Bahr El Ghazal		0.181	0.385	0.156	0.363	-0.025	0.025
Western Equatoria		0.163	0.370	0.149	0.356	-0.014	0.024
Observations		442		450			
Joint orthogonality					<i>F</i> -stat	<i>p</i> -value	
					0.673	0.644	

Notes: This table shows a balance test on baseline characteristics between attritors of the control group and attritors of the treatment group. Columns (1) and (2) show the control group mean and SD, columns (3) and (4) show the treatment group mean and SD, column (5) shows the difference in means, and column (6) the respective standard errors of the difference in means tests.* (**, ***) indicates statistical significance at the ten-percent (five-percent, one-percent) level.

Table A.4 – Difference between attritors and non-attritors

	(1)	(2)	(3)	(4)	(5)	(6)
	Non-attritors		Attritors			
	Mean	SD	Mean	SD	Diff.	SE
<i>Panel A: Baseline covariates</i>						
Age	27.630	4.794	27.424	5.207	-0.206	0.212
Female	0.604	0.489	0.611	0.488	0.007	0.021
Married	0.658	0.475	0.657	0.475	-0.001	0.020
Employment status	0.618	0.486	0.619	0.486	0.001	0.021
Business ownership	0.646	0.478	0.658	0.475	0.012	0.020
(Log) food consumption	5.388	1.187	5.328	1.299	-0.060	0.053
(Log) nonfood consumption	2.428	1.333	2.416	1.339	-0.012	0.057
Formal bank account	0.399	0.490	0.327	0.470	-0.071***	0.021
Formal loans in past 5 years	0.036	0.186	0.053	0.224	0.017**	0.009
Informal loans in past 5 years	0.161	0.367	0.133	0.340	-0.027*	0.015
Education level						
No education	0.192	0.394	0.209	0.406	0.016	0.017
Primary	0.302	0.459	0.355	0.479	0.053***	0.020
Secondary	0.399	0.490	0.373	0.484	-0.025	0.021
University	0.107	0.309	0.063	0.243	-0.044***	0.012
Literacy						
No English	0.239	0.427	0.280	0.449	0.041**	0.019
Some English	0.285	0.452	0.281	0.450	-0.004	0.019
Good English	0.476	0.500	0.438	0.496	-0.037*	0.021
Numeracy						
Low	0.231	0.422	0.260	0.439	0.029	0.018
Counting 0-100	0.109	0.312	0.127	0.333	0.017	0.014
Addition	0.046	0.210	0.050	0.219	0.004	0.009
Multiplication	0.613	0.487	0.563	0.496	-0.050**	0.021
Household size	7.324	3.330	7.224	3.469	-0.100	0.145
Number of children	3.204	2.295	3.127	2.318	-0.078	0.099
Number of elderly	0.099	0.348	0.096	0.324	-0.003	0.015
Number of rooms	3.159	1.691	3.094	1.739	-0.065	0.073
Number of buildings	3.587	1.982	3.638	1.794	0.051	0.082
(Log) distance to KCB branch	2.337	1.948	2.487	1.930	0.151*	0.084
Conflict exposure (2011-2014)	0.081	2.517	-0.020	0.907	-0.101	0.088
Conflict exposure (2015-2017)	0.093	2.218	0.018	0.993	-0.075	0.079
Observations	1,400		892			
Joint orthogonality					<i>F</i> -stat	<i>p</i> -value
					1.931***	0.004
<i>Panel B: States</i>						
Central Equatoria	0.174	0.379	0.158	0.365	-0.016	0.016
Eastern Equatoria	0.156	0.363	0.156	0.363	-0.001	0.016
Lakes	0.144	0.351	0.182	0.386	0.038**	0.016
Northern Bahr El Ghazal	0.174	0.379	0.170	0.376	-0.004	0.016
Western Bahr El Ghazal	0.168	0.374	0.177	0.382	0.009	0.016
Western Equatoria	0.184	0.387	0.157	0.364	-0.027	0.016
Observations	1,400		892			
Joint orthogonality					<i>F</i> -stat	<i>p</i> -value
					1.678	0.136

Notes: This table shows a balance test on baseline characteristics between non-attritors (i.e. endline participants) and attritors. Columns (1) and (2) show the control group mean and SD, columns (3) and (4) show the treatment group mean and SD, column (5) shows the difference in means, and column (6) the respective standard errors of the difference in means tests. * (**, ***) indicates statistical significance at the ten-percent (five-percent, one-percent) level.

Table A.5 – ρ -test for main estimation

	(1) geo controls	(2) geo + baseline controls
Employment index	0.049	0.040
Consumption index	0.390	0.423
Savings index	0.399	0.380
Business skills index	0.022	0.011
Psychological wellbeing index	0.353	0.306
Risk index	0.369	0.322
Trust index	0.210	0.207
Crime and violence index	0.237	0.216
Migration index	0.056	0.082

Notes: This table presents the ‘degree of endogeneity’-test following (Angrist and Kolesár, 2024). We report the respective test statistic $|\rho|$ for each of our main regressions from Table 4. Angrist and Kolesár (2024) argue that t -ratio inference in a just identified IV estimation is valid if the degree of endogeneity ρ is low. Column (2) presents the test statistics for regression with only geographic controls and column (2) for regressions with geographic and baseline controls. For all regressions, $|\rho|$ remains below the critical value of 0.56, implying that t -ratio inference is valid.

Interpretation. In the case of IV estimations the use of t -ratio inferences becomes a concern if the degree of endogeneity ρ is very large, making the usual t -test confidence intervals unreliable. In a single IV setting, this occurs if $|\rho|$ exceeds the critical value of 0.565 (Lee et al., 2022, Angrist and Kolesár, 2024). Table A.5 shows that in our main estimations, $|\rho|$ remains below this critical value.

Table A.6 – tF -confidence invertervals of main results from Table 4

	(1)	(2)	(3)	(4)	(5)	(6)
	geo controls			geo + baseline controls		
	D_i^{T-L}	D_i^T	(1) + (2)	D_i^{T-L}	D_i^T	(4) + (5)
<i>Panel A: Main socio-economic outcomes</i>						
Employment index	-0.024 (0.953) [-1.00 - 0.95]	0.173 (0.440)	0.149 (0.519) [-0.41 - 0.42]	-0.017 (0.967) [-1.02 - 0.98]	0.418 (0.170)	0.401 (0.161) [-0.28 - 1.09]
Consumption index	-0.879** (0.013) [-1.73 - -0.03]	0.495** (0.039)	-0.384** (0.030) [-0.81 - -0.18]	-0.974*** (0.005) [-1.80 - -0.15]	0.248 (0.456)	-0.725** (0.014) [-1.43 - -0.02]
Savings index	-0.946** (0.023) [-1.95 - 0.06]	0.820*** (0.003)	-0.126 (0.580) [-0.68 - 0.14]	-0.900** (0.026) [-1.87 - 0.07]	0.741** (0.014)	-0.159 (0.600) [-0.89 - 0.57]
Business skills index	-0.077 (0.811) [-0.86 - 0.71]	0.059 (0.791)	-0.018 (0.904) [-0.39 - 0.16]	-0.102 (0.756) [-0.88 - 0.68]	-0.267 (0.346)	-0.368 (0.198) [-1.05 - 0.32]
<i>Panel B: Main psychological and behavioral outcomes</i>						
Psychological index	-0.728** (0.044) [-1.60 - 0.15]	0.322 (0.158)	-0.406** (0.050) [-0.90 - -0.16]	-0.638* (0.068) [-1.47 - 0.20]	0.267 (0.349)	-0.371 (0.202) [-1.07 - 0.32]
Risk index	-0.773 (0.293) [-2.55 - 1.01]	0.504 (0.156)	-0.269 (0.531) [-1.31 - 0.24]	-0.690 (0.315) [-2.34 - 0.96]	0.618 (0.143)	-0.072 (0.861) [-1.06 - 0.92]
Trust index	-0.612* (0.075) [-1.44 - 0.22]	0.012 (0.953)	-0.600*** (0.002) [-1.06 - -0.38]	-0.613* (0.066) [-1.41 - 0.18]	-0.257 (0.446)	-0.869*** (0.003) [-1.58 - -0.16]
Crime index	-0.452 (0.287) [-1.48 - 0.58]	0.260 (0.297)	-0.192 (0.407) [-0.75 - 0.08]	-0.398 (0.350) [-1.42 - 0.62]	0.118 (0.719)	-0.281 (0.416) [-1.11 - 0.55]
Migration index	-0.296 (0.322) [-1.02 - 0.43]	0.149 (0.502)	-0.148 (0.292) [-0.49 - 0.02]	-0.355 (0.249) [-1.09 - 0.38]	0.559* (0.051)	0.204 (0.447) [-0.44 - 0.85]
Observations			1,378			1,378
Effective F -stat			25.592			26.500

Notes: This table reports results from 2 different specifications. Columns (1) - (3) report results conditional on geographic controls, including state-gender fixed effects, as well as (log) distance to the closest road, (log) distance to the closest city center, land gradient, conflict exposure, and the interactions of the geographic controls with the treatment group dummy. Columns (4) - (6) report results conditional on geographic controls and baseline individual-level controls including, education level, baseline numeracy skills, baseline literacy skills, and their interaction with the treatment group dummy. Standard errors are clustered at the baseline boma level. P -values are in parentheses displayed below the estimated coefficients. * (**, ***) indicates statistical significance at the ten-percent (five-percent, one-percent) level. tF -Confidence intervals are reported in square brackets.

Interpretation. Table A.6 replicates Table 4 while including tF -confidence intervals that assume worst-case scenarios for the degree of endogeneity following (Lee et al., 2022). The results show that the negative effects of “training, no loan” remain significant under these extreme assumptions.

Table A.7 – *Ex post* minimum detectable effect size: LATE estimates

	(1)	(2)	(3)	(4)	(5)	(6)
	geo controls			geo + baseline controls		
	D_i^{T-L}	D_i^T	(1) + (2)	D_i^{T-L}	D_i^T	(4) + (5)
Employment index	0.790	0.440	0.455	0.820	0.599	0.561
Consumption index	0.692	0.469	0.346	0.678	0.654	0.576
Savings index	0.813	0.537	0.447	0.793	0.589	0.597
Business skills index	0.637	0.438	0.299	0.642	0.556	0.561
Psychological wellbeing index	0.709	0.447	0.405	0.685	0.561	0.571
Risk index	1.446	0.697	0.845	1.351	0.829	0.812
Trust index	0.675	0.393	0.372	0.654	0.662	0.579
Crime and violence index	0.834	0.490	0.455	0.838	0.644	0.678
Migration index	0.588	0.435	0.276	0.605	0.562	0.528

Table A.8 – LATE estimations on main outcomes in the female sample

	(1)	(2)	(3)	(4)	(5)	(6)
	geo controls			geo + baseline controls		
	D_i^{T-L}	D_i^T	(1) + (2)	D_i^{T-L}	D_i^T	(4) + (5)
<i>Panel A: Main socio-economic outcomes</i>						
Employment index	-0.005 (0.991) [0.602]	0.284 (0.261) [0.444]	0.279 (0.342) [0.456]	-0.021 (0.969) [0.970]	0.267 (0.470) [0.752]	0.246 (0.486) [0.487]
Consumption index	-1.309* (0.058) [0.152]	0.399 (0.191) [0.379]	-0.910* (0.068) [0.272]	-1.369** (0.026) [0.155]	0.540 (0.253) [0.505]	-0.829 (0.102) [0.407]
Savings index	-1.202** (0.033) [0.126]	0.798 (0.013)** [0.100]	-0.404 (0.239) [0.456]	-1.056** (0.040) [0.155]	0.591 (0.123) [0.324]	-0.465 (0.228) [0.457]
Business skills index	-0.208 (0.727) [0.831]	-0.045 (0.870) [0.870]	-0.254 (0.527) [0.527]	-0.257 (0.673) [0.830]	-0.139 (0.726) [0.830]	-0.395 (0.437) [0.487]
<i>Panel B: Main psychological and behavioral outcomes</i>						
Psychological index	-0.764 (0.190) [0.377]	0.281 (0.317) [0.451]	-0.483 (0.209) [0.261]	-0.573 (0.253) [0.392]	-0.033 (0.925) [0.925]	-0.607* (0.093) [0.156]
Risk index	-2.109* (0.077) [0.349]	0.687 (0.133) [0.349]	-1.422* (0.076) [0.191]	-2.068** (0.049) [0.235]	1.017 (0.108) [0.235]	-1.051* (0.087) [0.156]
Trust index	-0.615 (0.141) [0.349]	-0.147 (0.492) [0.492]	-0.763*** (0.005) [0.025]	-0.656 (0.117) [0.235]	-0.055 (0.885) [0.925]	-0.711* (0.073) [0.156]
Crime index	-0.607 (0.361) [0.451]	0.211 (0.492) [0.492]	-0.396 (0.365) [0.366]	-0.744 (0.275) [0.392]	0.234 (0.562) [0.702]	-0.510 (0.332) [0.416]
Migration index	-0.789 (0.115) [0.349]	0.325 (0.304) [0.451]	-0.464 (0.126) [0.210]	-0.788 (0.119) [0.235]	0.813* (0.099) [0.235]	0.026 (0.939) [0.939]
Observations			831			831
F-stat			10.368			13.144

Notes: This table reports results from 2 different specifications on the female sample. Columns (1) - (3) report the results of our main estimation conditional on geographic controls, including state fixed effects, as well as (log) distance to the closest road, (log) distance to the closest city center, land gradient, conflict exposure, and the interactions of the geographic controls with the treatment group dummy. Columns (4) - (6) report results conditional on geographic controls and baseline individual-level controls including, education level, baseline numeracy skills, baseline literacy skills, and their interaction with the treatment group dummy. Standard errors are clustered at the baseline boma level. *P*-values are in parentheses displayed below the estimated coefficients. * (**, ***) indicates statistical significance at the ten-percent (five-percent, one-percent) level. Adjusted Benjamini-Hochberg *p*-values are reported in square brackets.

Table A.9 – LATE estimations on main outcomes in the male sample

	(1)	(2)	(3)	(4)	(5)	(6)
	geo controls			geo + baseline controls		
	D_i^{T-L}	D_i^T	(1) + (2)	D_i^{T-L}	D_i^T	(4) + (5)
<i>Panel A: Main socio-economic outcomes</i>						
Employment index	0.129 (0.858) [0.982]	0.086 (0.841) [0.982]	0.215 (0.606) [0.670]	0.354 (0.681) [0.775]	0.836 (0.172) [0.479]	1.190 (0.126) [0.253]
Consumption index	-0.571 (0.273) [0.555]	0.371 (0.348) [0.555]	-0.201 (0.390) [0.670]	-0.600 (0.328) [0.479]	-0.525 (0.360) [0.479]	-1.126* (0.076) [0.253]
Savings index	-0.844 (0.181) [0.555]	0.722 (0.127) [0.555]	-0.123 (0.670) [0.670]	-0.625 (0.338) [0.479]	0.910* (0.093) [0.479]	0.285 (0.589) [0.625]
Business skills index	0.130 (0.816) [0.982]	0.433 (0.287) [0.555]	0.562** (0.019) [0.077]	0.199 (0.774) [0.775]	-0.478 (0.324) [0.479]	-0.279 (0.624) [0.625]
<i>Panel B: Main psychological and behavioral outcomes</i>						
Psychological index	-0.762 (0.209) [0.818]	0.299 (0.426) [0.818]	-0.463 (0.182) [0.492]	-0.400 (0.546) [0.833]	0.650 (0.222) [0.833]	0.249 (0.715) [0.894]
Risk index	0.335 (0.594) [0.818]	0.031 (0.931) [0.932]	0.366 (0.295) [0.492]	0.521 (0.417) [0.833]	0.445 (0.424) [0.833]	0.966 (0.113) [0.282]
Trust index	-0.854 (0.255) [0.818]	0.392 (0.344) [0.818]	-0.462 (0.270) [0.492]	-1.014 (0.212) [0.833]	-0.657 (0.257) [0.833]	-1.671** (0.016) [0.083]
Crime index	-0.268 (0.667) [0.818]	0.168 (0.711) [0.818]	-0.100 (0.733) [0.733]	-0.162 (0.815) [0.833]	0.109 (0.833) [0.833]	-0.053 (0.913) [0.914]
Migration index	0.297 (0.644) [0.818]	-0.118 (0.736) [0.818]	0.179 (0.629) [0.733]	0.266 (0.717) [0.833]	0.224 (0.607) [0.833]	0.490 (0.426) [0.710]
Observations	547			547		
Effective F -stat	24.624			17.599		

Notes: This table reports results from 2 different specifications in the male sample. Columns (1) - (3) report results conditional on geographic controls, including state fixed effects, as well as (log) distance to the closest road, (log) distance to the closest city center, land gradient, conflict exposure, and the interactions of the geographic controls with the treatment group dummy. Columns (4) - (7) report results conditional on geographic controls and baseline individual-level controls including, education level, baseline numeracy skills, baseline literacy skills, and their interaction with the treatment group dummy. Standard errors are clustered at the baseline boma level. P -values are in parentheses displayed below the estimated coefficients. * (**, ***) indicates statistical significance at the ten-percent (five-percent, one-percent) level. Adjusted Benjamini-Hochberg p -values are reported in square brackets.

Gender heterogeneity. The difference in consumption between female (Table A.8) and male (Table A.8) participants is -0.7 (p -value = 0.245) with geographic controls and 0.29 (p -value = 0.685) with full controls. The difference in trust is -0.30 (p -value = 0.50) with geographic controls and 0.960 (p -value = 0.275) with full controls.

Table A.10 – Robustness of LATE results from Table 4 to different conflict measures

	(1)		(2)		(3)		(4)		(5)		(6)		(7)		(8)		(9)		(10)		(11)		(12)				
	D_i^{T-L}	D_i^T	D_i^{T-L}	D_i^T	(1) + (2)	D_i^{T-L}	D_i^T	(4) + (5)	D_i^{T-L}	D_i^T	D_i^{T-L}	D_i^T	(7) + (8)	D_i^{T-L}	D_i^T	(7) + (8)	D_i^{T-L}	D_i^T	(7) + (8)	D_i^{T-L}	D_i^T	(7) + (8)	D_i^{T-L}	D_i^T	(7) + (8)		
Employment index	-0.024 (0.954)	0.173 (0.447)	-0.291 (0.547)	0.336 (0.207)	0.149 (0.519)	-0.291 (0.547)	0.336 (0.207)	0.045 (0.859)	-0.052 (0.899)	0.195 (0.406)	0.144 (0.522)	-0.165 (0.694)	0.252 (0.298)	-0.052 (0.899)	0.195 (0.406)	0.144 (0.522)	-0.165 (0.694)	0.252 (0.298)	-0.052 (0.899)	0.195 (0.406)	0.144 (0.522)	-0.165 (0.694)	0.252 (0.298)	-0.052 (0.899)	0.195 (0.406)	0.144 (0.522)	
Consumption index	-0.879** (0.020)	0.495** (0.050)	-1.073** (0.024)	0.636** (0.042)	-0.384** (0.029)	-1.073** (0.024)	0.636** (0.042)	-0.437** (0.033)	-0.737** (0.046)	0.418** (0.092)	-0.319* (0.054)	-0.869** (0.041)	0.489* (0.089)	-0.737** (0.046)	0.418** (0.092)	-0.319* (0.054)	-0.869** (0.041)	0.489* (0.089)	-0.737** (0.046)	0.418** (0.092)	-0.319* (0.054)	-0.869** (0.041)	0.489* (0.089)	-0.737** (0.046)	0.418** (0.092)	-0.319* (0.054)	
Savings index	-0.946** (0.032)	0.820*** (0.006)	-1.138** (0.033)	0.940*** (0.010)	-0.126 (0.579)	-1.138** (0.033)	0.940*** (0.010)	-0.197 (0.433)	-1.079** (0.019)	0.894*** (0.005)	-0.186 (0.413)	-1.098** (0.022)	0.907*** (0.007)	-1.079** (0.019)	0.894*** (0.005)	-0.186 (0.413)	-1.098** (0.022)	0.907*** (0.007)	-1.079** (0.019)	0.894*** (0.005)	-0.186 (0.413)	-1.098** (0.022)	0.907*** (0.007)	-1.079** (0.019)	0.894*** (0.005)	-0.186 (0.413)	
Business skills index	-0.077 (0.813)	0.059 (0.793)	-0.076 (0.849)	0.058 (0.829)	-0.018 (0.904)	-0.076 (0.849)	0.058 (0.829)	-0.017 (0.919)	0.056 (0.867)	-0.008 (0.973)	0.048 (0.762)	0.127 (0.718)	-0.053 (0.824)	0.056 (0.867)	-0.008 (0.973)	0.048 (0.762)	0.127 (0.718)	-0.053 (0.824)	0.056 (0.867)	-0.008 (0.973)	0.048 (0.762)	0.127 (0.718)	-0.053 (0.824)	0.056 (0.867)	-0.008 (0.973)	0.048 (0.762)	
Psychological wellbeing index	-0.728* (0.056)	0.322 (0.171)	-0.947** (0.039)	0.470 (0.103)	-0.406** (0.049)	-0.947** (0.039)	0.470 (0.103)	-0.477** (0.031)	-0.454 (0.202)	0.177 (0.403)	-0.277 (0.157)	-0.408 (0.267)	0.141 (0.515)	-0.454 (0.202)	0.177 (0.403)	-0.277 (0.157)	-0.408 (0.267)	0.141 (0.515)	-0.454 (0.202)	0.177 (0.403)	-0.277 (0.157)	-0.408 (0.267)	0.141 (0.515)	-0.454 (0.202)	0.177 (0.403)	-0.277 (0.157)	
Risk index	-0.773 (0.304)	0.504 (0.169)	-1.175 (0.195)	0.768 (0.105)	-0.269 (0.530)	-1.175 (0.195)	0.768 (0.105)	-0.408 (0.392)	-0.772 (0.282)	0.505 (0.147)	-0.267 (0.519)	-1.002 (0.192)	0.631 (0.108)	-0.772 (0.282)	0.505 (0.147)	-0.267 (0.519)	-1.002 (0.192)	0.631 (0.108)	-0.772 (0.282)	0.505 (0.147)	-0.267 (0.519)	-1.002 (0.192)	0.631 (0.108)	-0.772 (0.282)	0.505 (0.147)	-0.267 (0.519)	
Trust index	-0.612* (0.088)	0.012 (0.953)	-0.642 (0.124)	0.045 (0.850)	-0.600*** (0.002)	-0.642 (0.124)	0.045 (0.850)	-0.597*** (0.004)	-0.536 (0.131)	-0.032 (0.875)	-0.568*** (0.003)	-0.527 (0.146)	-0.034 (0.871)	-0.536 (0.131)	-0.032 (0.875)	-0.568*** (0.003)	-0.527 (0.146)	-0.034 (0.871)	-0.536 (0.131)	-0.032 (0.875)	-0.568*** (0.003)	-0.527 (0.146)	-0.034 (0.871)	-0.536 (0.131)	-0.032 (0.875)	-0.568*** (0.003)	
Crime and violence index	-0.452 (0.298)	0.260 (0.308)	-0.612 (0.221)	0.354 (0.235)	-0.192 (0.406)	-0.612 (0.221)	0.354 (0.235)	-0.258 (0.300)	-0.165 (0.685)	0.118 (0.614)	-0.047 (0.831)	-0.104 (0.807)	0.066 (0.781)	-0.165 (0.685)	0.118 (0.614)	-0.047 (0.831)	-0.104 (0.807)	0.066 (0.781)	-0.165 (0.685)	0.118 (0.614)	-0.047 (0.831)	-0.104 (0.807)	0.066 (0.781)	-0.165 (0.685)	0.118 (0.614)	-0.047 (0.831)	
Migration index	-0.296 (0.332)	0.149 (0.509)	-0.254 (0.484)	0.113 (0.670)	-0.148 (0.291)	-0.254 (0.484)	0.113 (0.670)	-0.141 (0.352)	-0.260 (0.382)	0.142 (0.514)	-0.119 (0.391)	-0.381 (0.265)	0.199 (0.395)	-0.260 (0.382)	0.142 (0.514)	-0.119 (0.391)	-0.381 (0.265)	0.199 (0.395)	-0.260 (0.382)	0.142 (0.514)	-0.119 (0.391)	-0.381 (0.265)	0.199 (0.395)	-0.260 (0.382)	0.142 (0.514)	-0.119 (0.391)	
Observations					1,378			1,378			1,378					1,378					1,378						
F-stat					25.592			18.003			24.392					24.392											

Notes: This table tests the robustness of LATE results in columns (1) - (3) of Tables 4 against different measures for conflict exposure. Columns (1) - (3) repeat columns (1) - (3) of Table 4 with the main conflict measure used in all results of this paper. It consists of the distance-weighted number of fatalities based on UCDDP data. Columns (4) - (6) use the distance-weighted number of conflict events based on UCDDP data, columns (7) - (9) use distance weighted number of fatalities based on ACLED data, and columns (10) - (12) use the distance-weighted number of conflict events based on ACLED data. All regressions condition on geographic controls. Geographic controls include state-gender fixed effects, as well as (log) distance to the closest road, (log) distance to the closest city center, land gradient, conflict exposure and the interactions of the geographic controls with the treatment group dummy. Standard errors are clustered at baseline boma level. *P*-values are in parentheses displayed below the estimated coefficients. * (**, ***) indicates statistical significance at the ten-percent (five-percent, one-percent) level.

Table A.11 – Robustness of LATE effects to including replacement observation

	(1)	(2)	(3)	(4)	(5)	(6)
	geo controls			geo + baseline controls		
	D_i^{T-L}	D_i^T	(1) + (2)	D_i^{T-L}	D_i^T	(4) + (5)
Employment index	-0.039 (0.920) [0.920]	0.194 (0.366) [0.585]	0.155 (0.502) [0.618]	0.040 (0.922) [0.923]	0.439 (0.109) [0.217]	0.479 (0.110) [0.220]
Consumption index	-1.080*** (0.005) [0.015]	0.597** (0.024) [0.046]	-0.483** (0.011) [0.044]	-1.247*** (0.002) [0.012]	0.406 (0.212) [0.339]	-0.841*** (0.005) [0.020]
Savings index	-1.044*** (0.006) [0.015]	0.813*** (0.001) [0.008]	-0.232 (0.308) [0.616]	-1.019*** (0.004) [0.017]	0.781*** (0.008) [0.020]	-0.237 (0.328) [0.328]
Business skills index	-0.138 (0.636) [0.848]	0.066 (0.741) [0.848]	-0.072 (0.618) [0.618]	-0.195 (0.519) [0.692]	-0.091 (0.710) [0.811]	-0.286 (0.325) [0.328]
Psychological index	-0.835** (0.014) [0.134]	0.364 (0.111) [0.367]	-0.471** (0.013) [0.032]	-0.754** (0.018) [0.175]	0.377 (0.161) [0.320]	-0.378 (0.161) [0.403]
Risk index	-0.779 (0.273) [0.525]	0.444 (0.196) [0.487]	-0.335 (0.422) [0.515]	-0.710 (0.278) [0.428]	0.610 (0.120) [0.299]	-0.100 (0.807) [0.807]
Trust index	-0.611* (0.063) [0.313]	0.018 (0.925) [0.925]	-0.593*** (0.001) [0.006]	-0.649** (0.049) [0.221]	-0.332 (0.343) [0.428]	-0.981*** (0.001) [0.004]
Crime index	-0.359 (0.408) [0.525]	0.197 (0.410) [0.525]	-0.161 (0.515) [0.515]	-0.326 (0.472) [0.524]	0.096 (0.753) [0.753]	-0.230 (0.521) [0.651]
Migration index	-0.262 (0.420) [0.525]	0.112 (0.611) [0.679]	-0.150 (0.368) [0.515]	-0.326 (0.342) [0.428]	0.514* (0.067) [0.221]	0.188 (0.491) [0.651]
Observations			1,500			1,500
F-stat			26.744			28.434

Notes: This table replicates LATE results from Table 4 without excluding observations from the replacement pool. As in Table 4, we report results from 2 different specifications. Columns (1) - (3) report results conditional on geographic controls, including state-gender fixed effects, as well as (log) distance to the closest road, (log) distance to the closest city center, land gradient, conflict exposure, and the interactions of the geographic controls with the treatment group dummy. Columns (4) - (6) report results conditional on geographic controls and baseline individual-level controls including, education level, baseline numeracy skills, baseline literacy skills, and their interaction with the treatment group dummy. Standard errors are clustered at baseline boma level. *P*-values are in parentheses displayed below the estimated coefficients. * (**, ***) indicates statistical significance at the ten-percent (five-percent, one-percent) level. Adjusted Benjamini-Hochberg *p*-values are reported in square brackets.

B Methodological details on lab-in-field experiments

B.1 Lotteries

This study uses choices over lotteries that vary in expected return and variance to extract risk preferences. In the endline, data collection respondents were asked to choose between two or three alternative lotteries. The design of this experiment involved eight rounds, building on the research design by [Jakiela and Ozier \(2019\)](#). After choosing one option, the chosen lottery was played as a flip of a fair coin (50 percent chance of each outcome). The game started with two practice rounds to make participants familiar with the rules. After that, the participants had to play six additional rounds. At the end of the game, one round was selected at random and the lottery chosen by the participants was played and paid out. Participants were informed about these rules at the beginning of the game. The lotteries are set up as described below in table B.1.

Table B.1 – Pay-outs of lotteries, expected utility

	Lottery A		Lottery B		Lottery C	
	Heads	Tails	Heads	Tails	Heads	Tails
Practice						
Decision 1	100	100	150	150		
Decision 2	100	150	200	250		
Game						
Decision 3	100	100	100	120		
Decision 4	100	100	0	400		
Decision 5	30	340	100	100	0	400
Decision 6	100	100	55	240	30	340
Decision 7	30	230	60	170	90	110
Decision 8	10	200	70	160	90	110

The number of times respondents chose the riskiest lottery can be used as a proxy for their risk preferences. Given that respondents in these types of experiments often display choices that are inconsistent with constant relative risk aversion (CRRA), a non-parametric approach to measure risk aversion is more appropriate. Thus, following the approach put forward by [Jakiela and Ozier \(2019\)](#), the set of lottery choices can also be used to infer risk preferences in a less stringent and non-theoretic manner. One measure is created by counting how many times respondents choose the riskiest lotteries, i.e., lotteries with the largest spread, or the safest lotteries. In addition, the likelihood of choosing the riskier lottery during each decision round

was evaluated individually. The results are then compared to survey answers on risk preferences.

Test questions were included to detect biased answers that resulted from a lack of understanding. Due to the relatively low numeracy skills and the complexity of the lotteries, the study included 3 questions to test for monotonicity, i.e., if participants behaved like utility-maximizers (Andreoni and Sprenger, 2010). If participants answered more than 1 of these test questions in a way inconsistent with utility maximization, it is likely that they simply did not understand the nature of the decision problem.

B.2 Trust game

To assess treatment effects on trust levels, survey participants were asked to participate in two versions of a trust game. The basic structure of a trust game involves Player A receiving an endowment of X and choosing how much of this endowment to send to Player B, $Y \in [0, X]$. Player B receives $3Y$ – i.e., three times whatever A sends him – and must decide how much of this endowment to send back to A, $Z \in [0, 3Y]$. A ends with $X - Y + Z$ and B ends with $3Y - Z$. Y/X is used as a measure of trust. $Z/3Y$ is used as a measure of trustworthiness. The table below summarizes payouts for the two players:

Table B.2 – Trust game payouts

Player A			Player B		
Endowment	Sends	Payout	Endowment	Sends	Payout
X	Y	$X - Y + Z$	$3Y$	Z	$3Y - Z$

first, we asked participants to play a practice round which they played as Player A to ensure that all participants understood the rules of the game. Then, we asked participants to play the following two versions of a trust game.

Game 1: Participants vs. the World Bank.

In the first game, all survey participants were assigned to player type A and received an initial endowment of 100 SSP (approximately 0.83 USD). To obtain a measure of trust of the participant towards the World Bank – which participants may hold responsible for the (non-) payment of the business start-up loans - we framed the game as if Player 2 was the World

Bank. This enabled a direct measure of how willing participants are to partake in an interaction with the World Bank that could have financial consequences. Hence it can be interpreted as a measure of how participation in the Business Start-up Competition has influenced their level of trust and their willingness to interact with the World Bank. The reciprocal behavior of Player B was modeled to mirror the probability of non-disbursement of the cash loan. In 34 percent of the cases documented by the phone survey, participants received the loan. This information was used to define the reciprocal behavior of Player B. Player B played fairly 34 percent of the time – that is, Player B returns exactly half of what they obtained from the study participant. In 66 percent of the cases, Player B acted unfairly and kept all that was sent to them, regardless of what the respondent sent. In the end, the participant was paid out of the budget of Player A. This means they learned the result of game 1 before participating in game 2. Since game 2 pits them against a different player, learning applied if at all only to the rules of the game and not to the strategic behavior of the other player.

Game 2: Participants against each other

To obtain a more general measure of the participants' trust levels, and to accompany the first measure, a second game was played which pits the participants against each other. The survey participants were equally and randomly selected as Players A and B, stratified by treatment groups and treatment strands. Participants assigned to player type A received an initial endowment of 100 SSP (approximately 0.83 USD). Regarding the implementation of the games and pairing of the players, a lab-in-the-field experimental setup was impossible to organize because respondents had to be interviewed individually. This was primarily due to the complicated logistical circumstances surrounding fieldwork in South Sudan, in no small part due to rapidly deteriorating security conditions, but also due to constraints on the respondents' time. Respondents were, therefore, playing the games against a pre-loaded hypothetical distribution of responses. Enumerators explained to the respondents that the other player would be another survey participant elsewhere in South Sudan. The set of possible responses, in terms of the fraction of the endowment sent or returned, was equally distributed between $[0.1, 1]$ in increments of 0.1. In no cases was the fraction of endowment

sent or returned equally to zero. In the end, the participants received the corresponding pay-out from the game. The pay-out was independent of what they had received in the first game.

Since game 2 assigns half of all participants to player type A and half of all participants to player type B, player behavior measures two separate concepts. As outlined above, the behavior of player A is a measure of trust, while the behavior of player B is a measure of trustworthiness. We analyze these two concepts separately in the analysis in section 6. To allow for the inclusion of trust game results into the main index, we include both normalized measures as they are. While this might reduce the power of the estimation, it is unlikely that this will bias the results, since player type assignment was randomized by strata and treatment arm. Therefore, the aggregate index is more likely to respond to the treatment if the treatment affects both trust and trustworthiness in the trust game.

References

- Andreoni, J. and Sprenger, C. (2010), ‘Certain and uncertain utility: the Allais paradox and five decision theory phenomena’, *Levine’s Working Paper Archive* .
- Angrist, J. and Kolesár, M. (2024), ‘One instrument to rule them all: The bias and coverage of just-ID IV’, *Journal of Econometrics* **240**(2), 105398.
- Jakiela, P. and Ozier, O. (2019), ‘The impact of violence on individual risk preferences: Evidence from a natural experiment’, *Review of Economics and Statistics* **101**(3), 547–559.
- Lee, D. S., McCrary, J., Moreira, M. J. and Porter, J. (2022), ‘Valid t-ratio inference for IV’, *American Economic Review* **112**(10), 3260–3290.