Learning from meta-analyses and open questions on casual mechanisms, scalability and long-term effects

Tim Kaiser

G53 Financial Literacy & Personal Finance Bootcamp

Stanford University

June 25th, 2025

Rheinland-Pfälzische Technische Universität
Kaiserslautern
Landau

Learning from meta-analyses of RCTs

Meta-Analyses in Economics

- Until late, meta-analyses have been rarely conducted in economics. Yet, this has changed drastically!
- With the expansion of field experiments, **aggregating estimated treatment effects** from mutliple studies has become increasingly relevant
- Recent examples are meta-analyses on:
    - Active Labour Market Programs (Card et al. 2018, JEEA)
    - Microcredit expansions (Meager 2019, AEJ: Applied; Meager 2022, AER)
    - Gender differences in response to performance pay (Bandiera et al. 2021, AER: Insights)
    - US K-12 public school spending (Jackson and Mackevicius 2024, AEJ: Applied)
    - Informational nudges (Dellavigna and Linos 2022, ECTA)
    - Present bias and loss aversion (Imai et al. 2021, EJ; Brown et al. 2024, JEL)
    - Impact evaluations of development interventions (Vivalt 2020, JEEA)
    - …

RPTU

Previous meta-analyses on financial education

- The first meta-analysis by Fernandes, Lynch, and Netemeyer (2014, *ManSci*) has been widely cited to provide evidence of ineffectiveness of financial education in general:
  - “We find that interventions to improve financial literacy explain only 0.1% of the variance in financial behaviors studied” (page 1861)
  - “Intervention effects may decay over time – the case for ‘just in time financial education’.”(page 1866)

- Other meta-analyses with different foci (specific outcomes and target groups) (Miller et al. 2015, Kaiser and Menkhoff 2017, 2020) have been published since, but have not moved the priors of sceptics

Negative priors reflected in the news

THE QUEST TO IMPROVE AMERICA'S FINANCIAL LITERACY IS BOTH A FAILURE AND A SHAM

Financial literacy promotion may sound perfectly sensible—who wouldn't want to teach children and adults the secrets of managing money?—but in the face of recent res
initiative.

HELAINE OLEN · JAN 7, 2014

ECONOMIC VIEW

Financial Literacy, Beyond the Classroom

By Richard H. Thaler

Oct. 5, 2013

PostEverything • Perspective

More states are forcing students to study personal finance. It's a waste of time.

Study after study shows that financial-literacy courses don't change behavior.

BUSINESS // MICHAEL TAYLOR

Taylor: Is financial literacy a bad thing?

Michael Taylor | Dec. 6, 2019 | Updated: Dec. 6, 2019 1:05 p.m.

TIME

FINANCIAL EDUCATION

Financial Education Is All the Rage but Does it Work?

Reaching consumers with advice and information just before making a financial decision is the new target. But is that really more effective than teaching personal finance in K-12?

By Dan Kadlec @dankadlec | Oct. 25, 2013

Why financial literacy programs don't work

BY ATTY. DODO DULAY JANUARY 01, 2019

HOME / OPINION / OP-ED COLUMNS / WHY FINANCIAL LITERACY PROGRAMS DON'T WORK

HOME / MONEY / PERSONAL FINANCE / MY MONEY

Why Investor Education Doesn't Work – And How to Change That

Employer-sponsored 401(k) meetings aren't always effective.

CPFB head misguided in reliance on consumer education

BY LAUREN E. WILLIS, OPINION CONTRIBUTOR — 09/07/19 03:30 PM EDT
THE VIEWS EXPRESSED BY CONTRIBUTORS ARE THEIR OWN AND NOT THE VIEW OF THE HILL

A primer on meta-analysis

- Consider a set of *j* randomized experiments, each of them reporting *i* estimates of treatment effects
- Allow different experiments to result in different true effects as opposed to estimating a fixed parameter
- Goal of this aggregation is to arrive at a “general effect” (mean of a distribution) and a scale parameter

→ choose weights for each observation (treatment effect estimate) that reflect the precision ot the estimate (as function of random sampling error) and the differences in site-specific results (heterogeneity in true effects)

Meta-Model

$$\hat{y}_{ij} = \beta_0 + \upsilon_j + \epsilon_{ij}$$

$\hat{y}_{ij}$ is the ith treatment effect estimate within each study j.

β_0 is the mean of the distribution of true effects of interventions

υ_j is a study-level random effect with $\upsilon_j \sim N(0, \tau^2)$, i.e., the true effects can vary between (but not within) studies.

$\epsilon_{ij} \sim N(0, \hat{\sigma}_{ij}^2)$ is the residual of the ith treatment effect estimate within each study j

- We observe both $\hat{y}_{ij}$ and $\hat{\sigma}_{ij}^2$ from the data (we assume $\sigma_{ij}^2 = \hat{\sigma}_{ij}^2$)
- τ^2 needs to be estimated (e.g., via Restricted Maximum Likelihood)
- We next run WLS with weights defined $(\hat{\tau}^2 + \hat{\sigma}_{ij}^2)^{-1}$
- We cluster the standard errors at the study-level for inference

RPTU

Comparison of the updated evidence to the result in Fernandes et al. (2014)

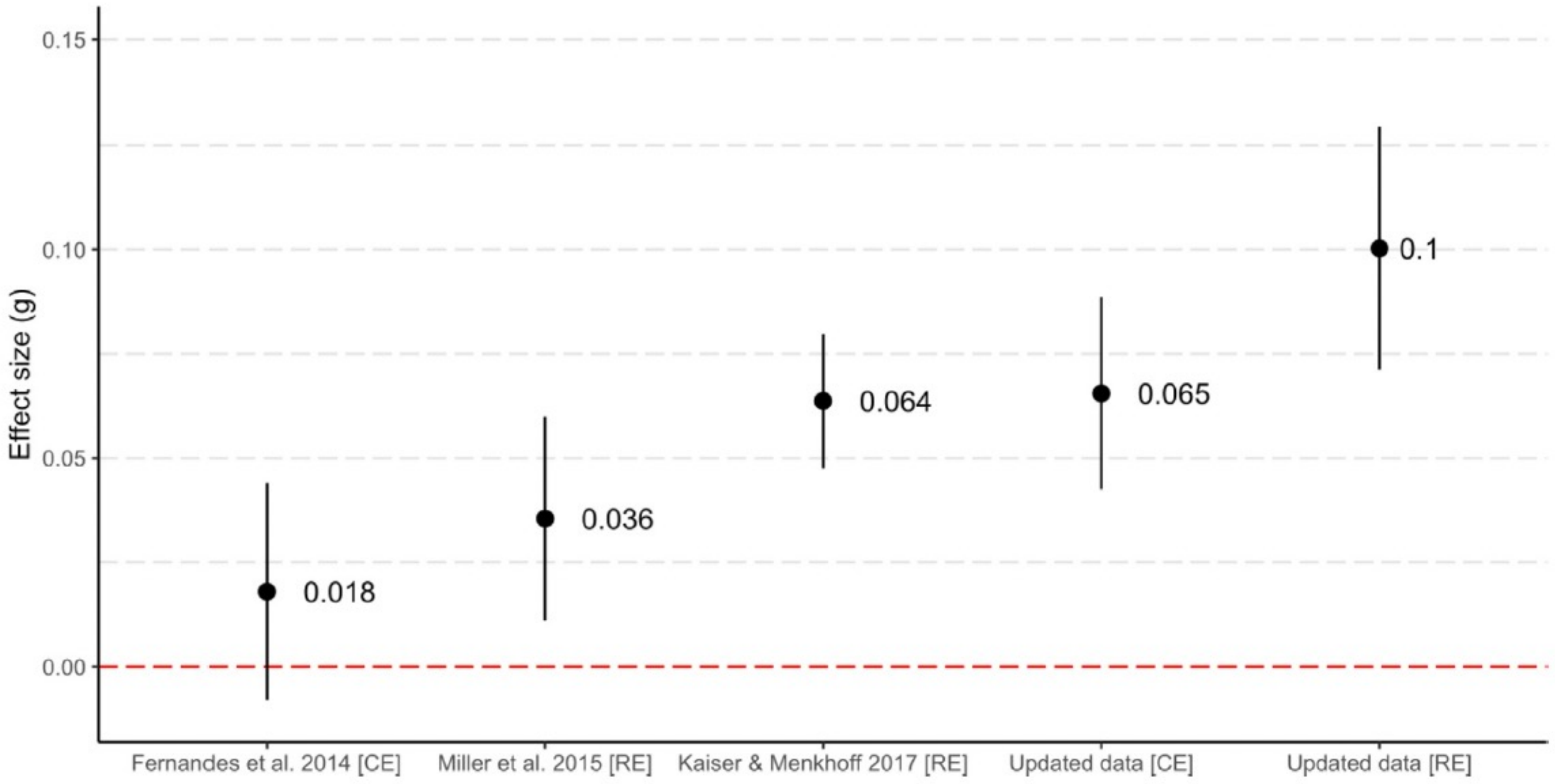


Fig. 3. Comparing the updated evidence to previous meta-analyses (treatment effects on *financial behaviors*).
Fernandes et al. (2014) report weighted least squares estimates with inverse variance weights (common-effect assumption) using 15 observations from 13 RCTs. Miller et al. (2015) use a random effects model and include results from 20 studies (13 quasi-experiments and seven RCTs). The result by Kaiser and Menkhoff (2017) is from a random effects model (RVE) using 349 observations from 90 studies (50 quasi-experiments and 40 RCTs). The results with updated data (458 treatment effect estimates from 64 RCTs) are from robust variance estimation in meta-regression with dependent effect size estimates (RVE) (Hedges et al., 2010) with $\tau^2 = 0$ in the common-effect case, and τ^2estimated via methods of moments in the random-effects case. Dots show the point estimates, and the solid lines indicate the 95% confidence interval.

Aggregated treatment effects by outcome domain

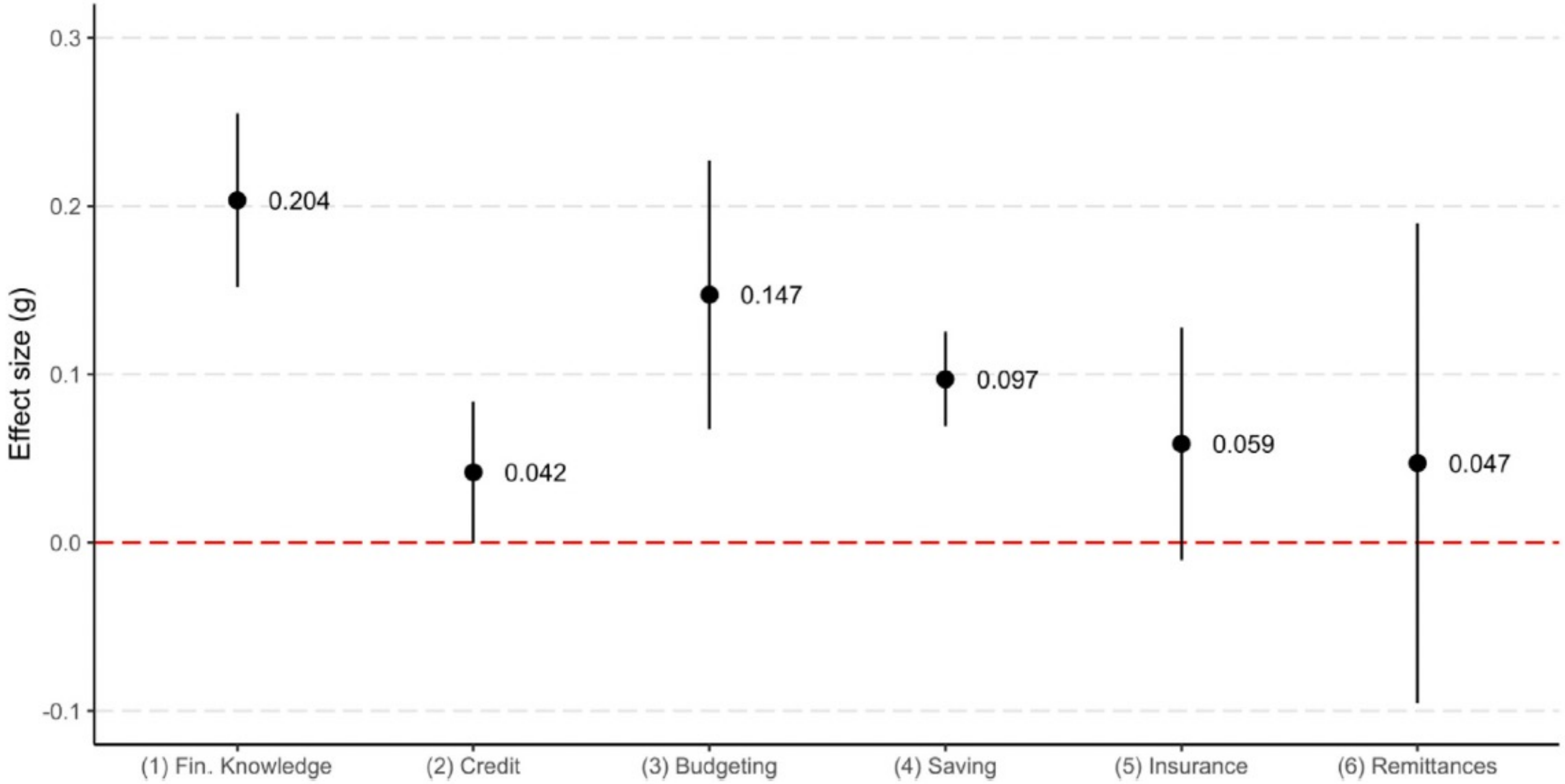


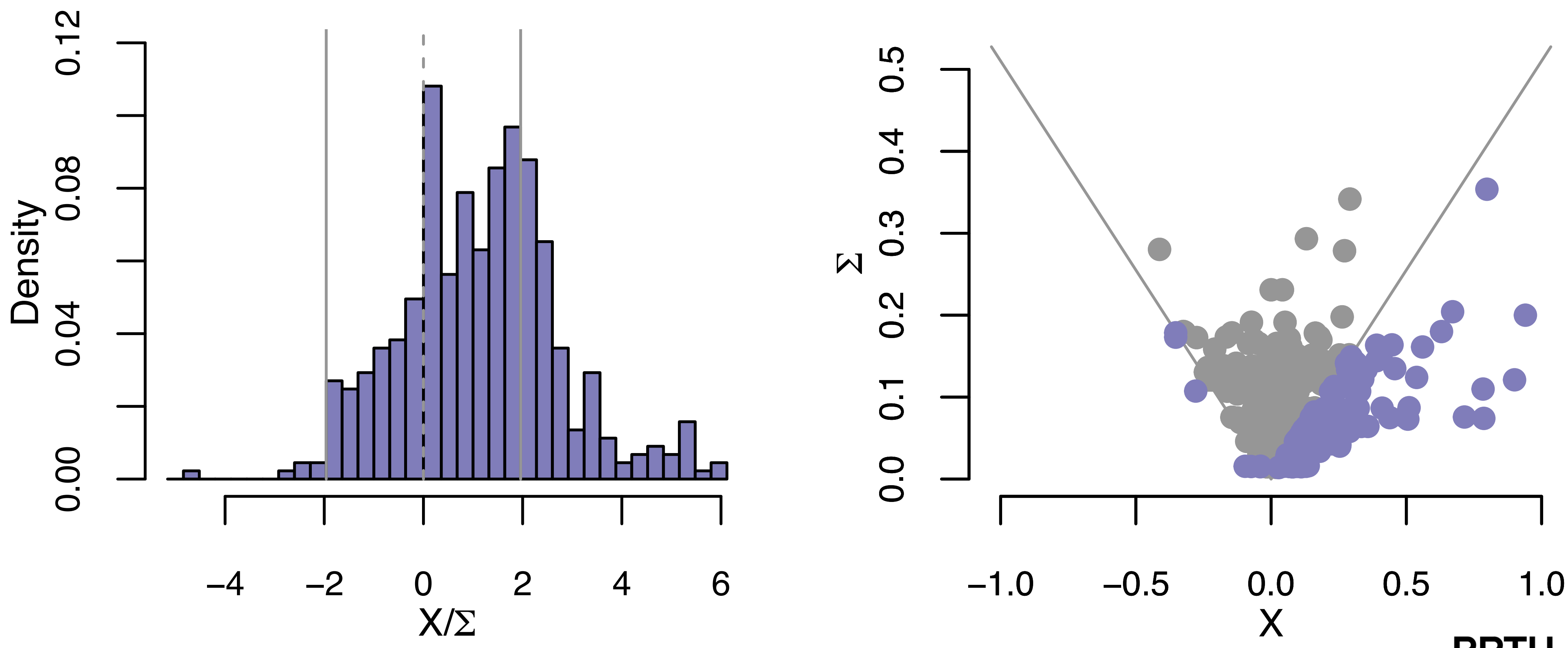
Fig. 4. Financial education treatment effects by outcome domain.
Results from robust variance estimation in meta-regression with dependent effect size estimates (RVE) (Hedges et al., 2010). The number of observations for the financial knowledge sample (1) is 215 effect size estimates within 50 studies. The number of observations for the credit behavior sample (2) is 115 within 22 studies. The number of effect size estimates for the budgeting behavior sample (3) is 55 within 23 studies. The number of observations in the saving and investing behavior (4) sample is 253 effect size estimates within 54 studies. The number of observations in the insurance behavior sample (5) is 18 effect sizes within six studies. The number of observations on remittance behavior (6) is 17 effect size estimates reported within six studies. Dots show the point estimates, and the solid lines indicate the 95% confidence interval.

RPTU

What about publication bias?

- Publication bias refers to the problem of authors (or journal editors) favoring the publication (selection) of statistically significant results
- Leaving this selection unadressed can lead to a biased assesement of a mean effect in a given literature
- Andrews and Kasy (2018, *AER*) develop a method for identifying and correcting publication bias using a step function approach

Results of Andrews and Kasy (2018) approach to this literature



RPTU

Results of Andrews and Kasy (2018) approach to this literature

Table 2: Identification of and correction for publitcation bias in the financial education literature

(a) Treatment effects on *financial behaviors*				(b) Treatment effects on *financial knowlege*			
(1) Selection on significance (cutoff of $Z = 1.96$)		(2) Selection on significance (cutoff of $Z = 1.65$)		(3) Selection on significance (cutoff of $Z = 1.96$)		(4) Selection on significance (cutoff of $Z = 1.65$)	
$\bar{\beta}_0$	λ_p	$\bar{\beta}_0$	λ_p	$\bar{\beta}_0$	λ_p	$\bar{\beta}_0$	λ_p
0.057	0.303	0.050	0.256	0.150	0.150	0.160	0.250
(0.001)	(0.071)	(0.007)	(0.051)	(0.037)	(0.126)	(0.040)	(0.190)

Notes: This table presents results from non-parametric identification of and correction for publication bias based on the method described in Andrews and Kasy (2018) (see Andrews and Kasy 2018, Appendix C). $\bar{\beta}_0$denotes the estimate of the true treatment effect in latent studies (i.e., the bias corrected treatmen effect) and $\lambda_p$denotes the estimated conditional publication probability (p) based on the Z-statistic (y_{ij}/σ_{ij}) as specified in the repective column header. Columns (1) and (3) show estimates for the treatment effects on financial behaviors and financial knowledge with $p(y_{ij}/\sigma_{ij}) = \lambda_p if\ |y_{ij}/\sigma_{ij}| < 1.96$ and $p(y_{ij}/\sigma_{ij}) = 1\ if\ |y_{ij}/\sigma_{ij}| \geq 1.96$, i.e., selection on significance at the 5%-level, repectively. Columns (2) and (4) show estimates for for the treatment effects on financial behaviors and financial knowledge with $p(y_{ij}/\sigma_{ij}) = \lambda_p if\ |y_{ij}/\sigma_{ij}| < 1.65$ and $p(y_{ij}/\sigma_{ij}) = 1\ if\ |y_{ij}/\sigma_{ij}| \geq 1.65$, i.e., selection on significance at the 10%-level, repectively. Standard errors (clustered at the study-level) are shown in parentheses.

How large are the effects?

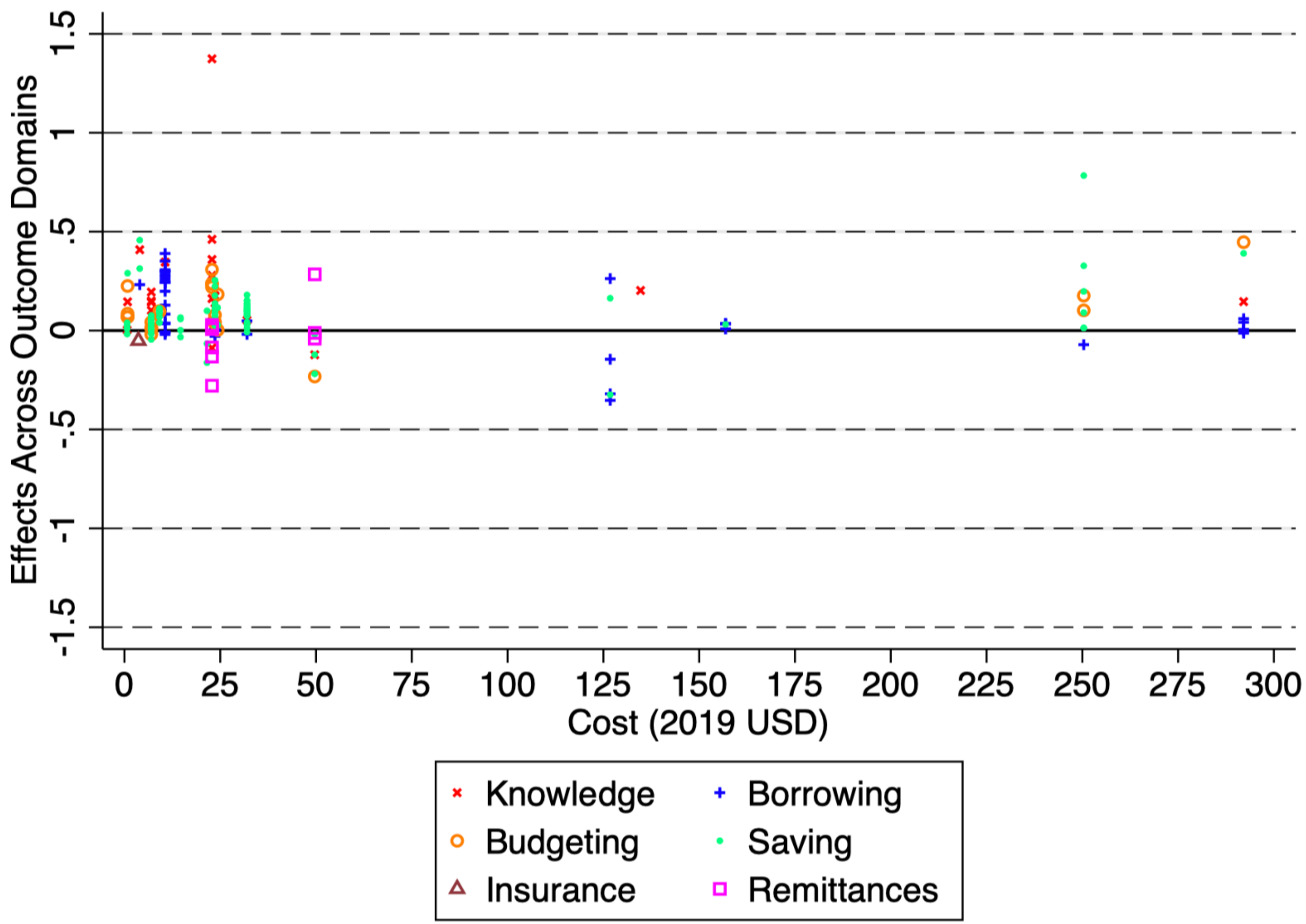
- Effects of financial education on *financial knowledge* are comparable to studies on math and reading (Hill et al. 2008; Cheung and Slavin 2016; Fryer 2016).
- Effects of financial education on *financial behaviors* are comparable to meta-analyses of behavior change interventions in other domains
  - anti-smoking (Rooney & Murray 1996)
  - tailored printed health interventions (Noar et al. 2017)
  - energy conservation (Karlin et al. 2015)

Interpreting effect sizes from causal studies (Kraft 2020, p. 250)

	Cost-Effectiveness Ratio				
	Cost Per Pupil				
Effect Size	Low (<$500)	Moderate ($500 to <$4,000)	High (≥$4,000)		Scalability
Small (<.05)	Small ES / low cost	Small ES / moderate cost	Small ES / high cost		Easy to scale
Medium (.05 to <.20)	Medium ES / low cost	Medium ES / moderate cost	Medium ES / high cost	&	Reasonable to scale
Large (≥.20)	Large ES / low cost	Large ES / moderate cost	Large ES / high cost		Hard to scale

Note. Green and red shading represent higher and lower cost-effectiveness ratios, respectively. Effect size and cost benchmarks provide empirically informed starting places that should be adapted based on the characteristics of individual studies. ES = effect size.

Costs and effect sizes of financial education interventions



Are interventions cost-effective?

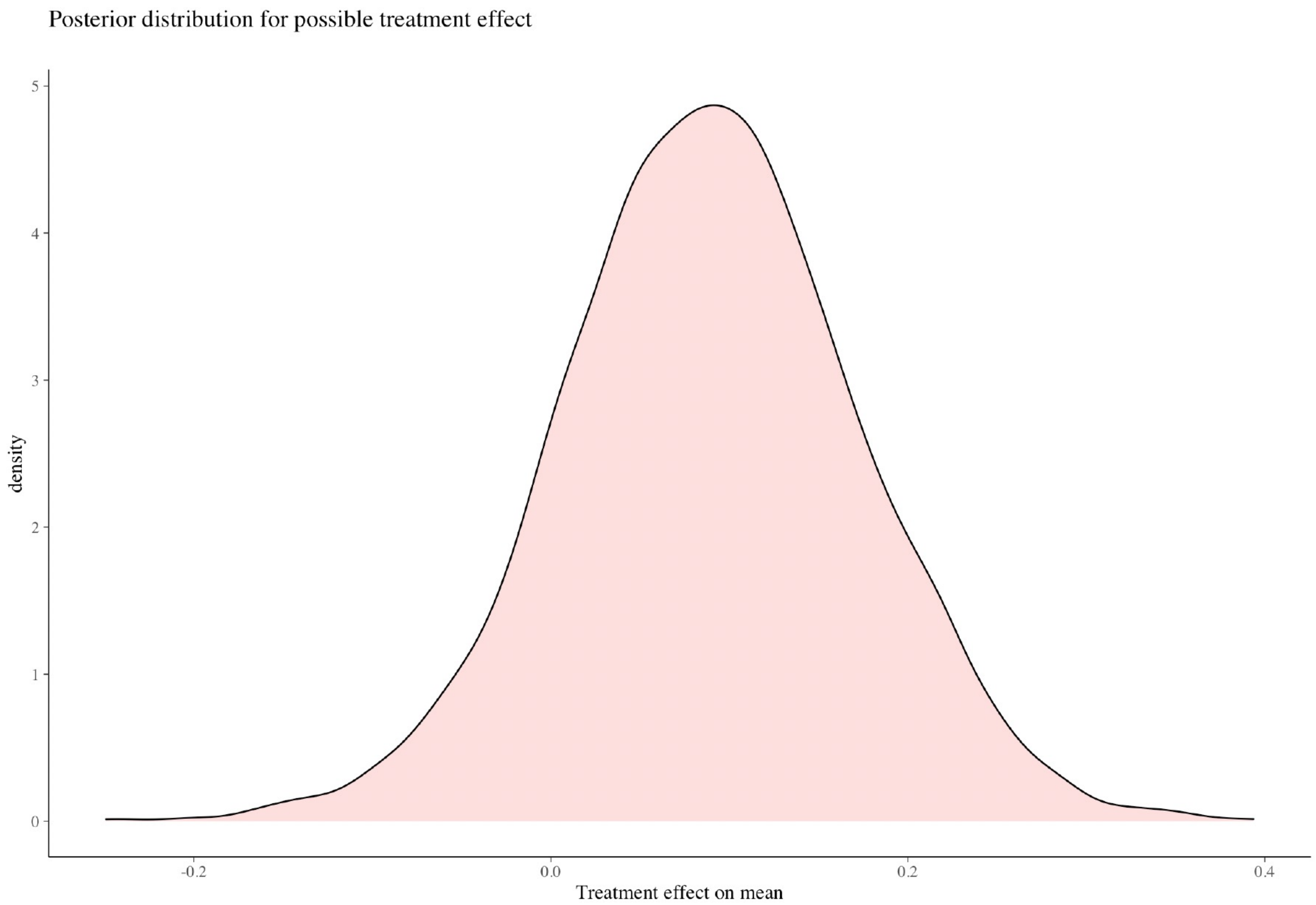
- Using Kraft’s (2019) scale of educational interventions, effects are "medium/large.”
- Average intervention has low cost per participant (mean costs are $60.40 and median costs are $22.90)
- With the data we have, for "medium effect sizes," Kraft’s educational intervention scale would say average cost per participant of $60 implies "low cost.”

Are effects fading out?

Subgroup	Effect size (g)	SE	95% CI Lower bound	95% CI Upper bound	n(Studies)	n(effects)
Panel A: Treatment effects on *financial behaviors*						
(a) By country income						
High income economies	0.1127	0.0316	0.0478	0.1777	32	129
Developing economies	0.0928	0.0130	0.0660	0.1195	32	329
(b) By respondent income						
Low-income individuals	0.0993	0.0194	0.0600	0.1387	43	367
General population	0.1035	0.0219	0.0571	0.1500	21	91
(c) By age of participants						
Children (< age 14)	0.0640	0.0186	0.0188	0.1091	9	36
Youth (age 14 to 25)	0.1203	0.0415	0.0250	0.2155	11	92
Adults (> age 25)	0.1068	0.0205	0.0653	0.1483	44	330
(d) By intensity of treatment						
< 5 hours	0.0817	0.0194	0.0407	0.1227	22	124
≥ 5 and < 20 hours	0.0992	0.0223	0.0533	0.1450	29	251
≥ 20 hours	0.2319	0.0664	0.0745	0.3893	8	54
(e) By delay between treatment and measurement of outcomes						
< 6 months	0.0991	0.0169	0.0645	0.1337	34	180
≥ 6 and < 18 months	0.0901	0.0181	0.0520	0.1283	23	211
≥ 18 months	0.0653	0.0192	0.0209	0.1098	10	49
((f) By type of intervention						
Classroom	0.1064	0.0181	0.0699	0.1428	50	331
Online	0.0796	0.0336	-0.0194	0.1786	5	55
Counseling	0.1595	0.0274	-0.1887	0.5077	2	48
Educative Nudge	0.0597	0.0206	0.0055	0.1138	8	24

Estimated posterior distribution of treatment effects on financial behaviors

Figure B.8: Posterior distribution of possible treatment effects on financial behaviors (BHM)



Notes: This figure shows the posterior distribution of possible treatment effects (in SD units) on the set of 64 studies with financial behaviors as the outcome. The result is based on the partial pooling Rubin model (with default priors) as specified in Table B4 with a posterior mean of 0.09.

RPTU

What works in financial education? Some observations from the literature (1/2)

- Treatment effects depend on responses from **heterogeneous consumers** (Lusardi, Michaud, and Mitchell 2017, JPE)

- **Implementation matters:** Long-term impacts depend on teacher training and delivery quality (Bruhn et al. 2016, AEJ: Applied; Brown et al. 2016, RFS; Urban et al. 2018, EconEdRev).

- **Keep it simple?:** “Rules of thumb” approaches yield larger effects (Drexler et al. 2014, AEJ: Applied, Skimmyhorn et al. 2016).

- **Tailoring interventions:** Targeted content improves relevance and outcomes (e.g., Doi et al. 2014; Seshan & Yang 2014; Abarcar et al. 2020, JDE).

What works in financial education? Some observations from the literature (2/2)

- **Personalization helps:** Counseling and individual feedback enhance effectiveness (Carpena et al. 2017, ManSci).

- **Innovative delivery formats:**
  - Mass media for broader reach and attitude shifts (Berg & Zia 2017, JEEA, Chopra 2023)
  - Experiential learning and debiasing approaches (Abel et al. 2020, WBER)
  - Digital tools and gamification (Attanasio et al. 2019; Sconti 2022)
  - …
- **Pedagogical innovations:** Active learning and group exercises outperform traditional lectures (Kaiser & Menkhoff 2022, JDE).

- **Decentralized delivery:** Localized instruction can be effective (Hakizimfura et al. 2020, JDE).

Open questions

- **What works, for whom, and – most importantly - why?**
- What are the causal mechanisms for behavior change? (Sayinzoga et al. 2016; Carpena and Zia 2020; Horn et al. 2020; Kaiser et al. 2022 WP)
- Why are some behaviors easier to change then others? (Kaiser et al. 2022, JFE)
- How sustainable are effects? (Horn et al. 2023, REStat, Bruhn et al. 2022; Frisancho 2025)
- What about scalability?
- What about general equilibirum effects? (e.g., Kosfeldt and Schuewer 2017, Ferreira ERC grant)
- Beyond directional changers in behavior: What are the welfare implications? (Ambuehl et al. 2022, AER; *Boyer et al. 2022; REStat)*)
- …

Experiment on Scalability

Financial education programs are implemented worldwide…

- The implementation of financial education programs has become a high priority for policymakers around the world (OECD 2015).

- Most evidence is from **randomized controlled trials (RCTs)** in developing economies, often with self-employed individuals (Kaiser, Lusardi, Menkhoff and Urban 2022, *JFE)*

- Financial education programs in developing economies are effective in…
  - improving **financial practices** of firms (Drexler et al. 2014*, AEJ: Applied*)
  - stimulating **investment** into micro-enterprise and diversification
  - fostering **household saving** and shifting savings to formal financial products (Doi et al. 2014, *JDE*, Bruhn et al. 2016, *AEJ: Applied;* Carpena et al. 2017, *ManSci;* Attanasio et al. 2019; Horn et al. 2023, *ReStat*)
  - improving borrowing behavior and **credit outcomes** (Berg and Zia 2017, *JEEA;* Bruhn et al. 2022)

…but RCTs are usually run on a small scale

- Median sample size is about 840 individuals in previous meta-analysis of 76 RCTs (Kaiser, Lusardi, Menkhoff and Urban 2022, *JFE)*
- But how do effects change when programs are run at a larger scale?
- Evidence for **smaller treatment effects at scale (i.e., "voltage drop")** in a variety of contexts (DellaVigna and Linos 2022, ECTA), Al-Ubaydli et al. 2017, 2019, 2023, Banerjee et al. 2017, Muralidharan and Niehaus 2017)

- Threats to scalability :
    - Statistical inference (e.g., small study effects and publication bias) (Andrews and Kasy 2018, AER, DellaVigna and Linos 2022, ECTA)
    - External validity (i.e., differences in response to treatment, take-up etc.)
    - Spillovers on the non-treated and the treated (Mckenzie and Puerto 2020, AEJ Applied)
    - General equilibirum effects
    - …

RPTU

New paper

- **Two-stage randomized saturation experiment** evaluating a financial education program in rural Uganda
- We exogenously vary the share of treated respondents relative to the target population
- In an affine model of treatment saturation, we find large effects on **mobile mone use**, **business investment**, and **saving behavior** on the *uniquely trated* but effects decline with increasing saturation
- We find no evidence of spillovers on the non-treated but the above phenomenon suggests **externalities on the treated (i.e., a specific form of “voltage drop”).**

CESifo WORKING PAPERS
11431
2024
October 2024

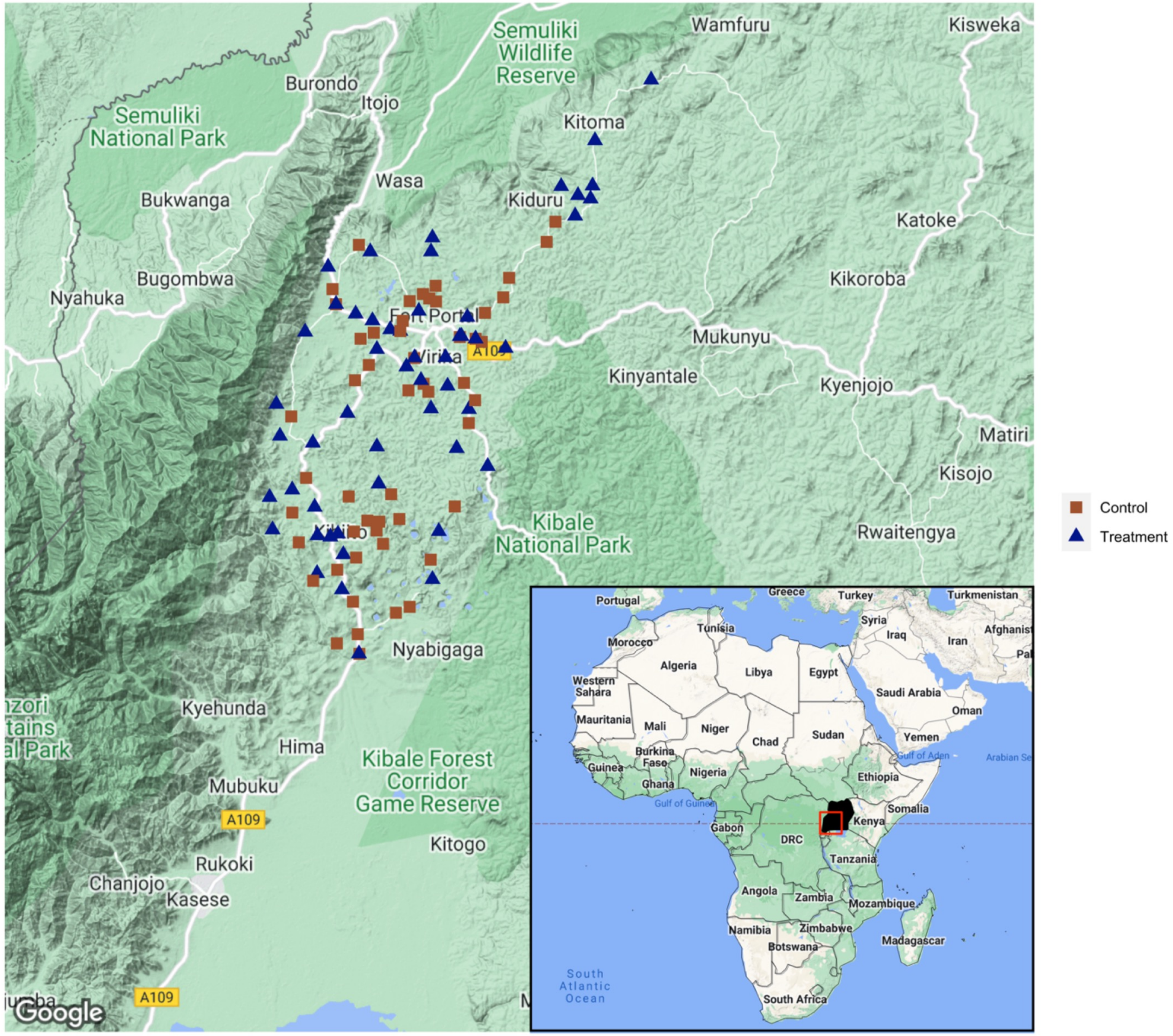
Scaling Financial Education Among Micro-Entrepreneurs: A Randomized Saturation Experiment

Jana S. Hamdan, Tim Kaiser, Lukas Menkhoff, Yuanwei Xu

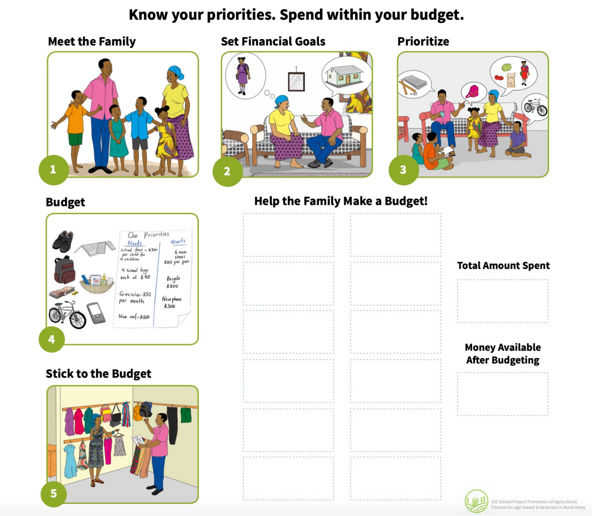
CESifo

The financial education intervention

- Cluster-RCT in 108 villages in Uganda
- Randomized half of the villages to a full-day financial education intervention developed by GIZ and Bank of Uganda with the following topics:
    - (i) budgeting and personal financial management,
    - (ii) saving and future consumption
    - (iii) credit and borrowing decisions
    - (iv) business investing
    - (v) mobile payments
- Training of trainers by Bank of Uganda
- Activities include group problem-solving and sharing of personal experiences
- Pilot previously evaluated in a separate sample RCT (efficacy trial) (Kaiser & Menkhoff 2022, JDE)



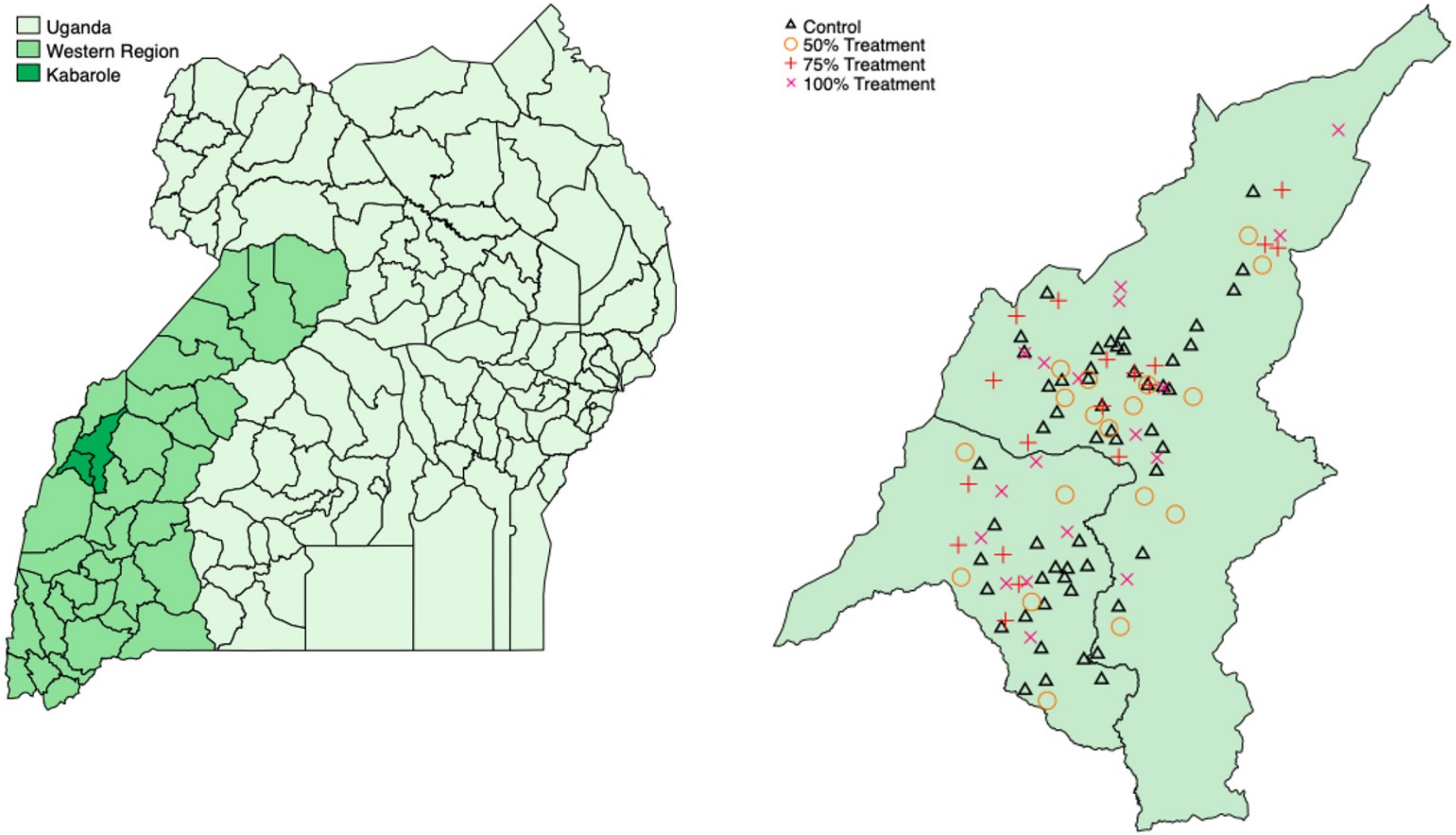
Impressions from the field



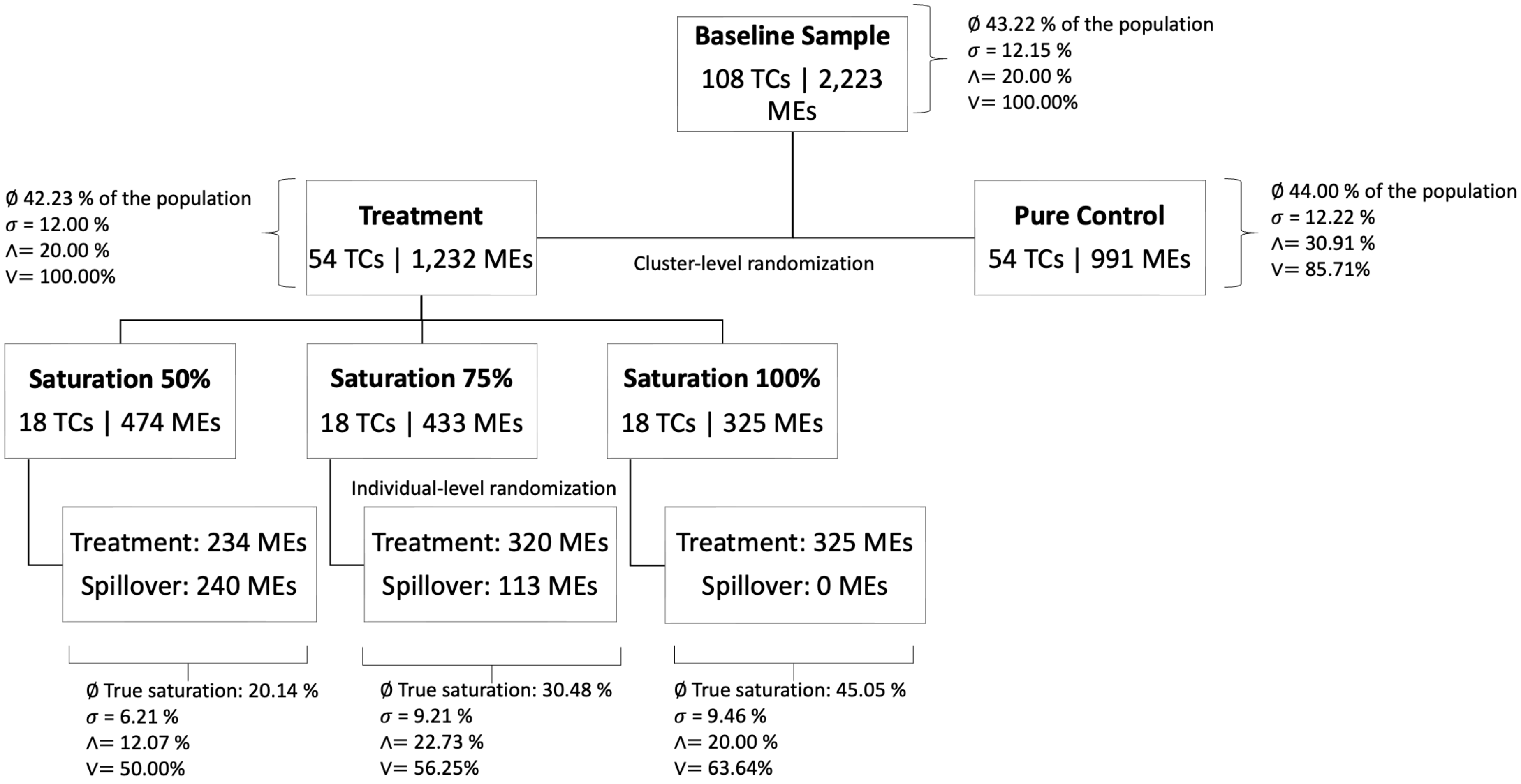
Field experiment setting

(a) Kabarole District

(b) Sampled Trading Centers

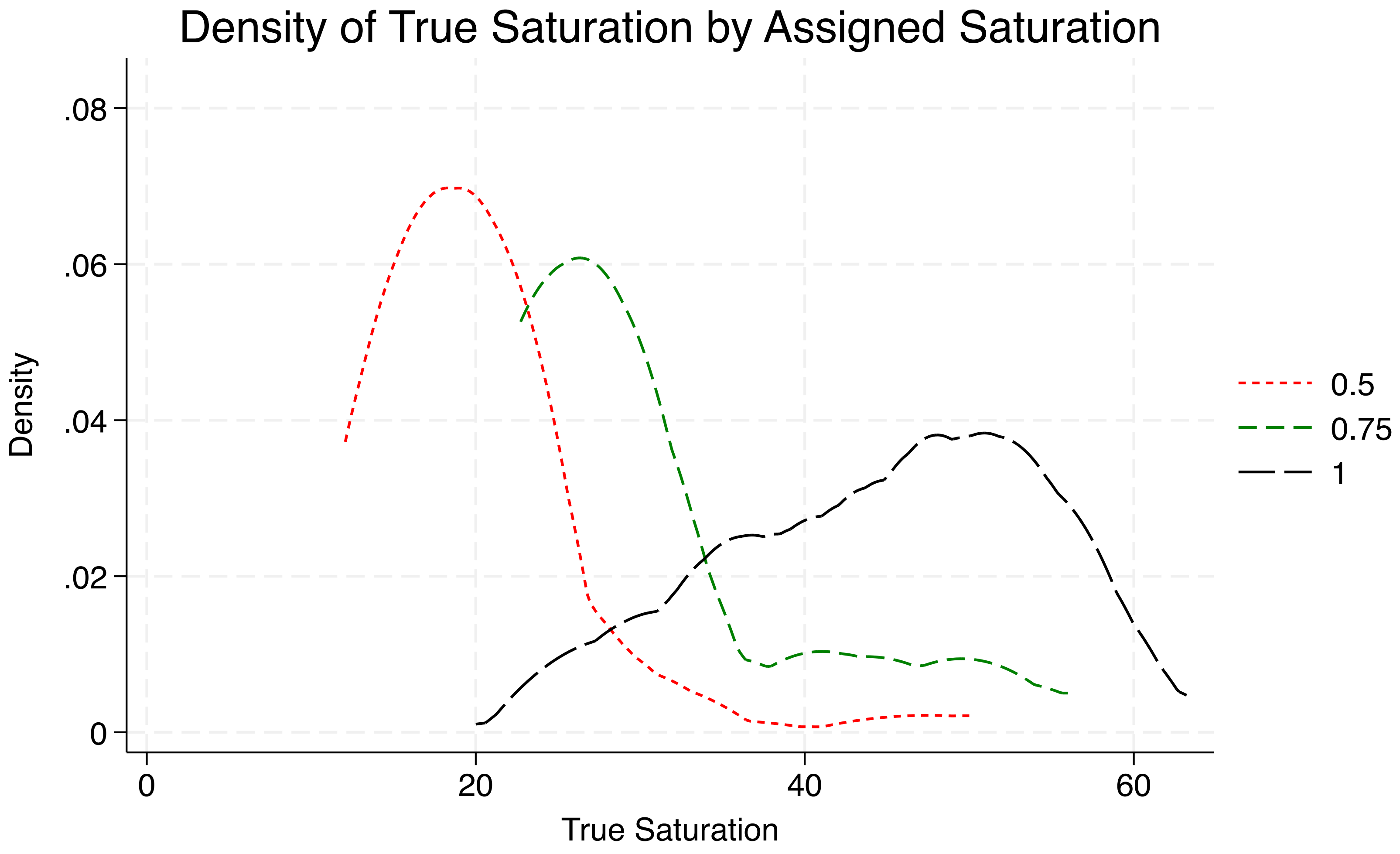


Randomized Saturation Experiment



RPTU

Distribution of true saturation by assigned saturation



RPTU

Timeline

	Q1/´19	Q2/´19	Q3/´19	Q4/´19	Q1/´20	Q2/´20	Q3/´20	Q4/´20	Q1/´21	Q2/´21
Baseline (field)	X	X								
Treatment			X							
Follow-Up (phone)								X		
Follow-Up (field)										X

Empirical strategy (see Baird et al. 2018, REStat)

- **Pooled Intention to treat and spillover effects**

$$y_{i,c(t)} = \alpha_0 + \beta_1 T_{i,c} + \beta_2 S_{i,c} + \delta_1 y_{ic(t-1)} + \lambda X_{s,i,c} + \varepsilon_{ict}$$

- **Non-parametric analysis**

$$y_{i,c(t)} = \alpha_0 + \beta_1 T_{i,c} \times 1(\pi_c = 50\%) + \beta_2 S_{i,c} \times 1(\pi_c = 50\%) + \beta_3 T_{i,c} \times 1(\pi_c = 75\%) + \beta_4 S_{i,c} \times 1(\pi_c = 75\%) + \beta_5 T_{i,c} \times 1(\pi_c = 100\%) + \delta_1 y_{ic(t-1)} + \lambda X_{s,i,c} + \varepsilon_{ict}$$

- **Linear saturation analysis**

$$y_{i,c(t)} = \alpha_0 + \beta_1 T_{i,c} + \beta_2 S_{i,c} + \beta_3 (T_{i,c} \times \Pi_c) + \beta_4 (S_{i,c} \times \Pi_c) + \delta_1 y_{ic(t-1)} + \lambda X_{s,i,c} + \varepsilon_{ict}$$

Partial endogeneity of true saturation (due to oversampling of population in smaller clusters). Thus, We run the above models in an IV-Setup where we instrument the true saturation Π_c with the assigned saturation π_c.

Results

Pooled ITT and Spillover Effects on Aggregate Indices

	(1) Mo. Money	(2) Savings	(3) Investment	(4) Budgeting	(5) Borrowing	(6) Summary
Panel A: Pooled Intention to Treat and Spillover Effects						
Assigned to Training	0.040	0.149***	0.103*	0.037	-0.045	0.092
	(0.073)	(0.054)	(0.054)	(0.061)	(0.046)	(0.058)
Spillover Group	-0.135*	-0.041	-0.021	-0.037	-0.077	-0.101
	(0.076)	(0.060)	(0.063)	(0.065)	(0.053)	(0.064)
Test of equality ($p - value$)	0.031**	0.004***	0.034**	0.183	0.499	0.002***

Notes: The dependent variables in column 1-6 are all standardized to have a mean of 0 and a standard deviation of 1 in the pure control group. Panel A shows intention to treat, and spillover effects pooled across all saturations (see Figure 2 for the distribution and support of the saturations).

Non-Parametric Analysis of True Saturation

	(1) Mo. Money	(2) Savings	(3) Investment	(4) Budgeting	(5) Borrowing	(6) Summary
Panel C: Non-Parametric Analysis of True Saturation						
Assigned to Training ×	0.193	0.214**	0.194**	0.069	-0.002	0.232**
$1(\Pi_c \geq 12\% < 30\%)$	(0.128)	(0.087)	(0.080)	(0.084)	(0.069)	(0.102)
Spillover Group ×	0.061	0.067	0.207*	0.108	-0.052	0.145
$1(\Pi_c \geq 12\% < 30\%)$	(0.135)	(0.114)	(0.115)	(0.097)	(0.083)	(0.118)
Assigned to Training ×	0.033	0.183	0.144	0.138	-0.106	0.132
$1(\Pi_c \geq 30\% \leq 45\%)$	(0.143)	(0.121)	(0.143)	(0.170)	(0.119)	(0.133)
Spillover Group ×	-1.535	-1.243	-1.767	-1.523	-0.292	-2.223
$1(\Pi_c \geq 30\% \leq 45\%)$	(1.419)	(1.089)	(1.343)	(1.141)	(0.680)	(1.498)
Assigned to Training ×	-0.235*	-0.025	-0.043	-0.012	-0.121*	-0.174
$1(\Pi_c > 45\% \leq 64\%)$	(0.126)	(0.116)	(0.124)	(0.141)	(0.071)	(0.134)
Spillover Group ×	-0.250	0.191	-0.432**	-0.056	0.209	-0.059
$1(\Pi_c > 45\% \leq 64\%)$	(0.281)	(0.249)	(0.193)	(0.178)	(0.188)	(0.252)

Notes: Panel C shows a non-parametric analysis of intention to treat and spillover effects for three bins of *true saturations* ($\Pi_c \geq 12\% < 30\%$), ($\Pi_c \geq 30\% \leq 45\%$), ($\Pi_c > 45\% \leq 64\%$). In this analysis, the true saturation (Π_c) is instrumented with the randomly assigned saturation (π_c)

RPTU

Affine Model of True Saturation

	(1) Mo. Money	(2) Savings	(3) Investment	(4) Budgeting	(5) Borrowing	(6) Summary
Panel D: Linear Analysis of True Saturation						
Assigned to Training	0.432*	0.447***	0.220	0.071	-0.029	0.370**
	(0.229)	(0.159)	(0.159)	(0.158)	(0.120)	(0.180)
Assigned to Training ×	-0.013*	-0.010**	-0.003	-0.001	-0.001	-0.009*
Π_c (% True Saturation)	(0.006)	(0.005)	(0.005)	(0.005)	(0.003)	(0.005)
Spillover Group	0.295	0.227	0.532**	0.393	-0.055	0.451*
	(0.339)	(0.263)	(0.228)	(0.246)	(0.190)	(0.258)
Spillover Group ×	-0.016	-0.010	-0.022**	-0.017*	-0.001	-0.021**
Π_c (% True Saturation)	(0.014)	(0.010)	(0.009)	(0.011)	(0.008)	(0.011)

Notes: Panel D shows results of a parametric analysis of treatment saturation, i.e., an affine model of how intention to treat and spillover effects change linearly with increasing true saturation. The true saturation (Π_c) is instrumented with the assigned saturation (π_c).

In economic terms

	Saving Index					
	(1) Any Savings	(2) Savings Amt. †	(3) Any MM Savings	(4) MM Savings Amt. †	(5) Any Bank Savings	(6) Bank Savings Amt.†
Control group mean (SD)	0.853	11.526 (5.247)	0.192	2.243 (4.705)	0.181	2.332 (5.222)
Assigned to Training	0.044	0.317	0.185***	1.945***	0.145***	1.919***
	(0.056)	(0.739)	(0.056)	(0.718)	(0.053)	(0.721)
Assigned to Training ×	-0.001	-0.002	-0.004**	-0.043**	-0.003**	-0.044**
Π_c (% True Saturation)	(0.002)	(0.021)	(0.002)	(0.021)	(0.001)	(0.020)
Spillover Group	0.027	0.891	0.140	1.614	0.002	0.433
	(0.096)	(1.505)	(0.090)	(1.090)	(0.094)	(1.298)
Spillover Group ×	-0.001	-0.051	-0.006*	-0.070*	0.000	-0.013
Π_c (% True Saturation)	(0.004)	(0.061)	(0.003)	(0.042)	(0.004)	(0.057)

Table 5: Probing mechanisms behind saturation effects

	(1) Take-Up	(2) Left early (conditional on take-up)	(3) Class-Size	(4) Fully adequate venue (binary rating)	(5) Outdoor venue	(6) Quality rating (training)
Reference group mean (SD)	0.000	0.013	15.977	0.752	0.563	9.135
			(6.996)			(0.202)
$1(\pi_c = 50\%)$						
$1\ (\pi_c = 75\%)$		-0.006	-2.240	-0.303**	0.094	0.053
		(0.015)	(1.550)	(0.115)	(0.187)	(0.050)
$1\ (\pi_c = 100\%)$		0.027*	2.121	-0.298**	0.310*	-0.042
		(0.016)	(2.166)	(0.123)	(0.178)	(0.084)
Assigned to Training (T) × $1(\pi_c = 50\%)$	0.708***					
	(0.053)					
Spillover Group (S) × $1(\pi_c = 50\%)$	0.302***					
	(0.057)					
Assigned to Training (T) × $1(\pi_c = 75\%)$	0.705***					
	(0.039)					
Spillover Group (S) × $1(\pi_c = 75\%)$	0.393***					
	(0.061)					
Assigned to Training (T) × $1(\pi_c = 100\%)$	0.729***					
	(0.033)					
Tests of equality (p-values):						
T × $(\pi_c = 50\%)$ = T × $(\pi_c = 75\%)$	0.954					
T × $(\pi_c = 50\%)$ = T × $(\pi_c = 100\%)$	0.742					
T × $(\pi_c = 75\%)$ = T × $(\pi_c = 100\%)$	0.648					
S × $(\pi_c = 50\%)$ = S × $(\pi_c = 75\%)$	0.266					
$1(\pi_c = 50\%) = 1\ (\pi_c = 75\%)$		0.675	0.154	0.011	0.618	0.293
$1(\pi_c = 50\%) = 1\ (\pi_c = 100\%)$		0.094	0.332	0.018	0.088	0.622
$1(\pi_c = 75\%) = 1\ (\pi_c = 100\%)$		0.069	0.046	0.972	0.233	0.315
N (individuals)	1,975	651	1,113	1,113	1,113	1,113
N (clusters)	108	54	54	54	54	54

Understanding the “voltage drop” in our setting

Selection

- Scaling is faced with potentially negative effects from selection of individuals, i.e., as a larger share is invited, peer effects may lead individuals to attend the sessions but be inattentive (i.e., driving down ITT and LATE) → no evidence in out setting.

Quality of training

- As a larger share of the population is being treated and the absolute number of treated increase, the quality may deteriorate. → <u>Classroom data does not suggest differences in quality by saturation</u>
- Class-size is not correlated with saturation in our setting, as we offered multiple sessions in high-saturation clusters
- Some evidence for lower quality venue

Understanding the “voltage drop” in our setting

Crowding out

- Improved financial practices may give micro-enterprises a competitive advantage
- Firms can improve their product and marketing activities in sectors allowing such differentiation (i.e., non-homogenous goods)

→ Saturation effect is <u>stronger for firms in service and manucfaturing relative to firms in retail.</u>

Summary

- Large effects if few individuals are treated per cluster (i.e., mobile money use, savings, investments)
- Evidence for a “voltage drop” in effectiveness if program is operated at a larger scale
- No evidence for negative spillovers on untreated peers
- Voltage drop likely the result of both social dynamics as well as institutional constraings

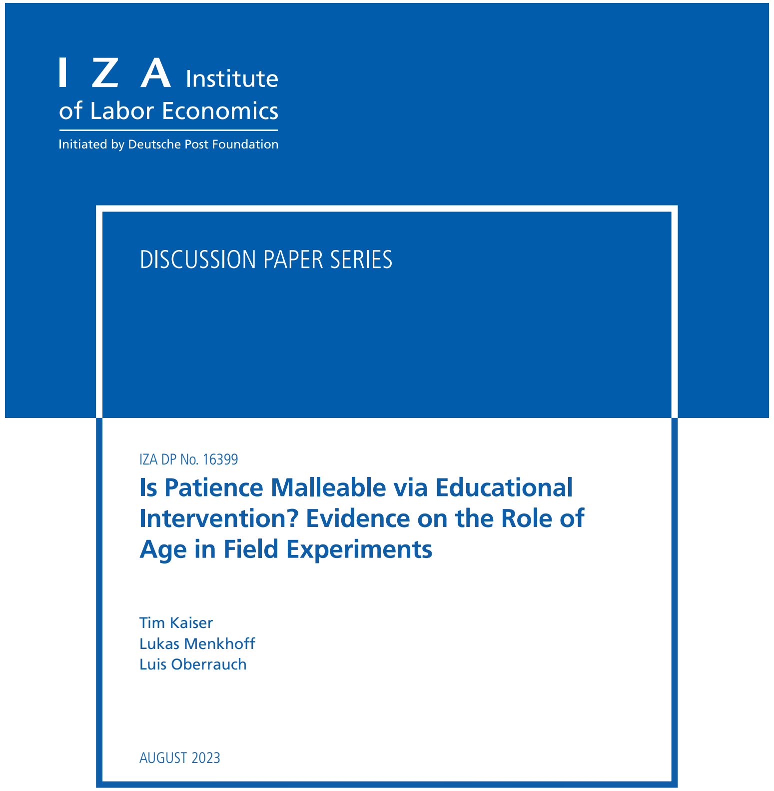
RPTU

Understanding causal mechanisms

New paper

1) We conduct **an RCT studying the effects of a financial education intervention** on time-preferences of both youth and adults in Uganda using the CTB protocol (Andreoni and Sprenger 2012, AER).
   - Study heterogenous treatment effects by age of respondents

2) We **combine our data with data from 10 earlier field experiments** studying the causal effects of (financial-) education interventions on impatience measured in incentivized tasks.
   - Meta-Analysis
   - Study the role of student age (and contextual features of the intervention) in explaining the heterogeneity in treatment effects across studies
   - Study the generalizability of (heterogenous) treatment effects


I Z A Institute of Labor Economics
Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 16399

Is Patience Malleable via Educational Intervention? Evidence on the Role of Age in Field Experiments

Tim Kaiser
Lukas Menkhoff
Luis Oberrauch

AUGUST 2023

Preview of results

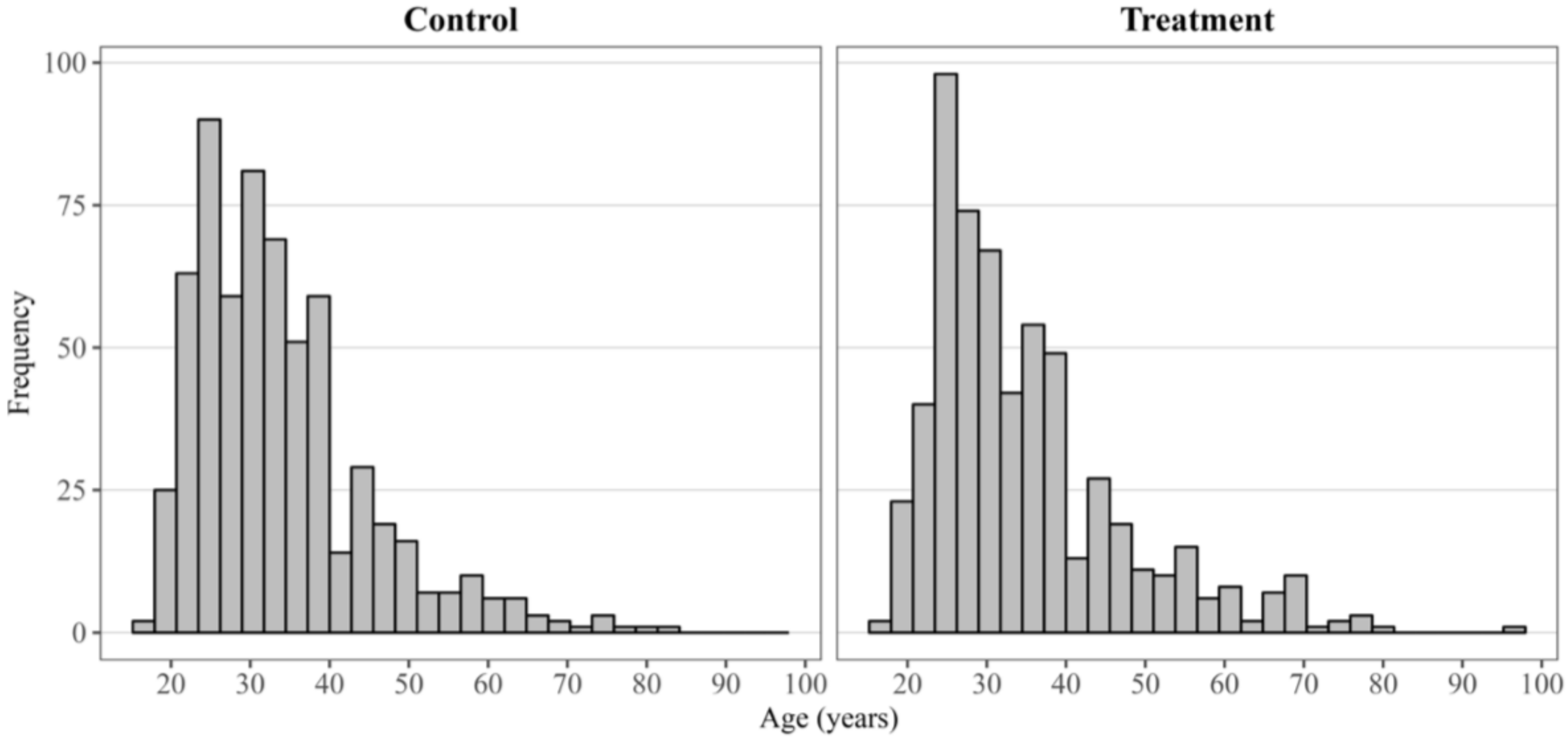
Field experiment:

- **Heterogenous effects** by age: adults' impatience measured in incentivized tasks is unaffected by the intervention after 15 months follow-up, but we observe **large effects on impatience, estimated discount factors,** and **field saving behavior for youth** in our setting

Meta-study:

- On average, the effect of interventions on reducing impatience may be positive but uncertain (-0.05 sigma). **The age of students and fade-out** appear to explain a large share of between-study heterogeneity in treatment effects.

Age distributions are balanced



RPTU

Time preference elicitation design (Carvalho et al. 2016, JDE)

Panel A: Time preference elicitation design

Budget	Sooner endowment (UGX)	Interior choice (split endowments) (UGX)	Later endowment (UGX)	t	$t+k$	$1+r$
1	5,400 | 0	2,700 | 3,000	0 | 6,000	0	1	1.11
2	5,400 | 0	2,700 | 3,000	0 | 6,000	1	2	1.11
3	5,000 | 0	2,500 | 3,000	0 | 6,000	1	2	1.20
4	5,000 | 0	2,500 | 3,000	0 | 6,000	1	6	1.20

- Conducted via phone and using mobile money
- Adding "thank-you payments" in two installments (500 UGX sooner and 500 UGX later) regardless of the experimental choices to equalize transaction costs

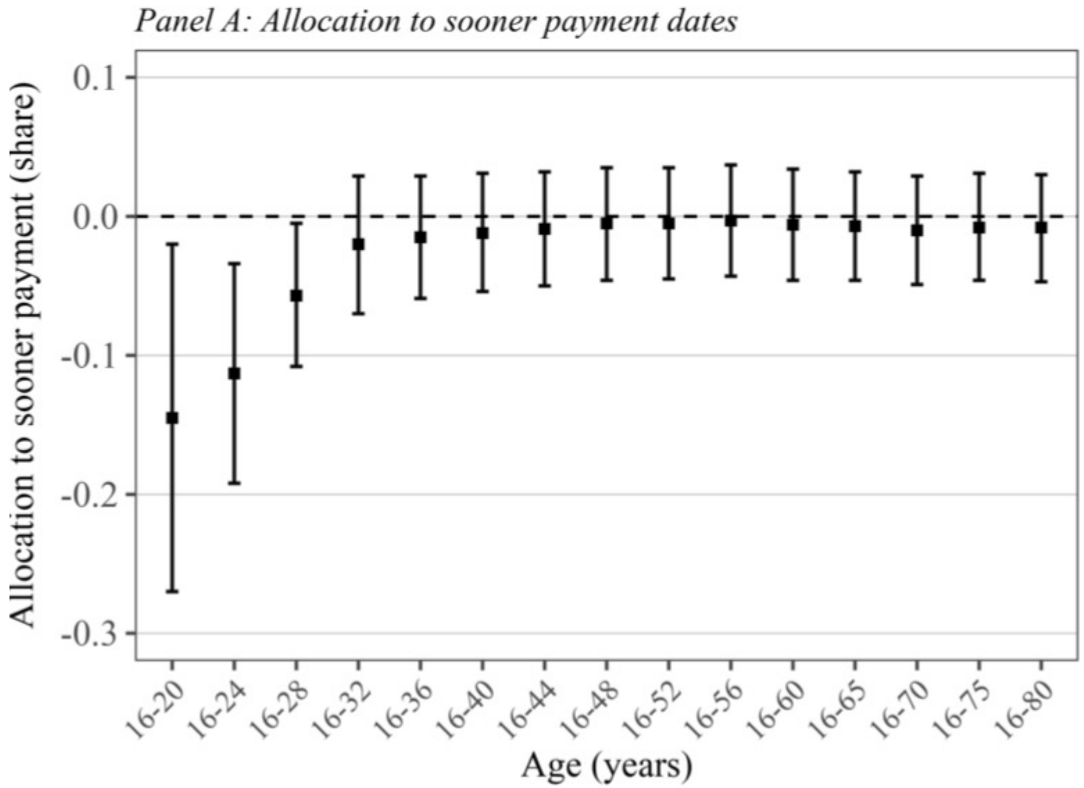
- Outcome variables:
  i. Share of the budget allocated to the sooner payment date
  ii. Binary indicator of choosing the early option (at the-choice-level)
  iii. Estimated individual discount factor $\hat{\delta}$ (and present bias $\hat{\beta}$) from a standard beta-delta utility function (Laibson 1997, QJE)

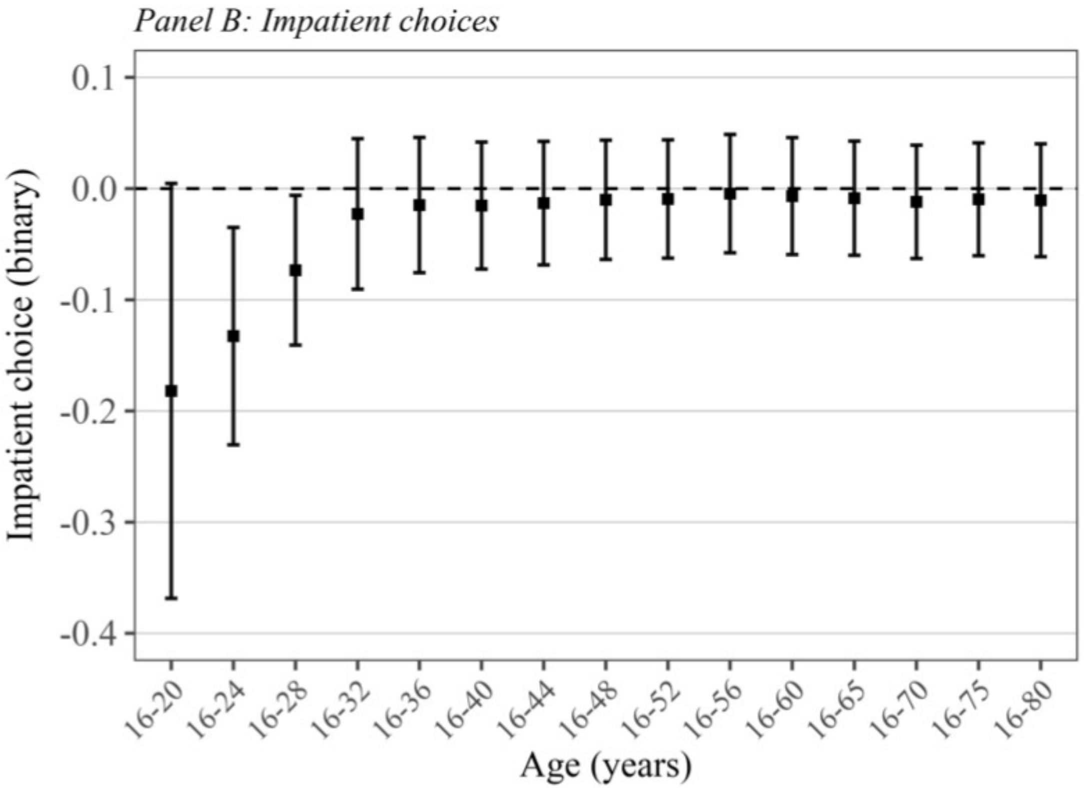
Treatment effects on allocation behaviors

	Average treatment effects (full sample)		Heterogeneous treatment effects (≤ 24 years of age)		Heterogeneous treatment effects (> 24 years of age)	
Panel A: Treatment effects on allocation behaviors						
	(1) Allocation to sooner payment (share)	(2) Impatient Choice (binary)	(3) Allocation to sooner payment (share)	(4) Impatient Choice (binary)	(5) Allocation to sooner payment (share)	(6) Impatient Choice (binary)
Treatment	-0.016	-0.023	-0.146***	-0.172***	0.013	0.009
	(0.024)	(0.032)	(0.045)	(0.058)	(0.027)	(0.035)
	[0.329]	[0.329]	[0.017]	[0.017]	[0.622]	[0.622]
Today ($t = 0\ days$)	0.097***	0.126***	0.068***	0.089***	0.104***	0.133***
	(0.014)	(0.017)	(0.018)	(0.025)	(0.016)	(0.020)
Delay ($k = 150$ days)	0.109***	0.129***	0.064**	0.063*	0.119***	0.143***
	(0.013)	(0.016)	(0.029)	(0.035)	(0.013)	(0.017)
Interest rate ($1 + r = 1.2$)	-0.057***	-0.062***	-0.052***	-0.054***	-0.057***	-0.064***
	(0.007)	(0.009)	(0.016)	(0.020)	(0.008)	(0.011)
Treatment × Today	-0.015	-0.013	0.006	-0.007	-0.020	-0.015
	(0.017)	(0.022)	(0.027)	(0.038)	(0.019)	(0.025)
Treatment × Delay	0.021	0.033	0.089*	0.092	0.006	0.020
	(0.020)	(0.028)	(0.047)	(0.058)	(0.020)	(0.028)
Treatment × Interest rate	0.014	0.014	0.029	0.033	0.012	0.011
	(0.010)	(0.012)	(0.021)	(0.025)	(0.011)	(0.014)
Permutation p-value	0.325	0.454	0.005	0.006	0.616	0.779
Control mean	0.687	0.710	0.770	0.815	0.669	0.687
Standardized effect size	-0.045	-0.051	-0.505	-0.442	0.036	0.019

RPTU

Heterogenous treatment effects on allocation behaviors by age

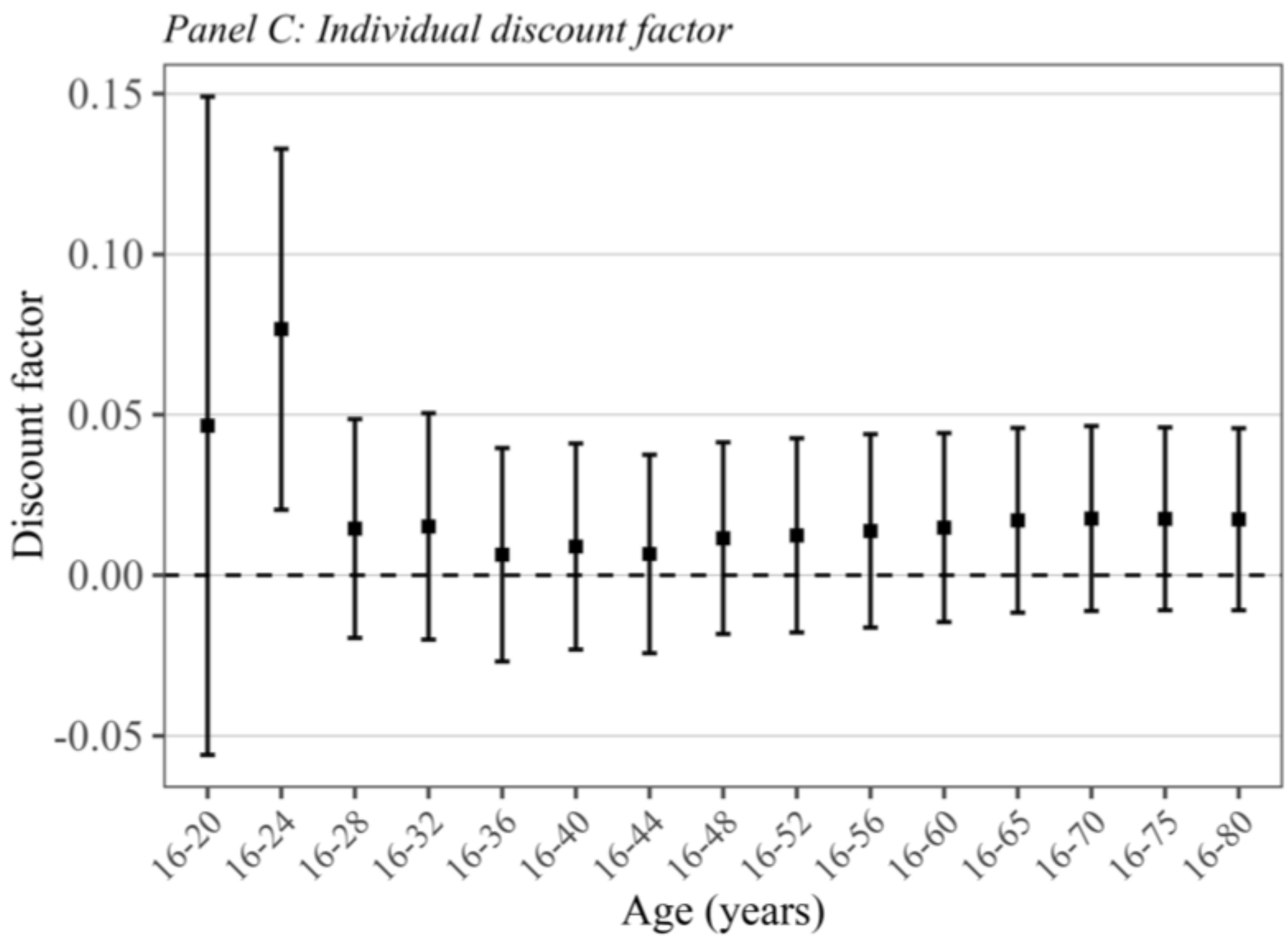




Treatment effects on utility parameters (Andersen 2008)

	Average treatment effects (full sample)		Heterogeneous treatment effects ($\leq$ 24 years of age)		Heterogeneous treatment effects (> 24 years of age)	
Panel B: Treatment effects on individual utility parameters						
	Discount factor $\widehat{\delta}_i$	Present bias $\widehat{\beta}_i$	Discount factor $\widehat{\delta}_i$	Present bias $\widehat{\beta}_i$	Discount factor $\widehat{\delta}_i$	Present bias $\widehat{\beta}_i$
Treatment	0.016	-0.007	0.077***	-0.022	0.004	-0.005
	(0.014)	(0.004)	(0.028)	(0.021)	(0.015)	(0.005)
	[0.313]	[0.175]	[0.017]	[0.313]	[0.789]	[0.510]
Permutation p-value	0.334	0.234	0.042	0.229	0.768	0.402
Control mean	1.063	0.993	1.030	0.999	1.064	0.995
Standardized effect size	0.079	-0.189	0.520	-0.275	0.020	-0.119

Heterogenous treatment effects by age: estimated individual discount factors



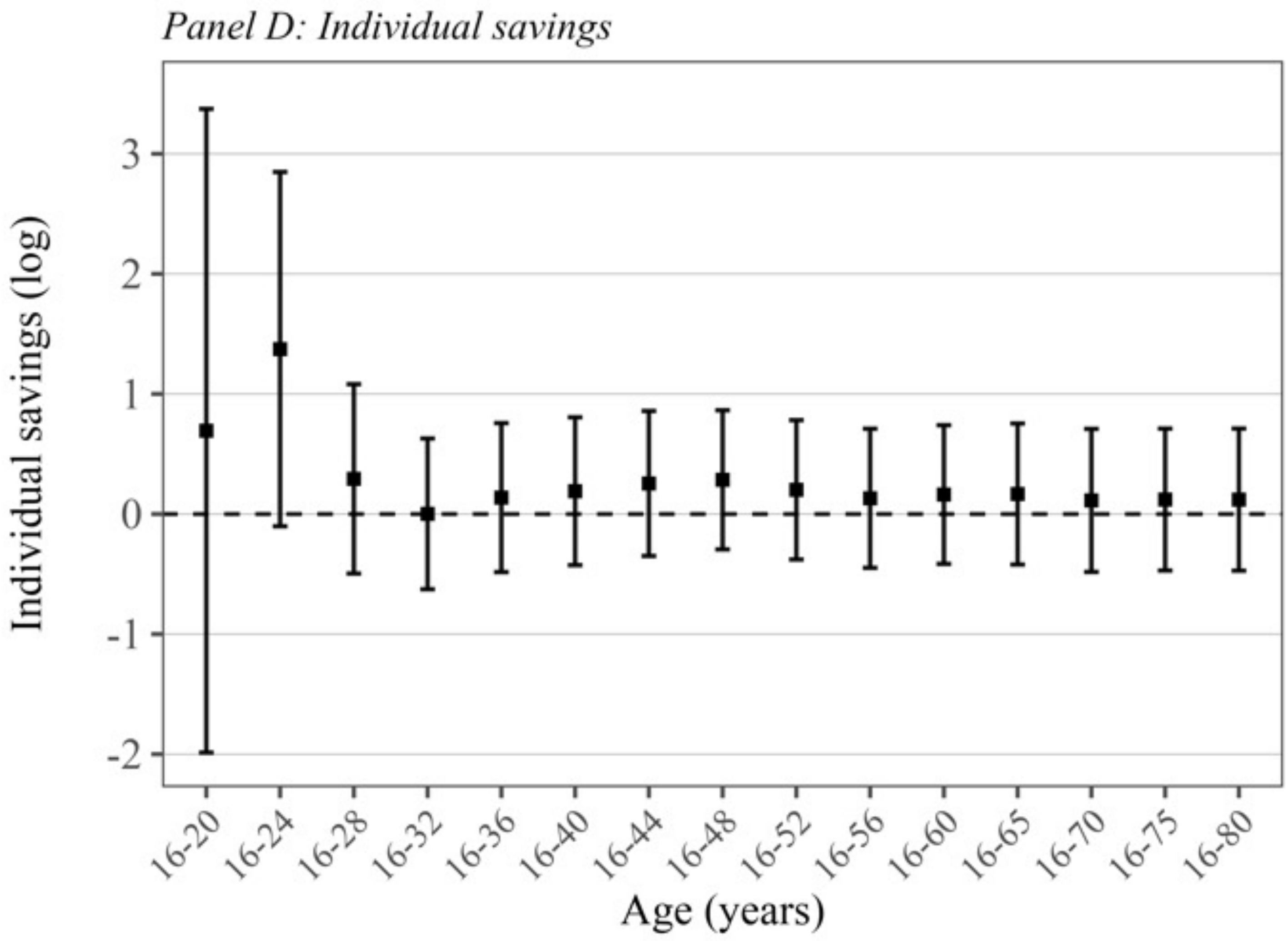
RPTU

Treatment effects on field saving behavior

	Average treatment effects (full sample)			Heterogeneous treatment effects (≤24 years of age)			Heterogeneous treatment effects (>24 years of age)		
	(1) Any	(2) log	(3) IHST	(4) Any	(5) log	(6) IHST	(7) Any	(8) log	(9) IHST
Treatment	0.010	0.132	0.133	0.102**	1.370*	1.432*	-0.008	-0.070	-0.078
	(0.021)	(0.298)	(0.313)	(0.051)	(0.744)	(0.780)	(0.021)	(0.300)	(0.315)
Control mean	0.843	10.781	11.354	0.821	10.392	10.950	0.847	10.865	11.442
Std. effect	0.027	0.026	0.025	0.266	0.256	0.255	-0.022	-0.014	-0.015
R^2	0.058	0.076	0.075	0.183	0.207	0.207	0.064	0.079	0.079
Observations	1,217	1,217	1,217	209	209	209	1,008	1,008	1,008
Clusters	108	108	108	81	81	81	107	107	107

Notes: All regressions included the lagged outcome at baseline and stratification fixed effects. Standard errors (in parentheses) are clustered at the village level. *** p<0.01, ** p<0.05, * p<0.1.

Treatment effects on field saving behavior



Ruling out alternative mechanisms

- ✓ No effects on **choice consistency**
- ✓ No evidence for decrease in **narrow bracketing** (intertemporal arbitrage)
- ✓ No treatment effects on **financial numeracy**
- ✓ Generally high **trust**, no evidence for treatment effects on trust, robust to excluding the few participants with low trust in payment
- ✓ No effects on income or background consumption (**liquidity constraints**)
- ✓ No difference **in take-up or attendence** by age
- ✓ No average or differential effects on **mobile money use**

➢ **Most plausible: A change in deep parameters? If so, is this finding general?**

Are these results generalizable across contexts? Meta-Analysis

Inclusion criteria

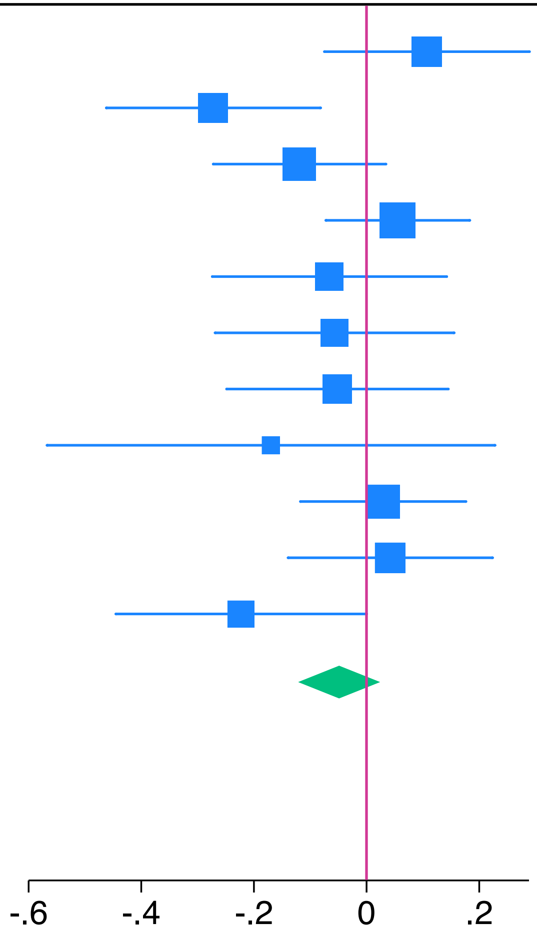
- RCT studying the effect of an educational intervention on a measure of impatience elicitied via incentivized decision experiments

Dataset:

- 11 RCTs and 45 treatment effect estimates
- Intensity ranges from **1 hour to 16 hours** of classroom exposure
- Within-study average age ranges from **8 to 49**
- Countries: Bangladesh, Germany, Italy, Liberia, Philippines, Spain, Tanzania, Turkey, Uganda
- Sample sizes from **165 to 4100**
- Delay between treatment and measurement of time preferences from **immediately after to about five years after**

Meta-Analysis: Forest Plot (Aggregated) (DV: TE on Impatience)

Study	Effect size with 95% CI	Weight (%)
Lührmann et al. 2018	0.11 [-0.07, 0.29]	9.51
Alan and Ertac 2018	-0.27 [-0.46, -0.08]	9.01
Bover et al. 2018	-0.12 [-0.27, 0.03]	11.51
Horn et al. 2020	0.06 [-0.07, 0.18]	13.66
Migheli and Moscarola	-0.07 [-0.27, 0.14]	8.03
Sutter et al. 2020	-0.06 [-0.27, 0.15]	7.83
Bjorvatn et al. 2020	-0.05 [-0.25, 0.14]	8.64
Berge et al. 2015	-0.17 [-0.57, 0.23]	2.91
Breitkopf et al. 2022	0.03 [-0.12, 0.18]	12.02
Blattman et al. 2017	0.04 [-0.14, 0.22]	9.54
Kaiser et al. 2023	-0.22 [-0.45, -0.00]	7.34
Overall	-0.05 [-0.12, 0.02]	



Heterogeneity: $\tau^2 = 0.01$, $I^2 = 40.00\%$, $H^2 = 1.67$

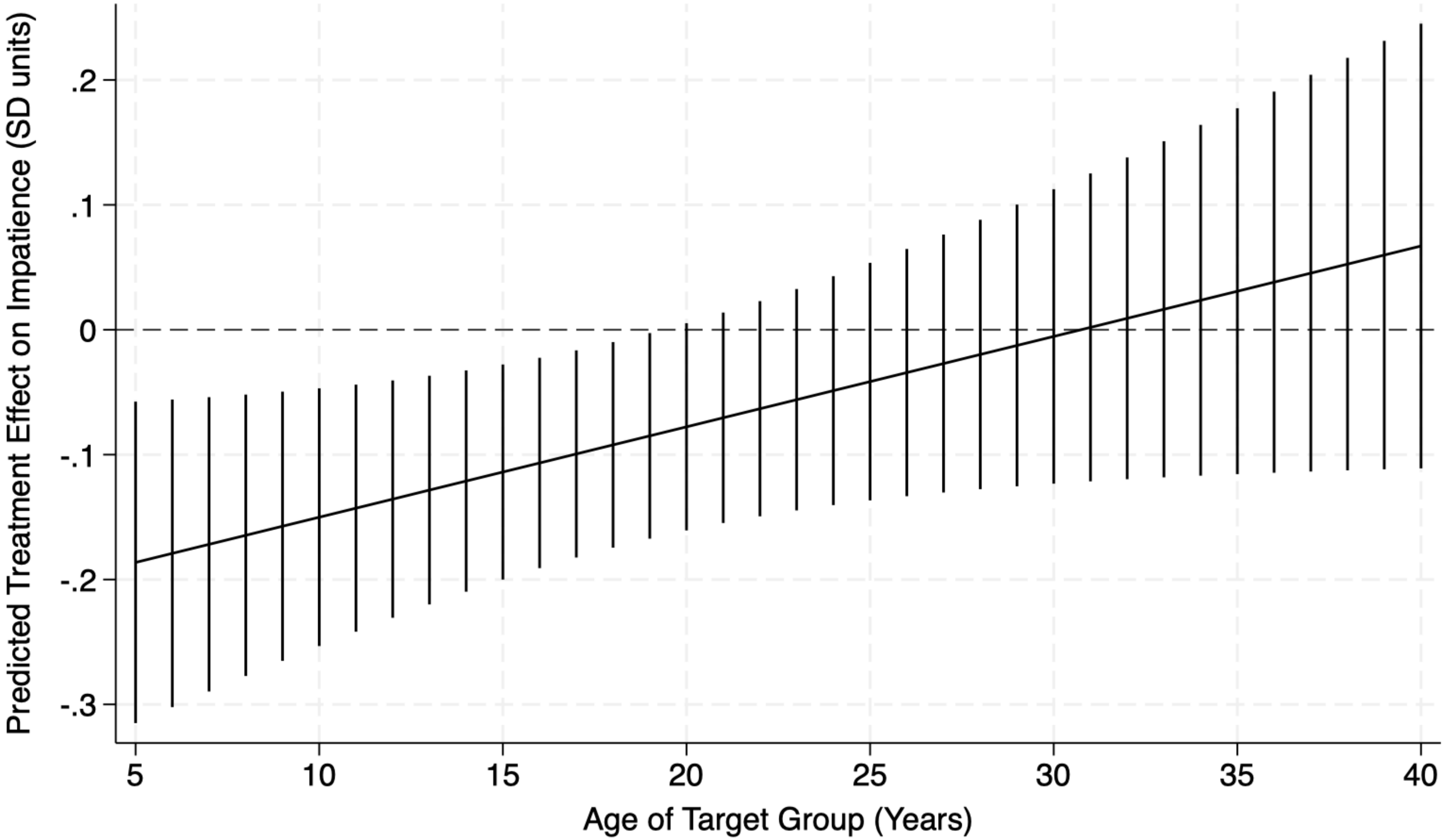
Test of $\theta_i = \theta_j$: $Q(10) = 16.18$, $p = 0.09$

Test of $\theta = 0$: $z = -1.32$, $p = 0.19$

Meta-Regression Analysis

	(1)	(2)	(3)	(4)	(5)	(6)
Age		0.007*	0.007*	0.008*	0.008**	0.019***
		(0.004)	(0.004)	(0.004)	(0.004)	(0.006)
Delay			0.002*	0.002**	0.003**	0.003**
			(0.001)	(0.001)	(0.001)	(0.001)
Intensity				-0.0002	0.001	-0.011
				(0.0025)	(0.003)	(0.024)
Convex Time Budget = 1					0.009	-0.173
					(0.068)	(0.024)
No. of choices					0.010	-0.021
					(0.007)	(0.024)
Developing country = 1						0.000
						(0.254)
Patience (GPS)						0.754
						(0.557)
Meta estimate ($\hat{\theta}$)	-0.056*	-0.190**	-0.204**	-0.218**	-0.259***	-0.364*
	(0.036)	(0.080)	(0.083)	(0.089)	(0.120)	(0.212)
$\hat{\tau}^2$	0.010	0.012	0.017	0.017	0.015	0.048
$\tau^2 = 0$ (p-value)	<0.001	<0.001	<0.001	<0.001	<0.001	0.183
I^2	65.77%	59.39%	60.33%	61.24%	39.91%	22.64%
n (Studies)	11	11	11	11	11	10
N (Treatment effects)	45	45	45	45	45	44

Predicted treatment effects by age from the meta-analysis model



Conclusion

- (Financial-) education interventions appear to be generally successful in fostering non-cognitive outcomes (i.e., time-preferences of children, youth and young adults)
  - We find causal effects on measures of impatience and estimated discount factors for youth
  - In contrast to Lührmann et al. 2018:
    - No effect on time-inconsistency (i.e., present bias)
    - No effect on choice consistency
    - No decrease in narrow bracketing
- This could be an important mechanism explaining part of the treatment effects of financial education on saving behavior documented in previous literature (Lusardi and Mitchell 2014, JEL; Kaiser et al. 2022, JFE)

Thank you!

Prof. Dr. Tim Kaiser
RPTU, IfW Kiel, CESifo & IZA
tim.kaiser@rptu.de

RPTU