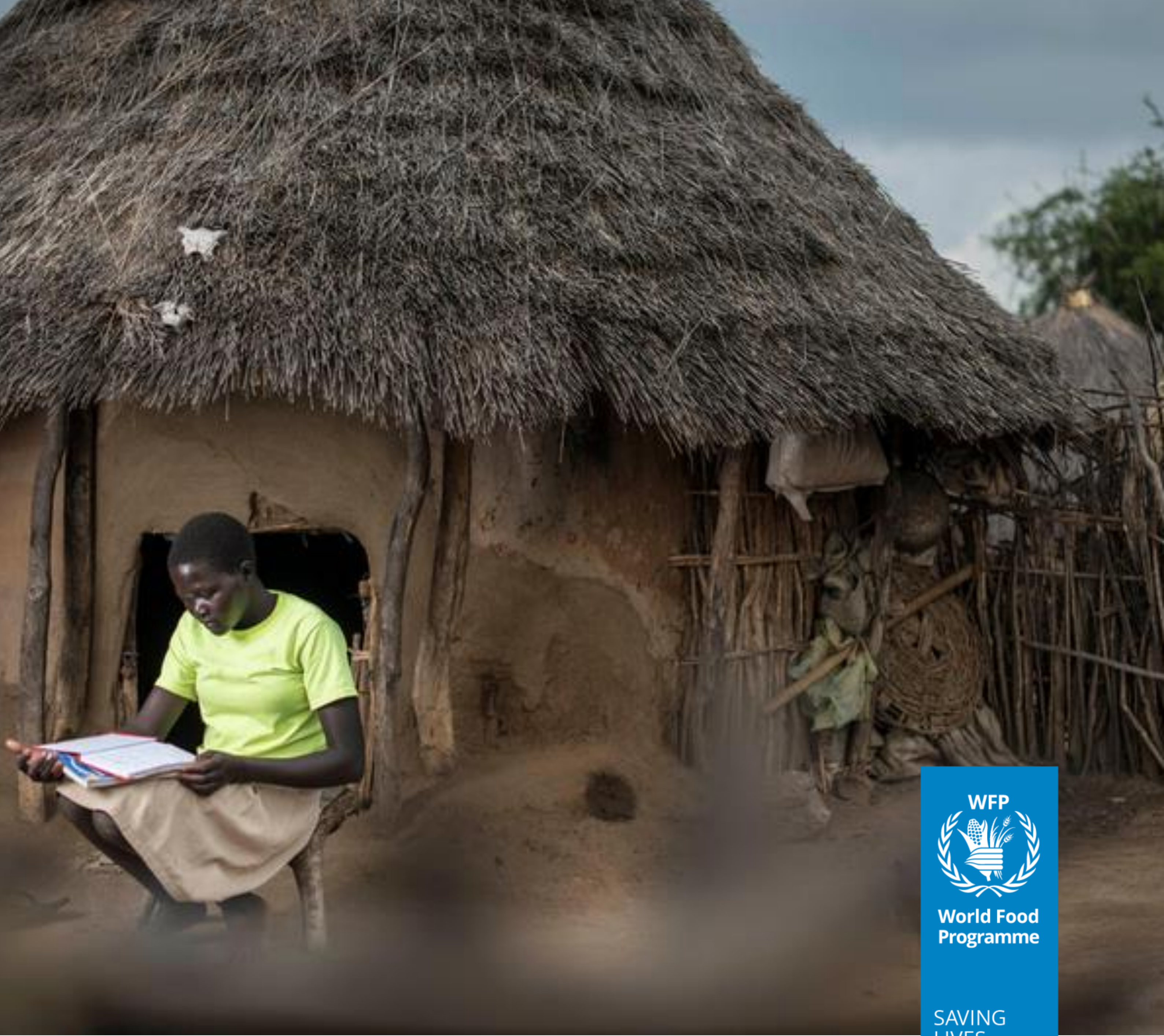


WFP
World Food Programme

SAVING LIVES
CHANGING LIVES

Financial Literacy in Refugee Settlements

An Evidence-Based Analysis

By Valerio Giuffrida[1], Viola Maone[2], Cristobal Mingo[3], and Denni Tommasi[4]

[1] Giuffrida: Research Assessment and Monitoring division, World Food Programme. Via Cesare Giulio Viola, 68/70, Roma, 00148, Italy. E-mail: valerio.giuffrida@wfp.org.
[2] Maone: Department of Economics, University of Bologna. Piazza Scaravilli 2, Bologna, 40126, Italy. E-mail: viola.maone2@unibo.it.
[3] Mingo: Programs and Policy division, World Food Programme. Via Cesare Giulio Viola, 68/70, Roma, 00148, Italy. E-mail: cristobal.mingo@wfp.org.
[4] Tommasi: Department of Economics, University of Bologna. Piazza Scaravilli 2, Bologna, 40126, Italy. E-mail: denni.tommasi@unibo.it.

Contents

Abstract

We present the first empirical evidence on the impact of a financial literacy training (FLT) program on financial knowledge and the adoption of digital financial services (DFS) among refugees living in Uganda. The program was a high-intensity intervention targeting households already receiving food assistance in the form of cash. Using a fuzzy Difference-in-Difference approach, we show the program significantly boosted financial knowledge. However, its impact on using DFS, like making digital transactions, was limited. Our key takeaway is that, while high-intensity programs work well in some contexts, they may not lead to substantial DFS adoption among refugees within a short timeframe. Further research is needed to optimize financial inclusion strategies for refugees and similar vulnerable populations.

Word count: 9,233 words
JEL Codes: D12, D91, I38, O12, Q12, O15.
Keywords: Refugee economies, Humanitarian aid, Cash transfers, Financial Literacy Training (FLT), Digital Financial Services (DFS).

We wish to gratefully acknowledge the continuous support of the Programme and Research Assessment and Monitoring teams for their unwavering assistance. In particular, we are deeply grateful to Wendy Alvarado for her substantial support, Margaret Katuutu for her practical contributions, and Rosemary Mwesigwa and Lilian Likicho for their invaluable insights and feedback. We extend our appreciation to the World Food Programme Country Office in Uganda and the Finnish Refugee Council for their exceptional professionalism. Special recognition is due to Elkhidir Daloum, the Country Director of the WFP Country Office in Uganda, for recognizing the value of this research and granting us access to the necessary information. Without his support, this work would not have been possible. We also thank Maddalena Davoli, Olivier Sterck and Till Wicker for very thoughtful comments that have improved an earlier version of the paper. All remaining errors are ours.

1 Introduction

The United Nations General Assembly's approach to refugee assistance underwent a significant change in 2016 with the adoption of the New York Declaration for Refugees and Migrants (UNHCR, 2018).[5] This marked a departure from the long-standing humanitarian model towards a development-focused approach, leading a number of host nations to initiate pioneering policies aimed at enhancing refugee agency and promoting local development (e.g., MacPherson and Sterck, 2021; Altındağ and O'Connell, 2023).[6] A popular form of intervention for promoting sustainable well-being is the provision of financial literacy trainings.[7] This is due to their ability to grant marginalized individuals access to digital financial services and credit markets, thereby enhancing their financial inclusion and stability (Lusardi and Mitchell, 2014). Despite the increasing amount of research on the interventions and outcomes for displaced individuals in low- and middle-income countries, there is still limited evidence on the effectiveness of financial literacy trainings for this population (Schuettler and Do, 2023). Addressing this gap in the literature is crucial in order to provide valuable insights to policymakers, aid organizations, and stakeholders involved in designing effective interventions for displaced populations.

We provide the first empirical evidence of a financial literacy training (FLT) program's impact on refugees' financial knowledge and digital financial services (DFS) adoption.[8] The FLT program was part of the World Food Programme's (WFP) financial inclusion campaign with the general scope of increasing financial inclusion of forcibly displaced people (WFP, 2022b). The targeted population was composed of refugees from South Sudan, the Democratic Republic of Congo, Somalia and Burundi. Since many refugees in Uganda settle semi-permanently (ECHO, 2024), developing financial skills is crucial for their integration. These targeted refugees already received cash transfers through WFP's general food assistance program, and a majority regularly used mobile phones at baseline.[9] Hence, the FLT program aimed to empower these vulnerable groups to utilize DFS by engaging in low-cost activities, like making digital transactions through e-wallets or bank accounts.

Our analysis is based on data collected before and after a short-term intervention conducted in four refugee settlements in Uganda from 2020-2021. Designed as a "high-intensity" program, with 20 hours of coursework in 5 weeks,[10] the implementation followed a phased-in approach, with all households becoming eligible for the program, but at different times. Settlements were divided into zones, and pre-specified geographic criteria were used to determine which zones would receive the training or serve as control groups. Approximately one third (37%) of households began training immediately, while the remaining two thirds (63%) were placed on a waiting list. With the program's phased implementation and simultaneous measurement of the outcome variables for treatment and control households, those waiting for the FLT served as the control group for those treated earlier.

The complexity of the setting made the implementation and evaluation of the experiment challenging. Firstly, the assignment process had imbalances in variables between treatment and control groups. Secondly, some participants did not comply with their assignment, with treated households not taking up the program and control households doing so. Additionally, some participants began a FLT program before the data collection started. To address the first problem, we employ a "sharp" difference-in-difference (DID) strategy, controlling for the

[5] The declaration acknowledges that refugees have the potential to be self-sufficient and make valuable contributions to their host communities, given equal access to job markets, education, healthcare, and livelihood opportunities without discrimination. The full text of the declaration is available at this link.

[6] The humanitarian model assumes that refugee situations are temporary, which suggests that aid efforts should primarily focus on protection, emergency relief, and shelter until the crises are resolved. However, this approach is based on flawed assumptions, as research has shown that only a small fraction of refugee crises (2.5 percent) are resolved within three years (Cosgrove et al., 2016).

[7] Financial literacy refers to an individual's capacity to comprehend and make well-informed decisions about financial planning, managing wealth, handling debt, and planning for retirement (OECD, 2021).

[8] The DFS that we consider include mobile money, agency banking, internet banking, online remittances.

[9] See Section 2 for baseline statistics.

[10] The term "high-intensity" is utilized in the study by Kaiser et al. (2022) to distinguish between the various types of educational interventions found within the literature on financial literacy education. These interventions span from providing informational brochures (e.g., Choi et al., 2010) to delivering intensive classroom instruction (e.g., Bruhn et al., 2016). According to the researchers, the average intensity of FLT interventions examined in the literature is 11.7 hours (with a median of 7 hours). Considering that our intervention extends for almost twice this duration, it positions us at the upper end of the spectrum in terms of intensity within the existing literature.

allocation rule used to decide treatment status and, in our preferred specification, we also control for household fixed effects. Although the DID strategy might lower precision, it is crucial in contexts of imperfect randomization where the presence of a baseline helps controlling for unobserved, time-invariant factors that could affect the outcome variables (e.g., McKenzie, 2012, page 211, footnote 2). This choice benefits our analysis by allowing us to address concerns that some observed differences in our balancing tests are not just due to chance. To address the second problem, we ex- tend our approach and utilize a “fuzzy” DID strategy, which considers non-compliance (de Chaisemartin and D’HaultfŒuille, 2017). The fuzzy DID estimator operates similarly to an Instrumental Variables (IV) estimation, where assignment and treatment status are differentiated. We focused on estimating the Local Average Treatment Effect (LATE) for “switchers,” measuring the effect on participants who showed a positive treatment rate between the first and the second period.

The FLT was primarily intended to promote financial inclusion by empowering refugees in the adoption of DFS. Notably, the FLT was not designed to help refugees accumulate financial capital through savings, given the precarious economic and living conditions of our study population. The adoption of DFS has proven to be a powerful tool for increasing financial inclusion among poor and marginalized communities that are typically excluded from the formal financial sector. Following Batista and Vicente (2013, 2020), we test the hypothesis that providing refugees with both training and access to safe and cheap DFS will induce a substitution effect - which would be unlikely to happen otherwise in such population - between formal financial services and traditional money management technologies, which are usually more expensive, risky and time consuming.

The results are fourfold. Firstly, focusing on four decision inputs related to the financial education program, we find that refugees who received the FLT experienced significant increases in their level of financial knowledge (0.65 SD), financial planning (0.74 SD), and agency in domestic finance matters (0.62 SD). These large effects are likely driven by the relatively short time between the intervention and the endline survey (3 months).[11] Secondly, when considering downstream financial behaviors and DFS adoption, we estimate a positive impact on households reporting a savings goal (0.59 SD) and regularly using budgets to control spending (0.37 SD). We also find suggestive, marginally significant, positive effects on mobile phone use (0.20 SD). However, the program did not significantly impact other outcomes, such as the number of digital transactions through e-wallets, or opening bank accounts and saving formally. Unlike the earlier results on decision inputs, the relatively short exposure to the program likely played against the potential effectiveness of the training on adoption. Thirdly, to gain deeper insights into the absence of DFS adoption, we explore potential explanations. Besides the short timeframe, low trust and low quality of financial institutions likely hindered stronger effects of the FLT on DFS adoption. Finally, we conduct comprehensive ex-post power calculations to pinpoint potential areas for improvement. We show that even with a sample size of 3200 observations collected over two periods and a plausibly credible research design, statistical power can be severely limited by the number of independent clusters. We conclude the paper by pointing out that donors and implementing organizations should take statistical power to detect effects on focal outcomes as a first-order consideration when designing and/or implementing quantitative evaluations of their programs.

We contribute primarily to the emerging literature studying the outcomes of refugees and internally displaced persons. A key finding from this literature is that the economic outcomes of these groups are strongly influenced by the conditions and policies they face upon reaching their destination (Schuettler and Do, 2023). Economists have explored various interventions to address the unique challenges faced by refugees, with cash transfer programs being a primary example of a policy utilized to partially offset the loss of assets and income (Hidrobo et al., 2014; Fiala, 2015; Chaaban et al., 2020; Aygün et al., 2021; MacPherson and Sterck, 2021; Quattrochi et al., 2022; Altındağ and O’Connell, 2023). Our study builds upon this research by examining the immediate impact of a financial literacy training program on the financial knowledge and DFS of refugees already receiving cash food assistance. Financial literacy training has shown positive effects on financial knowledge, financial services adoption and savings decisions. However, a recent meta-analysis by Kaiser et al. (2022) found heterogeneous effects based on program type and target population. We highlight the potential benefits of financial education in underdeveloped contexts, validating self-reliance assistance programs, and guiding policies that can potentially promote refugee financial inclusion.

Furthermore, a strand of literature has studied DFS adoption among poor and marginalized communities traditionally

[11] The average measurement horizon of the FLT interventions analyzed in the literature is 6 months (Kaiser et al., 2022). The choice of 3 months reflects the donor’s prioritization to study the effectiveness of a shorter-term evaluation.

excluded from the formal financial sector through targeted interventions providing both access and training to these technologies (e.g., Batista and Vicente, 2013, 2020; Breza et al., 2020). However, to the best of our knowledge, no study has investigated DFS adoption within a refugee context. Our contribution to this literature is demonstrating potential factors to consider when designing adoption campaigns in such contexts.

The paper is organized as follows. Section 2 presents the setting and the data. Section 3 discusses the empirical framework. Section 4 presents the main results. We conclude in section 5 with recommendations for future work in this area. Additional results and tables are included in the appendix.

2 Context, design and data

The United Nations High Commissioner for Refugees (UNHCR) reported that, in 2020, there were 103 million people forcibly displaced worldwide, with 32.5 million being refugees. Uganda is among the countries hosting the highest number of refugees, along with Türkiye, Germany, and Pakistan. Uganda accommodates over 1.5 million refugees, primarily from South Sudan, the Democratic Republic of Congo, and Burundi (UNHCR, 2021b). Over 80% of these refugees are hosted in 13 villages known as settlements situated in the North and South-Western regions, co-existing with local communities. For a significant share of these refugees, partial or full integration has transformed forced displacement from a temporary situation into a long-term reality (ECHO, 2024).

The context for refugees in Uganda have some very unique features. On one hand, the government provides households with land to farm and settle, freedom of movement, the right to search for employment and start business activities, and access to public services such as health and education (Betts et al., 2019). On the other hand, the WFP fulfills the essential requirements of refugees living in Uganda by offering unconditional and unrestricted food assistance through two main modalities: In-kind and Cash Based Transfers (CBT) (WFP, 2022a). Regarding the latter, cash assistance is distributed through both digital channels, such as agency banking and mobile money, as well as non-digital means, like cash-in-hand (UNHCR, 2021a). In the case of digital delivery mechanisms, cash is directly transferred to the beneficiaries' owned bank accounts or SIM cards. Once the transfer is completed, beneficiaries can then proceed to any agent or merchant point within the settlement to withdraw or use their cash as needed. To ensure convenient access to cash out points, Financial Service Providers (FSPs) actively participate in establishing agent points and providing them with the necessary liquidity within the target communities. Furthermore, to cater to the diverse needs of the recipients and ensure efficient assistance reaches those in need, the WFP collaborates closely with FSPs to facilitate the delivery of cash-in-hand to the designated beneficiaries at various Final Distribution Points (FDPs).

WFP general food assistance program in Uganda provides access to adequate nutritious food levels. At the time of the data collection (both baseline and endline) refugees were receiving 70% of the minimum expenditure basket ration across the country.[12] The food assistance provided was roughly equivalent to $18 USD per month per person registered in the household. Hence, the actual amount received by the household head (man or woman) varied depending on family size. It is important to note that while the WFP offered food assistance in both in-kind and cash, the FLT program only targeted households receiving cash. This distinction is crucial because the ability to sell received food likely impacts saving decisions. Unlike cash, selling large quantities of distributed food (like tomatoes) at once in a limited number of markets can significantly decrease its resale value. Therefore, focusing solely on cash recipients makes the program's impact on financial behavior more easily measurable.

Starting from 2020, the WFP Country Office in Uganda started implementing financial literacy activities with the scope of increasing the financial capability, knowledge, attitudes, skills and behavior of forcibly displaced persons and hosting communities. Particular attention has been given to the possibility of women emerging as economic actors and financial decision-makers within the household (WFP, 2022b).

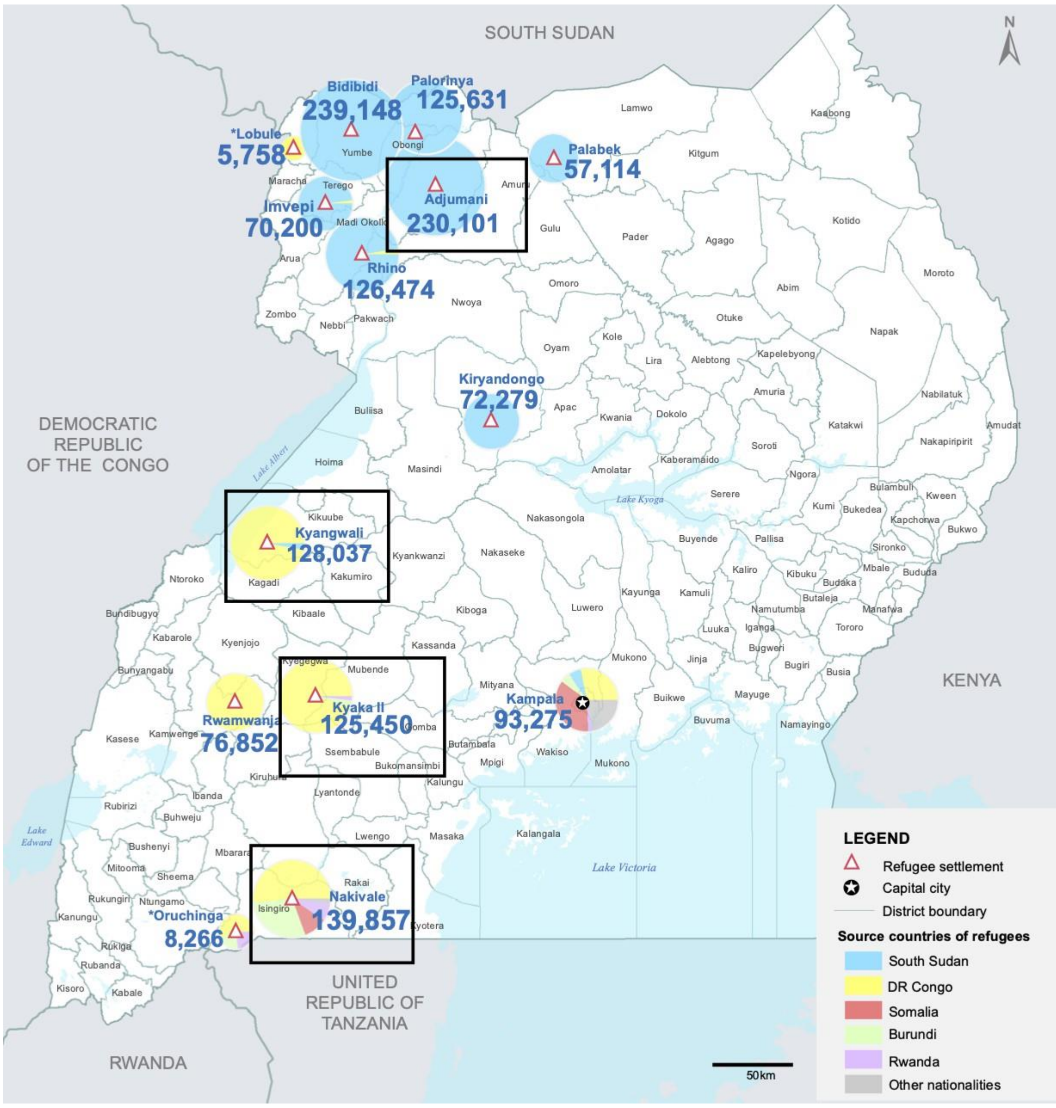
[12] Indeed, due to the declining long-term resourcing outlook, food rations were reduced from 100 to 70 percent in April 2020. For more details, see the WFP 2020 annual country reports for Uganda, available here. Further reductions in the assistance level were implemented when the data collection was already over (after November 2021), through a geographic prioritization scheme based on the relative socio-economic ability of refugees to meet their food consumption needs. Settlements classified into these groups received 70 percent, 60 percent, and 40 percent of the recommended dietary allowance, respectively.

2.1 Design

In 2020-2021, the WFP and its implementing partner, the Finnish Refugee Council (FRC), implemented a FLT program in 11 refugee settlements. The experimental sample considered in this study is made of eligible refugee households from 4 settlements depicted in Figure 1.[13] The design of the program followed a phase-in approach, meaning that all households receiving cash food assistance were also made eligible for the FLT, but at different times. The WFP divided the settlements into several smaller areas referred to as "zones". Subsequently, the implementing partner randomly selected which zones would be assigned to treatment, and then selected control zones that were located at a distance within the settlement to ensure that the training schedule would not clash with the monitoring activities. As a result, assignment probabilities varied among zones and depended on factors such as zone proximity and settlement size. This process resulted in 24 zones (approximately 37% of eligible households) being assigned to start the training immediately and 19 zones (approximately 63% of eligible households) being put on a waiting list. A baseline survey was conducted prior to the FLT, and an endline survey was conducted approximately 12 weeks after the completion of the training. Due to the staggered implementation of the program and the simultaneous measurement of treatment and control households, participants who were enrolled in the program but still waiting to receive the FLT served as the control group for those who had already completed the training.

[13] The participants were from Adjumani, Kyaka, Kyangwali, and Nakivale, and were selected by location and matched with the UNHCR registry of refugees to confirm their refugee status and to allow identification in the follow-up survey.

Figure 1: Refugee settlements in Uganda



<u>Notes</u>: Settlements identified by the black box are the ones selected for the experiment. The numbers included in the boxes indicate the total number of refugees in the corresponding settlement. Our elaboration based on data from UNHCR (2021b).

Several additional points about the design are worth mentioning. Firstly, to account for varying treatment assignment probabilities of zones, we control for the distance of the zones within the settlement and settlement fixed effects in all of our specifications. These covariates are treated as stratification variables. Settlement fixed effects also account for specific characteristics related to the fact that refugees in Adjumani, Kyangwali, Kyaka II, and Nakivale have been in the country for varying lengths of time and come from different countries of origin. Secondly, the results of the balancing tests reveal that the allocation process of zones to treatment and control statuses was imperfect. To mitigate the concerns that these differences may not be due to chance, we employ difference-in-difference (DID) strategies and control for household fixed effects. Thirdly, during the program's implementation, approximately 35% of households did not comply with their assignments. Some households assigned to the program did not participate, while others not initially assigned to it ended up taking part. This led to a "fuzzy" design which is incorporated into our empirical analysis. Fourthly, the follow-up survey for settlements in the North was delayed by approximately 20 weeks due to COVID-19-related shutdowns. This delay affected only one settlement, which we account for using settlement fixed effects.[14]

[14] The attrition rate for treated units stands at 7.2% while for control units at 5.2%, and the difference in characteristics are not statistically significant.

Lastly, our results remain valid assuming no interference between zones, an assumption that is unlikely to be violated given the short-term nature of the intervention.

2.2 The FLT program

The program is a 5-week course consisting of 20 hours, with sessions held twice a week for two hours.[15] Notably, when compared to the average intensity of financial literacy training interventions analyzed in the literature, which stands at 11.7 hours, our program qualifies as a "high-intensity" intervention (Kaiser et al., 2022). Its primary aim was to empower refugees in the use of DFS. It provided basic financial literacy training and focused on key aspects, such as raising awareness about digital financial technologies, effective savings strategies, and financial management techniques, as well as developing and implementing personal financial plans. The FLT followed a learner-centered approach, which previous research conducted in Uganda has demonstrated to be more effective compared to a lecture-centered approach (Kaiser and Menkhoff, 2022). The learner-centered approach is intended to offer problem-based learning experiences and foster discussions among participants. The exercises and materials are created to actively involve participants in the topic, such as through budgeting activities or organizing different savings options based on their safety and return characteristics. Moreover, participants are encouraged to share their personal experiences, while the trainer functions as a learning facilitator.[16]

2.3 Sample and outcomes

Columns (1)-(3) of Table 1 present baseline mean values of pre-treatment characteristics and outcome variables for our entire sample, which is the one used to estimate the main results of the paper. In Table A2 and A3 of the Appendix we report the same statistics for the two estimation samples employed in the empirical analysis that will follow.[17]

On average, the participants are 38 years old, with 60% being female household heads. The education level is particularly low, with an average of less than 3 years of schooling. More than half of the participants (52%) received no education, while only 15% obtained education beyond primary school. The average household size is almost 6 people, and there is at least one child under the age of 5 in each family. In terms of economic status, 35% of participants reported that no household member earns any form of income, 85% reported experiencing periods without cash income several times, and 65% reported rarely having an emergency fund to cover unplanned expenses. Finally, nearly half of the sample reported difficulties connecting to the local network.

Compared to UNHCR data, our sample is representative of refugees in Uganda in terms of settlements considered, household composition, households' main sources of income and debt, and general economic condition. Our sample's characteristics match other studies in the financial literacy literature in terms of female representation, age, household size, and number of children (e.g., Drexler et al., 2014; Abarcar et al., 2020; Horn et al., 2020; Kaiser and Menkhoff, 2022). However, there are also notable differences, particularly in the lower level of education and economic well-being. Therefore, compared to the previous literature, our data paints a picture of a highly disadvantaged population.[18]
Next, following Horn et al. (2020), we created 4 indexes of decision inputs covered in the financial education curriculum: financial knowledge, planning, agency, and trust. The financial knowledge index is the mean of 9 binary

This suggests that attrition is not a significant factor in potential bias.

[15] Each trainer managed four groups of ten participants, with a maximum of two participants per household. Attendees were given only an exercise book and a pen and received certificates of participation upon completion.

[16] This is different from the lecture-centered approach which instead is organized as an exposition-centered community lecture where the concepts are transmitted through frontal explanations and demonstrations, with the trainer being more of a lecturer than a facilitator.

[17] For the estimation of our main and preferred empirical model, we employ the entire sample. For our secondary empirical specification, we have to exclude from the estimation sample those participants who were already treated at baseline (103 units), see Section 3 for further explanations. Of the two samples we considered for estimation, the stricter one has a higher level of imbalance.

[18] For example, Drexler et al. (2014) reported that 35% of participants completed high school and continued their studies, while Abarcar et al. (2020) reported an average of 15 years of education. Studies by Horn et al. (2020) and Kaiser and Menkhoff (2022) focused on Ugandan participants. In the former case, household's heads reported an average of 10 years of education, a mean total income of approximately 140,000 Ugandan Shilling (~38$), and 37% of their sample owned a formal bank account, with 29% reporting frequent usage. In the latter case, the authors studied a sample where 28% of participants continued their education beyond primary school, and the reported average monthly income was around 220,000 Ugandan Shilling (~59$).

indicators that measure understanding of basic financial concepts such as budgeting, financial goals, savings, and financial products and services.[19] The baseline average for this index is 0.59.

Table 1: Baseline mean values and balancing (full sample)

	(1)	(2)	(3)	(4)	(5)
	Sample	*Control*	*Treatment*		
		Mean		p-value $H_0: T = C$	Normalized difference
Observables:					
Female	0.88	0.88	0.88	0.99	0.00
Age	38.45	38.17	38.93	0.10	0.01
Household head	0.82	0.79	0.86	0.43	0.03
Household head female	0.60	0.56	0.66	0.18	0.03
Years of education	2.89	3.02	2.68	0.55	0.01
Household head Education (> primary)	0.15	0.15	0.15	0.87	0.00
Household head no education (0 yrs)	0.52	0.50	0.55	0.41	0.01
N. of children 0-5	1.28	1.30	1.23	0.01	0.01
Household size	5.84	5.68	6.11	0.10	0.02
No income earners	0.35	0.28	0.46	0.05	0.06
Gone without cash income sometimes or often	0.85	0.87	0.83	0.22	0.02
Rarely or never has emergency fund	0.65	0.66	0.63	0.10	0.01
Bad connection experience	0.49	0.53	0.45	0.26	0.04
Key decision inputs:					
Financial knowledge index [0;1]	0.59	0.63	0.53	0.56	0.05
Financial planning index [0;1]	0.50	0.50	0.51	0.04	0.00
Financial agency index [0;1]	0.29	0.28	0.30	0.12	0.01
Financial trust index [0;1]	0.73	0.72	0.74	0.59	0.02
Downstream financial behaviors and DFS adoption:					
Budget use [0;1]	0.54	0.47	0.65	0.01	0.06
Saving goal [0;1]	0.48	0.47	0.50	0.11	0.01
Uses mobile phone [0;1]	0.68	0.68	0.69	0.28	0.00
Registered bank account	0.18	0.27	0.04	0.04	0.09
Saving formally [0;1]	0.04	0.06	0.00	0.10	0.04
Weekly DFS transactions [0;21]	0.22	0.18	0.29	0.02	0.03
Monthly DFS transactions [0;48]	0.83	0.82	0.86	0.31	0.00
Weekly E-wallet transactions [0;27]	0.07	0.05	0.09	0.40	0.02
Monthly E-wallet transactions [0;36]	0.34	0.30	0.41	0.28	0.01
Observations	*1,649*	*1,042*	*607*		

Notes: The table presents baseline mean values and balance tests over treatment assignment groups of pre-treatment outcome values and households' characteristics. We consider here the full sample. Columns (1) to (3) show the average values for the full sample, the treatment group, and the control group, respectively. Column (4) displays the results of a t-test for the impact of the treatment on each variable listed on the left side. Column (5) displays the normalized difference between treatment and control for each variable listed on the left side. Both the p-values and the normalized differences are calculated from a regression model that takes into account the treatment assignment and stratification variables performing the STATA command iebaltab. The standard errors are clustered at the zone level. The statistics reported are in absolute values.

[19] The knowledge of refugees is assessed through questions such as "Do you understand/know what a budget is?" and "Do you know how to create a

The financial planning index is the average of 9 binary indicators that measure habits and planning related to expenses, record-keeping, and budgeting.[20] The baseline average for this index is 0.50. The financial agency index is the mean of 4 indicators that assess the level of responsibility and decision-making power of respondents in household financial matters, with a mean value of 0.29. Finally, the financial trust index is based on 7 scores that measure the level of trust households have in various financial institutions, with a baseline average of 0.73. In Column (7) of Tables A4 to A7 of the Appendix, we report the individual values of each component that comprise the indices.

Finally, regarding other outcomes related to financial inclusion, we consider several measures of down- stream behavior and DFS adoption. Nearly 70% reported making use of mobile phone, which is a promising starting point for our experiment. However, in terms of engagement with traditional financial institutions, our data show that only 18% of respondents own a bank account. In terms of saving behaviors, only 4% saved through formal channels, 48% of respondents reported having a saving goal, and 54% consistently used budgets. These statistics suggest limited engagement with formal financial services, which is further supported by the fact that the vast majority of respondents reported zero transactions via digital financial services (DFS) or electronic wallets.

2.4 Balancing checks

In Table 1, we present the results of balancing checks for the entire sample to assess the comparability of our treatment assignment groups. We use p-values from t-tests and normalized differences. The results suggest that our sample exhibits some imbalance. Notably, some of the variables that are un- balanced, such as the higher proportion of individuals without income or unregistered bank accounts in the treatment group, are important for our analysis. As mentioned earlier, to address potential selection bias stemming from these differences, we employ DID strategies in our subsequent analysis, and in our preferred specification we also include household fixed effects, which absorb time-invariance unbalance of covariates[21]. While it is worth noting that the DID strategy may slightly reduce precision, in contexts of imperfect randomization the presence of a baseline helps controlling for unobserved, time-invariant factors that could affect the outcome variables (e.g., McKenzie, 2012, page 211, footnote 2).

3 Empirical framework

Building upon the insights of Batista and Vicente (2013, 2020), our study seeks to empirically test the hypothesis that providing refugees with both financial education and access to safe and affordable DFS can induce a substitution effect of money management practices. Specifically, we expect a transition from traditional informal methods, which are often more costly, risky, and time-consuming, to the adoption of formal digital financial technologies. To rigorously test this hypothesis, we employ both "sharp" and "fuzzy" difference-in-difference (DID) models. To ensure that program eligibility is plausibly random, we control for the distance of the zones within the settlement, settlement fixed effects and household fixed effects. To estimate the sharp DID model, we have to exclude participants who result already treated at baseline (103 units)[22].

Sharp difference-in-difference. The basic model we estimate is as follows:

$$Y_{i,j,t} = \alpha + \beta_1 G_{ij} + \beta_2 T_t + \delta DI\, D_{i,j,t} + \boldsymbol{Rule'_{ij}}\boldsymbol{\gamma} + \varepsilon_{i,j,t} \quad (1)$$

[20] Some of these indicators were originally discrete, but we converted them all to binary to create an index that takes values between 0 and 1. Survey questions used to create these binary indicators include "Do you make financial plans ahead regarding money?" and "How often do you create a budget for how to spend your income, whether it is earned through a job, received from the government, or from other people?".

[21] Controlling for household fixed effects in a DID regression is a more conservative and robust strategy than merely adjusting for individual characteristics, as it accounts for all potential individual-level unobserved, time-invariant factors that could bias the results.

[22] Please refer to Table A8 on this point.

where $Y_{i,j,t}$ represents the outcome of household i in settlement j at period t. T_t and G_{ij} represent dummy variables for the time period and treatment assignment, respectively. $DID_{i,j,t}$ is the difference-in- difference indicator obtained as the product of time and treatment assignment dummies. $\boldsymbol{Rule}_{ij}$ includes the distance of the zones within the settlement and a set of settlement dummies. $\varepsilon_{i,j,t}$ is the idiosyncratic error term. We also re-estimate the model by including a set of household fixed effects α_i, which captures time-invariant characteristics of households—both observable and unobservable—that might influence the outcome. In our context, the parameter δ captures the Intention-To-Treat (ITT) effect.

Fuzzy difference-in-difference.

The approach by de Chaisemartin and D'HaultfŒuille (2017) is suitable when no group experiences a sharp change in treatment and no group remains completely untreated. In such fuzzy designs, the DID indicator is not equal to the product of the time and group indicators (i.e., $DID \neq G \cdot T$). A more appropriate estimator of treatment effects for this case is the Wald (or "fuzzy") DID, which is the DID of the outcome divided by the DID of the treatment. The fuzzy DID estimator operates similarly to an Instrumental Variables (IV) estimation. In the fuzzy DID framework, we distinguish between treatment status D and treatment assignment Z. See section A.2 in the Appendix for further details. Individuals who receive the treatment in the second period are referred to as "switchers". For switchers, the Wald DID estimator targets the Local Average Treatment Effect (LATE):

$$\Delta = E(Y_{11}(1) - Y_{11}(0)|S).$$

where $Y_{11}(1) \equiv \{Y(1)|G = 1, T = 1\}$ is the potential outcome for switchers who received treatment at endline and $Y_{11}(0) \equiv \{Y(0)|G = 1, T = 1\}$ is the potential outcome if they do not receive treatment at endline.[23]

The Wald DID estimator identifies the LATE if two conditions are met. Firstly, the treatment's impact remains constant over time. Secondly, if the treatment's intensity increases in both the treatment and control groups, then the treatment's effect is the same in both groups. However, these conditions may be difficult to justify in the data. An alternative approach provided by the same authors is to use the time-corrected (TC) Wald ratio, which does not rely on any assumptions regarding treatment effects and can be applied when the proportion of treated units is stable in the control group. The TC-DID method assumes common trends within subgroups of units sharing the same treatment at the first date. Appendix A.2 provides more details on these fuzzy DID methods, including its underlying assumptions and why they are likely satisfied in our context.

4 Results

We present results for Sharp DID, Wald-DID, and TC-DID estimators. For each model, we provide estimates from both the basic specification and an augmented version of Equation (1), which includes household fixed effects. All output variables are standardized before estimation; thus, treatment effect parameters should be interpreted in terms of standard deviation (SD) units. Standard errors are clustered at the zone level, which is the level of assignment. To account for multiple hypothesis testing, we compute False Discovery Rate (FDR)-adjusted p-values for each outcome (Benjamini et al., 2006).

4.1 Impact on decision inputs

Table 2 reports treatment effect parameters for each of the four indexes.[24] Coefficients on financial knowledge, planning, and agency are positive and statistically significant at the 1% level in all "fuzzy" models. The magnitude

[23] It is worth noting that households with negative treatment rate between the two periods are excluded from the analysis as it likely indicates errors in the data collection. As shown in Table A9, including this sample of participants as a robustness check does not alter the results.

[24] Appendix Tables A4 to A7 report results separately for each index component.

of the coefficients tends to be larger for the “fuzzy” DID models compared to the sharp DID model, which implies that the ITT parameter underestimates the effect of the program likely due to compliance issues with the assignment variable. Our preferred estimates are obtained from the TC-DID estimator with household fixed effects because it relies on weaker assumptions. This approach shows that the treatment effect is 0.65 SD for financial knowledge, 0.74 SD for planning, and 0.62 SD for agency. We find smaller and statistically insignificant treatment effects on financial trust. This result suggests that confidence in financial institutions adjusts more slowly than knowledge, planning, or agency, and therefore it may require a longer intervention period to show significant effects.

Two key observations emerge from these initial results. Firstly, when compared to the mean effects reported in a recent meta-analysis by Kaiser et al. (2022), our study’s estimated effects appear both larger and less precise. Specifically, our effect sizes and standard errors correspond to the lower-right section of Figure 2 in their publication (page 262). To explain these results, we need to consider two important factors: (i) the studies included in the meta-analysis had twice the average measurement period compared to ours (6 months vs. 3 months), which could account for the larger short-term impact on decision inputs; (ii) the unique context of the refugee settlement camps likely had a substantial impact on both the magnitude and precision of our results. Indeed, our setting presents several challenges, such as uneven network coverage, low educational attainment, and severe economic conditions. These contextual elements may have introduced noise and variability into our results, making it more challenging to discern a definitive treatment effect. In our concluding remarks, we provide power calculations to further clarify these aspects.

Secondly, we consistently observed significant effects across the three indexes analyzed. Importantly, the inclusion of household fixed effects in the TC-DID models slightly reduced treatment effects but did not affect their significance, indicating that omitted variable bias is unlikely to significantly impact our analysis.

Table 2: Decision inputs covered in the financial education curriculum

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Sharp-DID		Wald-DID		Tc-DID		Mean Dep. Var.
Indexes:							
Financial knowledge	0.632***	0.654**	0.805***	0.638***	0.888***	0.652***	0.595
	(0.212)	(0.313)	(0.147)	(0.191)	(0.152)	(0.205)	
	[0.012]	[0.172]	[0.001]	[0.002]	[0.001]	[0.002]	
Financial planning	0.480**	0.501	0.737***	0.703***	0.842***	0.744***	0.503
	(0.210)	(0.313)	(0.165)	(0.183)	(0.146)	(0.190)	
	[0.035]	[0.181]	[0.001]	[0.001]	[0.001]	[0.001]	
Financial agency	0.412*	0.459	0.693***	0.571***	0.839***	0.617***	0.285
	(0.217)	(0.321)	(0.161)	(0.175)	(0.140)	(0.188)	
	[0.047]	[0.181]	[0.001]	[0.002]	[0.001]	[0.002]	
Financial trust	−0.178	−0.176	0.056	0.126	−0.115	0.126	0.726
	(0.257)	(0.378)	(0.169)	(0.199)	(0.173)	(0.217)	
	[0.140]	[0.256]	[0.228]	[0.152]	[0.146]	[0.164]	
Household FE	✗	✓	✗	✓	✗	✓	
Observations	*2,996*	*2,996*	*3,200*	*3,200*	*3,200*	*3,200*	

*<u>Notes</u>: Dependent variables: standardized composite indexes. Coefficients in column (1)-(2) were estimated using an OLS regression of dependent variables listed in the row headings on a dummy for treatment assignment, a dummy for time and an interaction between the two. Columns (3)-(4) and (5)-(6) report treatment effect estimates respectively from fuzzy DID regression using Wald-DID estimator and using Tc-DID estimator. Each of the variables in the row headings is regressed on a dummy for treatment status, a dummy for the group and one for time. All DID specifications are conducted with clustering at the unit of randomization (zones) and controlling for the distance of zones within the settlement as well as for settlement fixed effects. These controls are considered stratification variables because they help account for differences in treatment assignment probabilities across zones, which arose due to imperfections in the randomization process. Standard errors in (parentheses). FDR adjusted p-values in [square brackets], calculated using the two-stage procedure in Benjamini et al. (2006). Tables A4 to A7 report results separately for each index component. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.*

4.2 Impact on downstream financial behaviors and DFS adoption

The results from Table 3 demonstrate a significant positive impact of the treatment on individuals' likelihood of budget usage (0.37 SD) and of setting saving goals (0.59 SD) as revealed by the TC-DID in column (6). These results are in line with those documented in existing literature, although they are larger and somewhat less precisely estimated.[25] We also observe suggestive, marginally significant, effects for mobile phone usage (0.20 SD), which provides evidence for refugees willingness to take advantage of digital services.

[25] Among the set of financial behaviors considered in the meta-analysis by Kaiser et al. (2022), saving and budgeting showed the largest effects, ranging from 0.06-0.10 SD and 0.09-0.15 SD, respectively.

Table 3: Downstream financial behaviors and DFS adoption

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Sharp-DID		Wald-DID		Tc-DID		Mean Dep. Var.
Budget use	−0.103	−0.059	0.284**	0.376***	0.243	0.371**	0.537
	(0.195)	(0.285)	(0.134)	(0.140)	(0.150)	(0.152)	
	[0.358]	[1.000]	[0.121]	[0.020]	[0.489]	[0.033]	
Saving goal	0.230	0.278	0.715***	0.566***	0.873***	0.589***	0.482
	(0.164)	(0.235)	(0.168)	(0.163)	(0.150)	(0.174)	
	[0.468]	[1.000]	[0.121]	[0.020]	[0.339]	[0.038]	
Uses mobile phone	−0.057	−0.063	0.170*	0.181*	0.219*	0.198*	0.683
	(0.182)	(0.260)	(0.096)	(0.101)	(0.115)	(0.105)	
	[0.358]	[1.000]	[0.121]	[0.020]	[0.489]	[0.033]	
Registered bank account	−0.323*	−0.305	0.210	0.395	0.075	0.421	0.185
	(0.185)	(0.267)	(0.102)	(0.146)	(0.123)	(0.158)	
	[0.358]	[1.000]	[0.199]	[0.078]	[0.339]	[0.111]	
Saving formally	0.369	0.351	0.134	−0.010	0.177	−0.023	0.042
	(.09)	(.129)	(0.152)	(0.125)	(0.164)	(0.128)	
	[0.505]	[1.000]	[0.138]	[0.078]	[0.290]	[0.078]	
Weekly DFS transactions	0.198	0.191	0.110	0.303*	−0.055	0.237	0.220
	(0.317)	(0.458)	(0.229)	(0.178)	(0.235)	(0.183)	
	[0.001]	[0.063]	[0.311]	[0.144]	[0.405]	[0.284]	
Monthly DFS transactions	0.355	0.342	0.189	0.294	0.120	0.241	0.834
	(0.293)	(0.423)	(0.180)	(0.185)	(0.192)	(0.181)	
	[0.468]	[1.000]	[0.331]	[0.083]	[0.568]	[0.122]	
Weekly E-wallet transactions	1.219	1.209	0.872	1.146**	0.745	1.007*	0.069
	(0.903)	(1.308)	(0.580)	(0.558)	(0.745)	(1.007)	
	[0.422]	[1.000]	[0.266]	[0.092]	[0.489]	[0.122]	
Monthly E-wallet transactions	1.007	0.985	0.582	0.721*	0.595	0.571*	0.340
	(0.610)	(0.892)	(0.420)	(0.388)	(0.391)	(0.347)	
	[0.396]	[1.000]	[0.190]	[0.064]	[0.343]	[0.078]	
Household FE	✗	✓	✗	✓	✗	✓	
Observations	*2,996*	*2,996*	*3,200*	*3,200*	*3,200*	*3,200*	

*Notes: Dependent variables: standardized variables reported in the table. Coefficients in column (1)-(2) were estimated using an OLS regression of dependent variables listed in the row headings on a dummy for treatment assignment, a dummy for time and an interaction between the two. Columns (3)-(4) and (5)-(6) report treatment effect estimates respectively from fuzzy DID regression using Wald-DID estimator and using Tc-DID estimator. Each of the variables in the row headings is regressed on a dummy for treatment status, a dummy for the group and one for time. All DID specifications are conducted with clustering at the unit of randomization (zones) and controlling for the distance of zones within the settlement as well as for settlement fixed effects. These controls are considered stratification variables because they help account for differences in treatment assignment probabilities across zones, which arose due to imperfections in the randomization process. Standard errors in (parentheses). FDR adjusted p-values in [square brackets], calculated using the two-stage procedure in Benjamini et al. (2006). Tables A4 to A7 report results separately for each index component. *** p < 0.01, ** p < 0.05, * p < 0.1.*

In contrast to the findings regarding decision inputs, the relatively brief exposure to the program likely hindered the potential effectiveness of the training on engaging in DFS. In fact, we did not detect any significant impacts on the percentage of participants opening a bank account, engaging in formal savings, or conducting digital transactions through e-wallets or bank accounts. Similarly, Carpena et al. (2011) found that financial literacy positively influenced participants' awareness and attitudes towards financial products and financial planning but did not improve financial abilities.

4.3 Further analysis

Thus far, we have emphasized that high-intensity programs may not significantly affect individuals' use of DFS among marginalized communities within limited timeframes. In this section, we delve further into understanding the likely reasons behind the lack of noticeable changes in adoption resulting from the intervention.

Demand side: financial trust. From a demand-side perspective, FLT programs have the potential to improve financial inclusion by enhancing households' trust in financial instruments and services. However, the empirical analysis presented in Section 4 indicates that the training did not have any positive effects on the overall level of financial trust among participants. Table A7 provides a detailed breakdown of treatment effects on each component of the financial trust index, revealing no significant impacts along any dimension of DFS. These findings suggest that, when analyzing a population with little to no prior exposure to formal financial technologies, simply increasing their knowledge may not be enough to instill confidence in financial institutions. Further consideration should be given to the duration and intensity of the FLT, as it may not have been sufficient for refugees to build their trust levels.

Supply side: financial support. From a supply-side perspective, the effectiveness of a financial education program may be constrained among a population with low levels of education, particularly if they do not have access to high-quality financial support from financial professionals alongside the FLT lectures and especially once they are over. A financial professional provides expertise for clients' decisions around money matters, personal finances, and investments. For example bank agents, informal savings group staff, financial institution agents, stockbrokers, insurance agents, tax preparers, investment managers, and financial planners. From our data it emerges that the support offered to refugees from financial professionals seems to have failed in its purpose: at endline, a significant proportion of the sample (88%) reported a lack of understanding of the advice received, 84% expressed not receiving the necessary advice, 86% faced difficulties in comprehending the advisers' recommendations, and 93% reported unprofessional behaviour. Thus, despite the positive effects on financial knowledge, the adoption and use of digital technologies were likely to be hindered by the insufficient financial support.

5 Discussion and conclusion

The key conclusion is that a high-intensity FLT program can bolster the financial knowledge of refugees. However, it is essential to recognize that these programs may not produce equally substantial improvements in the rate of adoption of DFS, particularly when constrained by a short span of time. It is highly likely that changes such as greater adoption of digital transactions or increased rates of bank account registration, require more time to manifest and be accurately measured. These findings are potentially noteworthy for policymakers, as refugees can make substantial contributions to the economies of their host countries. We remain mindful of the limitations inherent in our experimental design, and we caution against drawing definitive conclusions regarding the long-term impact or cost-effectiveness of our interventions for marginalized communities.

To conclude this section, we engage in a thorough critique of our research design, employing ex-post power calculations to identify areas where enhancements to the design could be made. Table 4 summarizes the results of this investigation. In column (1) and (2), we provide the mean and standard deviation for each output variable. Column (3) showcases the coefficient of variation (CV), computed by dividing the standard deviation by the mean. This metric serves as a valuable indicator of the cross-sectional het- erogeneity within the refugee population. The greater the CV, the more diverse the sample, making it inherently more challenging to discern changes in average outcomes resulting from our treatment. The analysis reveals that, among the 9 economic outcomes examined, 6 exhibit a standard deviation that is twice the mean (CV > 2). This initial assessment highlights the need for future studies to refine our approach, specifically by considering a narrower mix of individuals when forming training classes.

Columns (4) and (5) shed light on additional design characteristics. The ICC (intra-cluster correlation) coefficient values fall within the range typically observed in the economics education literature, indicating a consistency with established standards. In contrast, the autocorrelation coefficient assesses the correlation between baseline measurements and the outcome. It is noteworthy that, with the exception of "registered bank account", the autocorrelation coefficient for all other measures is very low. This observation suggests that collecting multiple measurements of the outcomes, instead of relying solely on single follow-up measures, may be a more effective strategy to enhance statistical power for these economic outcomes.

Columns (6) and (7) present the estimated effect size and standard error of the program for each output variable. Given our assumption of 80% power and a 5% rejection rate, we can then calculate the ex-post Minimal Detectable Effect (MDE), with the results presented in column (8).[26] In our specific context, most MDEs fall within the range of 0.30-0.60 standard deviations (SD). In contrast, our estimated effect sizes in column (6) tend to be below these thresholds. In future studies, one can enhance the precision of the estimates by randomizing at the individual level or by increasing the number of clusters per treatment arm.

Finally, in columns (9)-(11), we present the results of a distinct analysis, inspired by the approach taken in McKenzie and Woodruff (2013). In this analysis, we utilize the actual observed data from the control group to retrospectively estimate the statistical power for detecting effect sizes of potential theoretical or policy significance.[27] An interesting result emerges. For each of the economic outcomes that held the highest importance for our donor—specifically, registered bank account activity, weekly/monthly DFS transactions, and weekly/monthly E-wallet transactions—detecting a significant effect would be quite challenging, even if the effect were substantial in magnitude. This implies that, when focusing on these variables, we would require a significantly larger sample size or an increased number of clusters for significant findings. Alternatively, future research could consider relaxing the stringent time frame constraints that guided our investigation and explore the impact on the same target population over an extended period, which would likely increase the effect size of the program. Eventually, we are able to provide evidence that donors and implementing organizations should take statistical power to detect effects on focal outcomes as a first-order consideration when designing and/or implementing quantitative evaluations of their programs.

[26] It is important to note that this calculation relies on the estimated standard error, not the true (population or expected) standard error, making this number somewhat imprecise. Nevertheless, it provides a valuable estimate of the MDE, which is crucial because it informs us about the effect sizes needed to meet cost-benefit thresholds or attract policy interest within a given experimental design.

[27] To clarify, we employ the actual sample size, the number of clusters, control group means, control group standard deviations, control ICC coefficients, and control group autocorrelation in output measures to compute the statistical power for detecting a 10%, 30%, or 50% increase in each of the output variables examined in this paper.

Table 4: Ex-post power analysis

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
	Summary statistics increase of:					Effect size			Power to detect:		
	Mean	SD	CV	ICC	Autoc.	SD (%)	SE	MDE	10%	30%	50%
						Decision inputs:					
Financial knowledge	0.59	0.33	0.56	0.43	0.41	0.65	0.20	0.57	0.16	0.83	1.00
Financial planning	0.50	0.32	0.64	0.28	0.29	0.74	0.19	0.53	0.17	0.84	1.00
Financial agency	0.29	0.18	0.62	0.22	0.16	0.62	0.19	0.53	0.20	0.91	1.00
Financial trust	0.72	0.26	0.36	0.19	0.02	0.13	0.22	0.61	0.51	1.00	1.00
						Downstream financial behaviors and DFS adoption:					
Budget use	0.54	0.50	0.93	0.20	0.10	0.37	0.15	0.43	0.12	0.63	0.97
Saving goal	0.48	0.50	1.04	0.16	0.13	0.59	0.17	0.49	0.11	0.62	0.96
Uses mobile phone	0.68	0.47	0.69	0.16	0.13	0.20	0.10	0.28	0.26	0.97	1.00
Registered bank account	0.18	0.39	2.17	0.32	0.30	0.42	0.16	0.44	0.05	0.13	0.28
Saving formally	0.04	0.20	5.00	0.13	0.02	-0.02	0.13	0.36	0.04	0.07	0.13
Weekly DFS transactions	0.22	0.69	3.14	0.10	0.04	0.24	0.18	0.51	0.05	0.14	0.32
Monthly DFS transactions	0.83	2.12	2.55	0.13	0.10	0.24	0.18	0.51	0.05	0.17	0.38
Weekly E-wallet transactions	0.07	0.35	5.00	0.27	0.06	1.01	1.01	2.82	0.03	0.06	0.09
Monthly E-wallet transactions	0.34	1.31	3.85	0.29	0.08	0.57	0.35	0.97	0.04	0.07	0.12

Notes: In Columns (1)-(4), we present the baseline mean, standard deviation, coefficient of variation (CV), and intra-cluster correlation (ICC) coefficient for each variable on the left-hand side. Column (5) provides the autocorrelation coefficient between baseline and endline data, exclusively for the control group. Columns (6)-(7) offer the effect size (expressed as a percentage of a standard deviation) and the corresponding standard error for the preferred specification outlined in Tables 2 and 3. In Column (8), we present the ex-post minimum detectable effect (MDE) using this information. For power calculations in Columns (9)-(11), we assume the presence of one baseline and one post-treatment survey, with 23 clusters per treatment arm and an average of 37 participants per cluster, alongside the remaining details reported in Columns (4) and (5).

References

ABARCAR, P., R. BARUA, AND D. YANG (2020): "Financial education and financial access for transnational households: Field experimental evidence from the Philippines," *Economic Development and Cultural Change*, 69, 373–404. [9]

ALTINDAĞ, O. AND S. D. O'CONNELL (2023): "The short-lived effects of unconditional cash transfers to refugees," *Journal of Development Economics*, 160, 102942. [2], [4]

ATHEY, S. AND G. W. IMBENS (2006): "Identification and Inference in Nonlinear Difference-in-Differences Models," *Econometrica*, 74, 431–497. [27]

AYGÜN, A. H., M. G. KIRDAR, M. KOYUNCU, AND Q. STOEFFLER (2021): "Keeping Refugee Children in School and Out of Work: Evidence from the World's Largest Humanitarian Cash Transfer Program," Tech. Rep. 14513. [4]

BATISTA, C. AND P. C. VICENTE (2013): "Introducing mobile money in rural Mozambique: Evidence from a field experiment," *NOVAFRICA Working Paper Series*. [3], [5], [11]

——— (2020): "Adopting mobile money: Evidence from an experiment in rural Africa," *AEA Papers and Proceedings*. [3], [5], [11]

BENJAMINI, Y., A. M. KRIEGER, AND D. YEKUTIELI (2006): "Adaptive linear step-up procedures that control the false discovery rate," *Biometrika*, 93. [13], [14], [15], [28], [29], [30], [31], [32], [33]

BETTS, A., I. CHAARA, N. OMATA, AND O. STERCK (2019): "Refugee economies in Uganda: What difference does the self-reliance model make?" Tech. rep., Oxford Department of International Development. [5]

BREZA, E., M. KANZ, AND L. F. KLAPPER (2020): "Learning to navigate a new financial technology: Evidence from payroll accounts," *National Bureau of Economic Research*. [5]

BRUHN, M., L. D. S. LEÃO, A. LEGOVINI, R. MARCHETTI, AND B. ZIA (2016): "The Impact of High School Financial Education: Evidence from a Large-Scale Evaluation in Brazil," *American Economic Journal: Applied Economics*, 8, 256–95. [3]

CARPENA, F., S. A. COLE, J. SHAPIRO, AND B. ZIA (2011): "Unpacking the causal chain of financial literacy," *World Bank Policy Research Working Paper*. [16]

CHAABAN, J., H. GHATTAS, N. SALTI, W. MOUSSA, A. IRANI, Z. JAMALUDDINE, AND R. AL MOKDAD (2020): "Multi-Purpose Cash Assistance in Lebanon: Impact Evaluation on the Well-Being of Syrian Refugees," Tech. rep., World Food Programme, Beirut, Lebanon. [4]

CHOI, J., D. LAIBSON, AND B. MADRIAN (2010): "Why Does the Law of One Price Fail? An Experiment on Index Mutual Funds," *The review of financial studies*, 23, 1405–1432. [3]

COSGROVE, J., N. CRAWFORD, AND I. MOSEL (2016): "10 things to know about refugees and displacement," Tech. rep., ODI: London. [2]

DE CHAISEMARTIN, C. AND X. D'HAULTFŒUILLE (2017): "Fuzzy Differences-in-Differences," *The Review of Economic Studies*, 85, 999–1028. [3], [12], [25], [27]

DREXLER, A., G. FISCHER, AND A. SCHOAR (2014): "Keeping it simple: Financial literacy and rules of thumb," *American Economic Journal: Applied Economics*, 6, 1–31. [9]

ECHO (2024): "Forced displacement: Refugees, asylum-seekers, and internally displaced persons (IDPs)," Tech. rep., European Civil Protection and Humanitarian Aid Operations. [2], [5]

FIALA, N. (2015): "Economic Consequences of Forced Displacement," *The Journal of Development Studies*, 51, 1275–1293. [4]

HIDROBO, M., J. HODDINOTT, A. PETERMAN, A. MARGOLIES, AND V. MOREIRA (2014): "Cash, food, or vouchers? Evidence from a randomized experiment in northern Ecuador," *Journal of Development Eco- nomics*, 107, 144–156. [4]

HORN, S., J. JAMISON, D. KARLAN, AND J. ZINMAN (2020): "Does lasting behavior change require knowledge change? Evidence from savings interventions for young adults," Tech. rep., National Bureau of Economic Research. [9], [26]

KAISER, T., A. LUSARDI, L. MENKHOFF, AND C. URBAN (2022): "Financial education affects financial knowledge and downstream behaviors," *Journal of Financial Economics*, 145, 255–272. [3], [4], [5], [8], [13], [15]

KAISER, T. AND L. MENKHOFF (2022): "Active learning improves financial education: Experimental evidence from Uganda," *Journal of Development Economics*, 157, 102870. [8], [9]

LUSARDI, A. AND O. S. MITCHELL (2014): "The Economic Importance of Financial Literacy: Theory and Evidence," *Journal of Economic Literature*, 52, 5–44. [2]

MACPHERSON, C. AND O. STERCK (2021): "Empowering refugees through cash and agriculture: A regression discontinuity design," *Journal of Development Economics*, 149, 102614. [2], [4]

MCKENZIE, D. (2012): "Beyond baseline and follow-up: The case for more T in experiments," *Journal of Development Economics*, 99, 210–221. [3], [11]

MCKENZIE, D. AND C. WOODRUFF (2013): "What Are We Learning from Business Training and Entrepreneurship Evaluations around the Developing World?" *The World Bank Research Observer*, 29, 48–82. [18]

OECD (2021): "PISA 2021 Financial Literacy Framework,". [2]

QUATTROCHI, J., G. BISIMWA, P. VAN DER WINDT, AND M. VOORS (2022): "Cash-like vouchers improve psychological well-being of vulnerable and displaced persons fleeing armed conflict," *PNAS Nexus*, 1. [4]

SCHUETTLER, K. AND Q.-T. DO (2023): "Outcomes for Internally Displaced Persons and Refugees in Low and Middle-Income Countries," Tech. rep., The World Bank. [2], [4]

UNHCR (2018): "UNHCR Global Compact on Refugees," Tech. rep., The UN Refugee Agency. [2]

——— (2021a): "Financial Service Provider (FSP) Assessment report for Cash and Voucher Assistance in Uganda," Tech. rep., UNHCR. [6]

——— (2021b): "UNHCR Global Trends 2021: Forced Displacement in 2021," Tech. rep., The UN Refugee Agency. [5], [7]

WFP (2022a): "Digital Financial Inclusion through Cash Transfers," Tech. rep., World Food Programme. [5]

——— (2022b): "Prioritizing Women," Tech. rep., World Food Programme. [2], [6]

A Online Appendix

A.1 Outcomes

Table A1: Outcome definitions

Outcome category	Definition
A Decision inputs	
1) Financial knowledge	Standardized score of financial knowledge questions. Basic knowledge of: budgeting, financial goals, savings options, financial products and services
2) Financial planning	Standardized score of habits and planning related to expenses, record-keeping, and budgeting
3) Financial agency	Standardized score of financial agency behaviors: having a personal goal; being in charge of account decisions, E-wallet decisions and day-to-day decisions
4) Financial trust	Standardized score of trust in bank, agent bank and a set of financial institutions
B Downstream financial behaviors & DFS adoption	
1) Budget use	Are you using a budget currently? - Yes - No
2) Saving goal	Do you currently have a saving goal? - Yes - No
3) Uses a mobile phone	Do you use a mobile phone and/or have a personal one? - Yes - No
4) Registered bank account	Do you personally have a bank account that is registered in your name and active? - Yes - No
5) Saving formally	Paying money into a <savings/deposit> account - Yes - No
6) Weekly DFS transactions	How many transactions does this household make through DFS based on a recall period weekly?
7) Monthly DFS transactions	How many transactions does this household make through DFS based on a recall period monthly?
8) Weekly E-wallet transaction	How many transactions does this household make through E-wallet based on a recall period weekly?
9) Monthly E-wallet transaction	How many transactions does this household make through E-wallet based on a recall period monthly?

Table A2: Baseline mean values and balancing (restricted *sharp* sample)

	(1)	(2)	(3)	(4)	(5)
	Sample	*Control*	*Treatment*	p-value $H_0 : T = C$	Normalized difference
	Mean				
Observables:					
Female	0.89	0.89	0.91	0.56	0.02
Age	38.57	38.10	39.50	0.00	0.02
Household head	0.83	0.79	0.89	0.01	0.04
Household head female	0.60	0.56	0.69	0.12	0.04
Years of education	2.74	2.99	2.24	0.03	0.03
Household head Education (> primary)	0.14	0.15	0.12	0.13	0.01
Household head no education (0 yrs)	0.53	0.50	0.59	0.07	0.03
N. of children 0-5	1.29	1.30	1.25	0.01	0.01
Household size	5.83	5.67	6.15	0.16	0.02
No income earners	0.35	0.28	0.49	0.08	0.07
Gone without cash income sometimes or often	0.85	0.87	0.83	0.31	0.02
Rarely or never has emergency fund	0.67	0.66	0.67	0.20	0.00
Bad connection experience	0.50	0.52	0.46	0.81	0.03
Key decision inputs:					
Financial knowledge index [0;1]	0.58	0.63	0.47	0.79	0.08
Financial planning index [0;1]	0.48	0.50	0.46	0.16	0.02
Financial agency index [0;1]	0.28	0.28	0.28	0.28	0.00
Financial trust index [0;1]	0.73	0.72	0.77	0.68	0.03
Downstream financial behaviors and DFS adoption:					
Budget use [0;1]	0.52	0.47	0.62	0.06	0.05
Saving goal [0;1]	0.46	0.47	0.45	0.31	0.01
Uses mobile phone [0;1]	0.67	0.68	0.66	0.75	0.01
Registered bank account	0.18	0.27	0.01	0.06	0.11
Saving formally [0;1]	0.04	0.06	0.00	0.09	0.05
Weekly DFS transactions [0;21]	0.19	0.18	0.22	0.07	0.01
Monthly DFS transactions [0;48]	0.76	0.81	0.65	0.55	0.01
Weekly E-wallet transactions [0;27]	0.05	0.05	0.04	0.78	0.01
Monthly E-wallet transactions [0;36]	0.27	0.30	0.21	0.14	0.02
Observations	*1,546*	*1,033*	*513*		

Notes: The table presents a summary of statistics and balance tests over treatment assignment groups for outcomes and covariates. We consider here the restricted estimation sample, that is the one which excludes participants who results already treated at baseline (103 units). Columns (1) to (3) show the average values for the full sample, the treatment group, and the control group, respectively. Column (4) displays the results of a t-test for the impact of the treatment on each variable listed on the left side. Column (5) displays the normalized difference between treatment and control for each variable listed on the left side. Both the p-values and the normalized differences are calculated from a regression model that takes into account the treatment assignment and stratification variables performing the STATA command iebaltab. The standard errors are clustered at the zone level. The statistics reported are in absolute values.

Table A3: Baseline mean values and balancing (unrestricted *fuzzy* sample)

	(1)	(2)	(3)	(4)	(5)
	Sample	*G0*	*G1*		
	Mean			p-value $H_0 : T = C$	Normalized difference
Observables:					
Female	0.88	0.87	0.90	0.99	0.00
Age	38.45	37.90	39.13	0.10	0.01
Household head	0.82	0.82	0.82	0.43	0.03
Household head female	0.60	0.61	0.59	0.18	0.03
Years of education	2.89	3.10	2.65	0.55	0.01
Household head Education (> primary)	0.15	0.16	0.13	0.87	0.00
Household head no education (0 yrs)	0.52	0.50	.54	0.41	0.01
N. of children 0-5	1.28	1.19	1.39	0.01	0.01
Household size	5.84	5.56	6.18	0.10	0.02
No income earners	0.35	0.37	0.31	0.05	0.06
Gone without cash income sometimes or often	0.85	0.85	0.86	0.22	0.02
Rarely or never has emergency fund	0.65	0.68	0.61	0.10	0.01
Bad connection experience	0.49	0.49	0.50	0.26	0.04
Key decision inputs:					
Financial knowledge index [0;1]	0.59	0.55	0.65	0.56	0.05
Financial planning index [0;1]	0.50	0.47	0.54	0.04	0.00
Financial agency index [0;1]	0.29	0.27	0.30	0.12	0.01
Financial trust index [0;1]	0.73	0.74	0.70	0.59	0.02
Downstream financial behaviors and DFS adoption:					
Budget use [0;1]	0.54	0.53	0.55	0.01	0.06
Saving goal [0;1]	0.48	0.46	0.51	0.11	0.01
Uses mobile phone [0;1]	0.68	0.66	0.71	0.28	0.00
Registered bank account	0.18	0.14	0.24	0.04	0.09
Saving formally [0;1]	0.04	0.03	0.05	0.10	0.04
Weekly DFS transactions [0;21]	0.22	0.21	0.24	0.02	0.03
Monthly DFS transactions [0;48]	0.83	0.80	0.87	0.31	0.00
Weekly E-wallet transactions [0;27]	0.07	0.08	0.06	0.40	0.02
Monthly E-wallet transactions [0;36]	0.34	0.33	0.36	0.28	0.01
Observations	*1,649*	*909*	*740*		

Notes: The table presents baseline mean values and balance tests of pre-treatment outcome values and households' characteristics. We consider here the (unrestricted) estimation sample, that is the one which includes participants who result already treated at baseline (103 units). Columns (1) to (3) show the average values for the full sample, the treatment group, and the control group, respectively. Column (4) displays the results of a t-test for the impact of the treatment on each variable listed on the left side. Column (5) displays the normalized difference between treatment and control for each variable listed on the left side. Both the p-values and the normalized differences are calculated from a regression model that takes into account the treatment assignment and stratification variables performing the STATA command iebaltab. The standard errors are clustered at the zone level. The statistics reported are in absolute values.

A.2 Fuzzy difference-in-difference: details

To be clear about how the de Chaisemartin and D'HaultfŒuille (2017)'s fuzzy DID methods work in our context, we introduce some notation and define our target parameter and estimands.

Let Y be the output variable, D a binary treatment status (1 for treated, 0 for untreated), and T the time period (1 for post-intervention, 0 for pre-intervention). $Y(1)$ and $Y(0)$ are the potential outcomes of the same unit with and without treatment. Whereas, $D(0)$ and $D(1)$ are the potential treatment statuses of an individual before and after the intervention. Differently from before, let G represent whether or not an individual's treatment status increased between two time periods (1 for increased, 0 for unchanged or decreased). Individuals who become treated in the second period are referred to as "switchers", $S \equiv \{D(0) < D(1), G = 1\}$. Similarly, individuals who become untreated in the second period are referred to as "control group switchers," $S' \equiv \{D(0) \neq D(1), G = 0\}$. In our study, we excluded control group switchers from the main analysis as it likely indicates errors in the data collection. For switchers, the DID estimator in fuzzy settings targets the Local Average Treatment Effect (LATE), which is the effect on participants who transition to the treatment group in the second period:

$$\Delta = E(Y_{11}(1) - Y_{11}(0) | S).$$

The following set of assumptions are fulfilled by the definition of the group indicator G and are maintained throughout the text. Furthermore, for any random variable R, the random variable $Rg\,t \sim R|G = g, T = t$.

Assumption A.1. Y, D, T *and* G *satisfy*

(i) *(Fuzzy design)* $E(D_{11}) > E(D_{10})$ *and* $E(D_{11}) - E(D_{10}) > E(D_{01}) - E(D_{00})$*;*

(ii) *(Stable percentage of treated units in the control group)* $0 < E(D_{01}) = E(D_{00}) < 1$*;*

(iii) *(Treatment participation equation)* $D = 1\{V \geq vGT\}$*, with* $V \perp T | G$*;*

Assumption A.1-(i) states that the proportion of treated individuals in the treatment group increases more than in the control group between the two periods. Assumption A.1-(ii) states that the proportion of treated individuals in the control group remains constant across periods. Assumption A.1-(iii) defines the treatment participation equation. This is determined by a latent index model, where an individual's likelihood of being treated is represented by a variable V, and the threshold for treatment depends on both time and group. This implies that, within each group, treatment can only change in one direction between the two periods. [28]

The following set of assumptions are necessary for the validity of both the sharp and fuzzy difference- in-difference methods.

Assumption A.2. Y*, D, T and G satisfy*

(i) *(Common trend)* $E(Y(0)|G, T = 1) - E(Y(0)|G, T = 0)$ *does not depend on* G*;*

(ii) *(Stable treatment effect) For all* $d \in S(D), E(Y(1) - Y(0)|G, T = 1, D(0) = 1) = E(Y(1) - Y(0)|G, T = 0, D(0) = 1)$.

Assumption A.2-(i) is the common trend assumption and it likely holds in our case due to the short du- ration of the intervention. Assumption A.2-(ii) is arguably stronger and it states that the treatment effect remains stable over time, meaning that the effect of the treatment does not change over the course of the study. In our case, it is not

[28] Assumption A.1-(i) and Assumption A.1-(ii) are satisfied by construction given that G is defined in terms of treatment status rate: G=1 for individuals who show a positive treatment rate between baseline and endline and G = 0 for individuals who show constant or negative treatment rate. Thus, according to such definition always treated units fall into the control group. Assumption A.1-(iii) holds in our case given that households with negative treatment rate between the two periods are excluded from the analysis as it likely indicates errors in the data collection.

simple to determine the accuracy of assumption A.2-(ii) because we compute treatment effect parameters for various outcomes. However, studies on financial education have shown that both financial knowledge and behaviors decline over time after training (e.g., Horn et al., 2020). Therefore, assuming that the treatment effects measured immediately after treatment and 3 months later would be the same for households who received the treatment at baseline may be unlikely in this context. However, we note that the time between $T = 0$ and $T = 1$ in our study is relatively short compared to other financial literacy programs, so it may be reasonable to assume that the average effect of receiving the FLT at baseline would remain constant when estimated 3 months later. Under Assumptions A.1 and A.2, the Wald-DID estimator identifies the target parameter:

$$W_{DID} = \frac{E(Y_{11}) - E(Y_{10}) - \left(E(Y_{01}) - E(Y_{00})\right)}{E(D_{11}) - E(D_{10}) - \left(E(D_{01}) - E(D_{00})\right)} = \Delta \qquad \text{(A1)}$$

From equation (A1) is clear that the basic idea behind the fuzzy approach is to adjust the difference- in-differences (DID) estimate of the outcome based on the DID of the treatment status. Note that, since Assumption A.2-(ii) may be questionable in our context, as an alternative (preferred) method we employ an estimator that requires a weaker assumption about the stability of the treatment effect parameter.

Assumption A.3.
(Conditional common trend) For all $d \in S(D), E(Y(d)|G, T = 1, D(0) = d) - E(Y(d)|G, T = 0, D(0) = d)$ *does not depend on G.*

Assumption A.3 states that the mean of $Y(1)$ and $Y(0)$ evolves similarly over time among treatment and control units that were, respectively, treated and not treated at the baseline.
Under Assumptions A.1 and A.3, the Time-corrected (Tc) Wald DID estimator identifies the target Parameter:

$$W_{TC} = \frac{E(Y_{11}) - E\left(Y_{10} + \delta_{D_{10}}\right)}{E(D_{11}) - E(D_{10})} = \Delta \qquad \text{(A2)}$$

where $\delta_d = E(Yd_{01}) - E(Yd_{00})$ denotes the change in the mean outcome between period 0 and 1 for control group units with treatment status *d*.[29]

[29] Note that de Chaisemartin and D'HaultfŒuille (2017) introduce a third estimator, called the Wald-CIC, which is based on the changes-in-changes estimator developed by Athey and Imbens (2006). The Wald-CIC requires an additional assumption of monotonicity and time invariance of unobservables. When assumption A.2-(ii) is problematic, the choice of a second estimator depends on certain conditions, with the TC or CIC being selected accordingly. The CIC requires a new assumption that imposes restrictions on the entire outcome distribution but is invariant to scaling of the outcome, while assumption A.3 only restricts the mean and is not invariant to scaling of the outcome. In this case, the Wald-TC appears to be the most suitable choice since the distribution of the outcomes conditional on *D* in the first period are quite similar.

A.3 Treatment effects on single components of the indexes

Table A4: Financial knowledge and its components

	(1)	(2)	(3)	(4)	(5)	(6)	(6)
	Sharp-DID		Wald-DID		Tc-DID		Mean Dep. Var.
Financial Knowledge Index - Same as Table 2 Row 1	0.632***	0.654**	0.805***	0.638***	0.888***	0.652***	0.595
	(0.212)	(0.313)	(0.147)	(0.191)	(0.152)	(0.205)	
	[0.007]	[0.101]	[0.001]	[0.002]	[0.001]	[0.003]	
Financial products/services knowledge	0.392*	0.392	0.377***	0.282	0.408***	0.280	0.523
	(0.227)	(0.335)	(0.131)	(0.201)	(0.152)	(0.214)	
	[0.039]	[0.156]	[0.002]	[0.077]	[0.004]	[0.093]	
Digital products/services knowledge	0.505***	0.482***	0.322***	0.208	0.359***	0.205	0.571
	(0.124)	(0.186)	(0.110)	(0.138)	(0.130)	(0.145)	
	[0.001]	[0.101]	[0.002]	[0.071]	[0.003]	[0.086]	
Electronic payment instruments awareness	0.628***	0.621**	0.362**	0.212	0.452***	0.213	0.349
	(0.196)	(0.293)	(0.166)	(0.186)	(0.164)	(0.192)	
	[0.007]	[0.101]	[0.005]	[0.083]	[0.003]	[0.103]	
Understanding of what a budget is	0.130	0.139	0.321***	0.271*	0.289***	0.265*	0.749
	(0.210)	(0.308)	(0.108)	(0.139)	(0.112)	(0.144)	
	[0.104]	[0.244]	[0.002]	[0.031]	[0.004]	[0.039]	
Savings and how to save knowledge	0.390**	0.424	0.589***	0.508***	0.630***	0.512***	0.777
	(0.179)	(0.267)	(0.112)	(0.122)	(0.124)	(0.127)	
	[0.026]	[0.107]	[0.001]	[0.001]	[0.001]	[0.001]	
Range of savings options knowledge	0.347**	0.401*	0.758***	0.632***	0.836***	0.645***	0.649
	(0.149)	(0.217)	(0.134)	(0.141)	(0.140)	(0.153)	
	[0.021]	[0.101]	[0.001]	[0.001]	[0.001]	[0.001]	
Setting savings goals knowledge	0.369*	0.411	0.771***	0.550***	0.848***	0.549***	0.631
	(0.203)	(0.297)	(0.129)	(0.191)	(0.140)	(0.212)	
	[0.036]	[0.143]	[0.001]	[0.004]	[0.001]	[0.009]	
Making financial goals knowledge	0.469**	0.492*	0.825***	0.634***	0.886***	0.659***	0.530
	(0.190)	(0.285)	(0.148)	(0.183)	(0.165)	(0.204)	
	[0.017]	[0.101]	[0.001]	[0.001]	[0.001]	[0.003]	
Making weekly/monthly spending plan knowledge	0.652***	0.661**	0.646***	0.646***	0.765***	0.699***	0.574
	(0.211)	(0.314)	(0.166)	(0.175)	(0.162)	(0.185)	
	[0.007]	[0.101]	[0.001]	[0.001]	[0.001]	[0.001]	
Household FE	✗	✓	✗	✓	✗	✓	
Observations	*2,996*	*2,996*	*3,200*	*3,200*	*3,200*	*3,200*	

Notes: Dependent variables: standardized variables displayed on the row headings. To calculate the index in the first row, we take the mean of its non-missing components displayed in the following row headings (each of which has control group mean 0 and SD 1) and then restandardize to SD = 1 so that treatment effect parameters are in standard deviation units. Coefficients in column (1)-(2) were estimated using an OLS regression of dependent variables listed in the row headings on a dummy for treatment assignment, a dummy for time and an interaction between the two. Columns (3)-(4) and (5)-(6) reports treatment effect estimates respectively from fuzzy DID regression using Wald-DID estimator and using Tc-DID estimator. Each of the variables in the row headings is regressed on a dummy for treatment status, a dummy for the group and one for time. All DID specifications are conducted with clustering at the unit of randomization (zones) and controlling for the distance of zones within the settlement as well as for settlement fixed effects. These controls are considered *stratification variables because they help account for differences in treatment assignment probabilities across zones, which arose due to imperfections in the randomization process. Standard errors in (parentheses). FDR adjusted p-values in [square brackets], calculated using the two-stage procedure in Benjamini et al. (2006). *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.*

Table A5: Financial planning and its components

	(1)	(2)	(3)	(4)	(5)	(6)	(6)
	Sharp-DID		Wald-DID		Tc-DID		Mean Dep. Var.
Financial Planning Index - Same as Table 2 Row 2	0.480**	0.501	0.737***	0.703***	0.842***	0.744***	0.503
	(0.210)	(0.313)	(0.165)	(0.183)	(0.146)	(0.190)	
	[0.050]	[0.295]	[0.001]	[0.001]	[0.001]	[0.001]	
Having written budgets	0.058	0.088	0.278*	0.211	0.347**	0.240	0.619
	(0.188)	(0.270)	(0.159)	(0.162)	(0.175)	(0.176)	
	[0.299]	[0.425]	[0.017]	[0.075]	[0.011]	[0.075]	
Making plans ahead	0.327**	0.331	0.543***	0.356**	0.657***	0.365*	0.577
	(0.163)	(0.241)	(0.139)	(0.181)	(0.136)	(0.197)	
	[0.063]	[0.295]	[0.001]	[0.029]	[0.001]	[0.038]	
Monitoring expenses	0.140	0.134	0.285*	0.310*	0.242	0.333*	0.716
	(0.205)	(0.303)	(0.164)	(0.181)	(0.204)	(0.199)	
	[0.282]	[0.414]	[0.017]	[0.041]	[0.050]	[0.044]	
Record keeping habits	0.620***	0.635**	0.773***	0.871***	0.831***	0.914***	0.269
	(0.209)	(0.315)	(0.153)	(0.146)	(0.159)	(0.144)	
	[0.016]	[0.295]	[0.001]	[0.001]	[0.001]	[0.001]	
Weekly/monthly plans on income/expenses	0.410**	0.417	0.419**	0.522***	0.524***	0.563***	0.493
	(0.188)	(0.279)	(0.167)	(0.161)	(0.156)	(0.162)	
	[0.050]	[0.295]	[0.007]	[0.002]	[0.001]	[0.001]	
Keeping spending notes	0.359**	0.346	0.546***	0.572***	0.581***	0.611***	0.338
	(0.165)	(0.247)	(0.182)	(0.154)	(0.198)	(0.160)	
	[0.050]	[0.295]	[0.003]	[0.001]	[0.002]	[0.001]	
Having an emergency fund	0.171	0.207	0.329***	0.396***	0.405***	0.388**	0.355
	(0.224)	(0.329)	(0.093)	(0.150)	(0.103)	(0.164)	
	[0.282]	[0.395]	[0.001]	[0.007]	[0.001]	[0.016]	
Keeping track of money	0.553***	0.554**	0.479***	0.392*	0.516***	0.428*	0.675
	(0.186)	(0.277)	(0.182)	(0.230)	(0.187)	(0.240)	
	[0.016]	[0.295]	[0.005]	[0.041]	[0.003]	[0.039]	
Having a saving goal	0.230	0.278	0.715***	0.566***	0.873***	0.589***	0.482
	(0.164)	(0.235)	(0.168)	(0.163)	(0.150)	(0.174)	
	[0.102]	[0.372]	[0.001]	[0.001]	[0.001]	[0.001]	
Household FE	✗	✓	✗	✓	✗	✓	
Observations	*2,996*	*2,996*	*3,200*	*3,200*	*3,200*	*3,200*	

*Notes: Dependent variables: standardized variables displayed on the row headings. To calculate the index in the first row, we take the mean of its non-missing components displayed in the following row headings (each of which has control group mean 0 and SD 1) and then restandardize to SD = 1 so that treatment effect parameters are in standard deviation units. Coefficients in column (1)-(2) were estimated using an OLS regression of dependent variables listed in the row headings on a dummy for treatment assignment, a dummy for time and an interaction between the two. Columns (3)-(4) and (5)-(6) reports treatment effect estimates respectively from fuzzy DID regression using Wald-DID estimator and using Tc-DID estimator. Each of the variables in the row headings is regressed on a dummy for treatment status, a dummy for the group and one for time. All DID specifications are conducted with clustering at the unit of randomization (zones) and controlling for the distance of zones within the settlement as well as for settlement fixed effects. These controls are considered stratification variables because they help account for differences in treatment assignment probabilities across zones, which arose due to imperfections in the randomization process. FDR adjusted p-values in [square brackets], calculated using the two-stage procedure in Benjamini et al. (2006). *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.*

Table A6: Financial agency and its components

	(1)	(2)	(3)	(4)	(5)	(6)	(6)
Sharp-DID						Wald-DID Tc-DID Mean Dep. Var.	
Financial Agency Index - Same as Table 2 Row 3	0.412*	0.459	0.693***	0.571***	0.839***	0.617***	
	(0.217)	(0.321)	(0.161)	(0.175)	(0.140)	(0.188)	0.285
	[0.065]	[0.342]	[0.001]	[0.003]	[0.001]	[0.003]	
Take day to day decisions	0.114	0.138	0.241*	0.212	0.228*	0.229	0.759
	(0.206)	(0.306)	(0.123)	(0.173)	(0.136)	(0.188)	
	[0.409]	[0.484]	[0.026]	[0.096]	[0.019]	[0.091]	
Having a personal goal	0.295**	0.326	0.570***	0.412***	0.768***	0.445***	0.371
	(0.131)	(0.200)	(0.146)	(0.116)	(0.110)	(0.124)	
	[0.065]	[0.342]	[0.001]	[0.002]	[0.001]	[0.002]	
Taking account decisions	−0.353	−0.282	0.398*	0.431*	0.488***	0.493*	0.008
	(0.279)	(0.411)	(0.216)	(0.257)	(0.189)	(0.269)	
	[1.000]	[1.000]	[0.027]	[0.050]	[0.005]	[0.035]	
Taking e-wallet decisions	2.551**	2.591*	1.638***	1.567***	1.793***	1.646***	0.003
	(1.057)	(1.525)	(0.556)	(0.601)	(0.518)	(0.609)	
	[0.065]	[0.342]	[0.004]	[0.010]	[0.001]	[0.007]	
Household FE	✗	✓	✗	✓	✗	✓	
Observations	*2,996*	*2,996*	*3,200*	*3,200*	*3,200*	*3,200*	

*Notes: Dependent variables: standardized variables displayed on the row headings. To calculate the index in the first row, we take the mean of its non-missing components displayed in the following row headings (each of which has control group mean 0 and SD 1) and then restandardize to SD = 1 so that treatment effect parameters are in standard deviation units. Coefficients in column (1)-(2) were estimated using an OLS regression of dependent variables listed in the row headings on a dummy for treatment assignment, a dummy for time and an interaction between the two. Columns (3)-(4) and (5)-(6) report treatment effect estimates respectively from fuzzy DID regression using Wald-DID estimator and using Tc-DID estimator. Each of the variables in the row headings is regressed on a dummy for treatment status, a dummy for the group and one for time. All DID specifications are conducted with clustering at the unit of randomization (zones) and controlling for the distance of zones within the settlement as well as for settlement fixed effects. These controls are considered stratification variables because they help account for differences in treatment assignment probabilities across zones, which arose due to imperfections in the randomization process. FDR adjusted p-values in [square brackets], calculated using the two-stage procedure in Benjamini et al. (2006). *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.*

Table A7: Financial trust and its components

	(1)	(2)	(3)	(4)	(5)	(6)	(6)
Sharp-DID						Wald-DID Tc-DID Mean Dep. Var.	
Financial Trust Index - Same as Table 2 Row 4	−0.178	−0.176	0.056	0.126	−0.115	0.126	0.726
	(0.257)	(0.378)	(0.169)	(0.199)	(0.173)	(0.217)	
	[1.000]	[1.000]	[1.000]	[1.000]	[1.000]	[1.000]	
Trust in bank	−0.596	−0.563	0.023	0.427	−0.204	0.445	0.599
	(0.349)	(0.518)	(0.158)	(0.266)	(0.160)	(0.283)	
	[1.000]	[1.000]	[1.000]	[1.000]	[1.000]	[1.000]	
trust in bank's agent'	0.037	0.073	0.013	−0.066	−0.003	−0.062	0.573
	(0.206)	(0.290)	(0.110)	(0.119)	(0.098)	(0.127)	
	[1.000]	[1.000]	[1.000]	[1.000]	[1.000]	[1.000]	
Trust in banks on the wheels	−0.073	−0.113	0.139	0.087	0.061	0.062	0.351
	(0.262)	(0.381)	(0.187)	(0.166)	(0.210)	(0.175)	
	[1.000]	[1.000]	[1.000]	[1.000]	[1.000]	[1.000]	
Trust in MTN mobile money	−0.131	−0.152	−0.035	−0.062	−0.062	−0.094	
	(0.147)	(0.210)	(0.116)	(0.124)	(0.147)	(0.140)	0.513
	[1.000]	[1.000]	[1.000]	[1.000]	[1.000]	[1.000]	
Trust in Airtel Mobile Money	−0.052	−0.036	0.263	0.099	0.190	0.093	0.783
	(0.243)	(0.357)	(0.212)	(0.175)	(0.225)	(0.188)	
	[1.000]	[1.000]	[1.000]	[1.000]	[1.000]	[1.000]	
Trust in Mpesa	−0.130	−0.123	0.049	0.169	−0.030	0.215	0.874
	(0.277)	(0.410)	(0.205)	(0.232)	(0.189)	(0.242)	
	[1.000]	[1.000]	[1.000]	[1.000]	[1.000]	[1.000]	
Trust in Africell mobile money	−0.131	−0.134	0.113	0.177	−0.170	0.148	0.874
	(0.268)	(0.397)	(0.212)	(0.216)	(0.192)	(0.239)	
	[1.000]	[1.000]	[1.000]	[1.000]	[1.000]	[1.000]	
trust in UTL mobile money	−0.076	−0.073	0.108	0.183	−0.062	0.202	0.883
	(0.268)	(0.399)	(0.208)	(0.231)	(0.191)	(0.246)	
	[1.000]	[1.000]	[1.000]	[1.000]	[1.000]	[1.000]	
Trust in Micro Finance Institution	−0.136	−0.116	0.002	0.154	−0.166	0.169	0.867
	(0.249)	(0.369)	(0.188)	(0.208)	(0.191)	(0.224)	
	[1.000]	[1.000]	[1.000]	[1.000]	[1.000]	[1.000]	
Trust in VSLA	0.080	0.023	−0.214	−0.257	−0.346	−0.290	0.612
	(0.135)	(0.209)	(0.140)	(0.107)	(0.187)	(0.129)	
	[1.000]	[1.000]	[1.000]	[1.000]	[1.000]	[1.000]	
Trust in Insurance agencies	−0.172	−0.159	0.060	0.173	−0.034	0.194	0.877
	(0.253)	(0.374)	(0.187)	(0.213)	(0.182)	(0.226)	
	[1.000]	[1.000]	[1.000]	[1.000]	[1.000]	[1.000]	
Trust in Private lending agencies	0.022	0.052	−0.007	0.060	−0.053	0.099	0.903
	(0.298)	(0.440)	(0.215)	(0.230)	(0.205)	(0.244)	
	[1.000]	[1.000]	[1.000]	[1.000]	[1.000]	[1.000]	
Household FE	✗	✓	✗	✓	✗	✓	
Observations	*2,996*	*2,996*	*3,200*	*3,200*	*3,200*	*3,200*	

Notes: Dependent variables: standardized variables displayed on the row headings. To calculate the index in the first row, we take the mean of its non-missing components displayed in the following row headings (each of which has control group mean 0 and SD 1) and then restandardize to SD = 1 so that treatment effect parameters are in standard deviation *units. Coefficients in column (1)-(2) were estimated using an OLS regression of dependent variables listed in the row headings on a dummy for treatment assignment, a dummy for time and an interaction between the two. Columns (3)-(4) and (5)-(6) report treatment effect estimates respectively from fuzzy DID regression using Wald-DID estimator and using Tc-DID estimator. Each of the variables in the row headings is regressed on a dummy for treatment status, a dummy for the group and one for time. All DID specifications are conducted with clustering at the unit of randomization (zones) and controlling for the distance of zones within the settlement as well as for settlement fixed effects. These controls are considered stratification variables because they help account for differences in treatment assignment probabilities across zones, which arose due to imperfections in the randomization process. Standard errors in (parentheses). FDR adjusted p-values in [square brackets], calculated using the two-stage procedure in Benjamini et al. (2006). *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.*

A.4 Treatment effects - Robustness checks

Table A8: Decision inputs - Excluding 103 already treated at baseline

	(1)	(2)	(3)	(4)	(5)
	Sharp-DID		Wald-DID		Mean Dep. Var.
Indexes:					
Financial knowledge	0.632***	0.654**	0.888***	0.652***	0.575
	(0.212)	(0.313)	(0.157)	(0.203)	
	[0.012]	[0.172]	[0.001]	[0.002]	
Financial planning	0.480**	0.501	0.842***	0.744***	0.485
	(0.210)	(0.313)	(0.133)	(0.186)	
	[0.035]	[0.181]	[0.001]	[0.001]	
Financial agency	0.412*	0.459	0.839***	0.617***	0.279
	(0.217)	(0.321)	(0.132)	(0.187)	
	[0.047]	[0.181]	[0.001]	[0.002]	
Financial trust	−0.178	−0.176	−0.115	0.126	0.733
	(0.257)	(0.378)	(0.179)	(0.192)	
	[0.140]	[0.256]	[0.150]	[0.148]	
Household FE	✓	✗	✗	✓	✗
Observations	*2,996*	*2,996*	*3,097*	*3,097*	*3,097*

*Notes: As a robusteness check, the table reports treatment effects on financial knowledge and behaviors excluding from fuzzy estimation samples the subsample of households which result already treated at baseline (103 units). In this case, Wald-DID estimator equals the Tc-DID since treatment takes only one value in the control group. Dependent variables: standardized composite indexes. Coefficients in column (1)-(2) were estimated using an OLS regression of dependent variables listed in the row headings on a dummy for treatment assignment, a dummy for time and an interaction between the two. Columns (3)-(4) and (5)-(6) report treatment effect estimates respectively from fuzzy DID regression using Wald-DID estimator and using Tc-DID estimator. Each of the variables in the row headings is regressed on a dummy for treatment status, a dummy for the group and one for time. All DID specifications are conducted with clustering at the unit of randomization (zones) and controlling for the distance of zones within the settlement as well as for settlement fixed effects. These controls are considered stratification variables because they help account for differences in treatment assignment probabilities across zones, which arose due to imperfections in the randomization process. Standard errors in (parentheses). FDR adjusted p-values in [square brackets], calculated using the two-stage procedure in Benjamini et al. (2006). *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.*

Table A9: Decision inputs - Including downward switchers

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Sharp-DID		Wald-DID		Tc-DID		Mean Dep. Var.
Indexes:							
Financial knowledge	0.632***	0.654**	0.802***	0.636***	0.895***	0.639***	.598
	(0.212)	(0.313)	(0.146)	(0.183)	(0.152)	(0.203)	
	[0.012]	[0.172]	[0.001]	[0.001]	[0.001]	[0.002]	
Financial planning	0.480**	0.501	0.732***	0.695***	0.828***	0.723***	.506
	(0.210)	(0.313)	(0.161)	(0.176)	(0.144)	(0.189)	
	[0.035]	[0.181]	[0.001]	[0.001]	[0.001]	[0.001]	
Financial agency	0.412*	0.459	0.672***	0.552***	0.813***	0.593***	.286
	(0.217)	(0.321)	(0.161)	(0.168)	(0.140)	(0.186)	
	[0.047]	[0.181]	[0.001]	[0.001]	[0.001]	[0.002]	
Financial trust	−0.178	−0.176	0.042	0.111	−0.070	0.135	.723
	(0.257)	(0.378)	(0.167)	(0.192)	(0.162)	(0.211)	
	[0.140]	[0.256]	[0.252]	[0.165]	[0.200]	[0.150]	
Household FE	✗	✓	✗	✓	✗	✓	
Observations	*2,996*	*2,996*	*3,242*	*3,242*	*3,242*	*3,242*	

*Notes: As a robusteness check, the table reports treatment effects on financial knowledge and behaviors from two fuzzy DID models including in the estimation sample the subsample of households reporting negative treatment rate between the two periods which were excluded from the main specifications. Dependent variables: standardized composite indexes. Coefficients in column (1)-(2) were estimated using an OLS regression of dependent variables listed in the row headings on a dummy for treatment assignment, a dummy for time and an interaction between the two. Columns (3)-(4) and (5)-(6) report treatment effect estimates respectively from fuzzy DID regression using Wald-DID estimator and using Tc-DID estimator. Each of the variables in the row headings is regressed on a dummy for treatment status, a dummy for the group and one for time. All DID specifications are conducted with clustering at the unit of randomization (zones) and controlling for the distance of zones within the settlement as well as for settlement fixed effects. These controls are considered stratification variables because they help account for differences in treatment assignment probabilities across zones, which arose due to imperfections in the randomization process. Standard errors in (parentheses). FDR adjusted p-values in [square brackets], calculated using the two-stage procedure in Benjamini et al. (2006). *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.*