Group versus Individual Coaching for Rural Social Protection Programs: Evidence from Uganda, Philippines, and Bangladesh *

Emily A. Beam, Lasse Brune, Narayan Das, Stefan Dercon, Nathanael Goldberg, Rozina Haque, Dean Karlan, Maliha Noshin Khan, Doug Parkerson, Ashley Pople, Yasuyuki Sawada, Christopher Udry, and Rocco Zizzamia

January 12, 2026

Abstract

Multifaceted social protection programs in low-income countries often include both capital grants and informational and behavioral support on the premise that households face simultaneous and multiple frictions. To tackle informational and behavioral constraints, programs typically deploy either individual or group coaching from field agents. The relative efficacy of individual versus group coaching could provide insights into the underlying mechanisms, such as strengthened peer connections, through which coaching changes household decisions. In three similar randomized evaluations in Uganda, the Philippines, and Bangladesh, we find no differences in efficacy across coaching modalities. Given its 15–20% lower costs, group coaching is more cost-effective.

JEL: I38, O12, O15
Keywords: Social protection, coaching, ultra-poor graduation, cost-effectiveness

*The Human Subjects Committee for Innovations for Poverty Action provided oversight for the Uganda and the Philippines sites, "ADB Graduation of the Ultra Poor Pilot," protocol #14304 and "Uganda Graduation Pilot" #14649. The Uganda National Council for Science and Technology also provided oversight for the Uganda site, "Evaluation of the Graduation to Resilience Program in Ramwanja Refugee Settlement and Kamwenge District" #SS 4684. The Oxford Economic Department's Research Ethics Committee (DREC) provided oversight for the Bangladesh site, "Experimental assessment of varying coaching intensity in the "ultra-poor graduation" model in Bangladesh", ECONCIA21-22-12. The Philippines site was pre-registered at the AEA RCT Registry as "Ultra Poor Graduation Pilot in Philippines," #AEARCTR-0004658 (Beam et al., 2020), the Uganda site as "Randomized Impact Evaluation of a Graduation Program on Livelihoods in Refugee and Host Communities in Uganda," #AEARCTR-0004080 (Brune et al., 2024), and the Bangladesh site as "The role of coaching within bundled ultra-poor graduation programs: Experimental evidence from Bangladesh" #AEARCTR-0009618 (Das et al., 2022). Funding provided by the Asian Development Bank, the USAID Bureau of Humanitarian Assistance Technical and Operational Performance Support (TOPS) Uganda Graduation Randomized Control Trial Associate Award, UK International Development from the UK government awarded through the J-PAL Crime and Violence Initiative, and the ERINN Program under the Wellspring Philanthropic Fund. We are thankful for excellent research assistance and support from Nathalie Basto Aguirre, Marie Therese David, Maria Isabel Fernandez, Jorge Franco, Nathan Haddon, Manushi Kumar, Allan Lalisan, and Monserrat Lara; for the support of DOLE leadership and staff, including Ahmma Charisma Lobrin-Satumba and Assistant Secretary Alex Avila; for the support of BRAC USA, specifically Marlowe Popes, Dristy Shrestha, and Lauren Whitehead; for the support of AVSI, specifically Jackie Aldrette, Rita Larok, Massimo Zucca. We are very grateful to Christopher Adam, Raphael Bradenbrink, Dennis Egger, David Evans, Douglas Gollin, Vatsal Khandelwal, Pramila Krishnan, Muhammad Meki, Sandra Sequeira, and Verena Wiedemann, as well as participants at the IPA/GPRL Annual Researcher Gathering, the Oxford CSAE Conference, the CREB ADE Conference, the LUMS Pathways to Development Conference, the NOVAfrica Conference on Economic Development, and The Ohio State University, for helpful comments and discussions. Any errors are the authors'.

1 Introduction

Graduation-style social protection programs aim to lift people out of extreme poverty under the premise that multiple interventions are needed to address the multiple constraints households face. These programs demonstrate marked improvements in household well-being in both the short and long run (Banerjee et al., 2015, 2021a; Barker et al., 2024). Graduation programs typically provide a comprehensive package of assistance, including regular consumption support, coaching and mentoring, a one-time asset transfer to support livelihood development, and improved access to savings. The combination of supports—especially regular coaching and mentoring—is argued to shift households into higher-return economic activities (Blattman et al., 2016; Balboni et al., 2022; Sedlmayr et al., 2020; Banerjee et al., 2022; Roelen and Devereux, 2019).

We evaluate the delivery modality of coaching—individual versus group based—through randomized controlled trials in three countries: Uganda, the Philippines, and Bangladesh. Households are assigned to an individual coach who meets with them regularly to provide training, mentoring, and personalized guidance on livelihood management. Across our three sites, coaches followed a structured curriculum covering business development, agronomic or enterprise-specific skills, financial literacy, and broader life-skills topics such as sanitation, nutrition, and goal-setting. The delivery modality may impact program effectiveness through several mechanisms. Individual coaching may allow messages to be tailored to a household's specific productive asset choice, constraints, and living conditions. Coaches can observe challenges directly within the home and provide advice in a more easily digestible way. Group coaching, in contrast, may facilitate peer learning, help shift norms through discussion with others facing similar constraints, and strengthen social ties or improve access to informal support networks. The relative efficacy of each modality may shed light on the underlying frictions: if group-based coaching generates larger treatment effects than individual-based coaching, this suggests that social norms and complementarities across households can be leveraged to improve outcomes. If individual coaching dominates, this indicates that highly heterogeneous information challenges across households may be more important to address (or more addressable) than information challenges common to many in the community.

As with any program, quality of implementation—not merely program design—may drive impact. The promising results of graduation programs have motivated rapid expansion by non-governmental organizations and, increasingly, governments, with versions implemented in at least 75 countries reaching more than 92 million people (Andrews et al., 2021; Banerjee et al., 2024). This push to scale requires an understanding of which program components are essential and which can be adapted to improve effectiveness or reduce cost. Coaching, a stan-

dard feature of graduation programs, relies on a well-trained labor force and organizational capacity to deliver complex, tailored support. This is costly, and finding skilled coaches grows harder with scale, unlike other program components such as cash transfers. One way to reduce program costs and complexity is to shift away from individual-based coaching delivered through home visits to group-based coaching, in which a single coach works simultaneously with a group of participants. The considerably lower cost of group coaching means that if both generate similar effects, group coaching would be most cost-effective. However, in practice, as in education and active labor market programs more generally, there is no clear consensus on whether individual- or group-based instruction modalities are more effective (McKenzie et al., 2023; Brooks et al., 2018; Ross and Begeny, 2011; Mühlberger and Traut-Mattausch, 2015; Miles et al., 2022; Roelen and Devereux, 2019).

We experimentally measure the differential impact of group versus individual coaching within graduation programs in three sites. We also test the impact of a lower intensity individual coaching arm in Bangladesh. Each site uses a cluster-randomized design, assigning communities to a control group that receives no program, to the graduation program with individual coaching, or to the program with group coaching. Coaching sessions—whether individual or group—are typically fortnightly. We vary the delivery modality but keep the syllabus and discussion topics identical. We measure short-run effects approximately one year after program conclusion in Uganda and the Philippines and longer-term four-year effects in Bangladesh.

Overall treatment effects are strong and consistent with prior research (Banerjee et al., 2015; Bandiera et al., 2017; Banerjee et al., 2021b; Barker et al., 2024). Relative to the control group, graduation increased household consumption (0.17–0.36 standard deviations (s.d.)); reduced food insecurity (0.31–0.80 s.d. in Uganda and the Philippines, not statistically significant in Bangladesh), increased value of productive assets (0.22–0.50 s.d.), increased monthly household income (0.22–0.44 s.d. in Uganda and Bangladesh; not significant statistically in the Philippines) and improved subjective well-being (0.06–0.67 s.d.).

In terms of the relative effectiveness of group versus individual coaching, we find no evidence of differences in treatment effects between group and individual coaching on a range of economic outcomes when examining sites separately or pooling across all three sites. Additionally, when considering the ratio of impact to cost, we find that group coaching is more cost-effective in raising our aggregate outcomes index and key welfare outcomes, and we reject the null of equality of these group vs. individual ratios at the 5% level.

We also examine impacts on social connections and trust, with the hypothesis that group coaching should affect those more. We find suggestive evidence that group coaching promotes social connection, but impacts are modest. We further test for heterogeneity, noting that

while the two approaches perform similarly on average, this could mask underlying—and offsetting—heterogeneity. Following the generic machine learning approach in Chernozhukov et al. (2025), we find no evidence that the difference in group versus individual coaching impacts vary across a broad set of baseline measures.

While common experimental features across these three sites ensure comparability, the diverse study contexts support the external validity of our findings and strengthen the case for scaling group coaching. Specifically, we provide evidence over shorter (Philippines and Uganda) and longer time horizons (Bangladesh); for graduation programs delivered at various scales, from a scoping study in the Philippines to a scaled intervention in Bangladesh; for refugee (Uganda) and non-refugee populations (Uganda, Philippines and Bangladesh); for graduation programs with large lump-sum cash transfers (Uganda) and for those with in-kind productive asset transfers (Philippines and Bangladesh); and across different implementation partners.

A limitation of our study is the absence of a treatment arm in which recipients received the graduation program without any coaching component. Budget constraints and statistical power concerns, and the fact that implementers consider the coaching component essential, precluded inclusion of such a treatment arm. Supporting the implementers' conviction, several experimental "unbundling" studies show that coaching or training-like components are important determinants of the effectiveness of graduation-style programs. In Ghana, when Banerjee et al. (2022) experimentally removed training and coaching elements, the remaining asset-only and savings-only interventions generated weak or non-persistent effects, whereas the full package—with coaching—produced substantial and sustained improvements in consumption, income, and assets. Similarly, in Uganda, Malawi, and Niger, simplified variants of multifaceted programs that excluded training (Sedlmayr et al., 2020; Burchi and Strupat, 2018), mentorship (Blattman et al., 2016), or psychosocial elements (Bossuroy et al., 2022), respectively, delivered smaller or dissipating impacts compared to their full intervention treatment arms. Furthermore, complementary evidence from our Uganda site suggests that the non-asset transfer components of the program are important: IPA (2022a) reports that a "No Asset" treatment arm, which received all components of the individual coaching graduation package—including cash consumption support and savings group promotion—except for the asset transfers, experienced smaller but still substantial improvements in household well-being. This indicates that the effects are not driven entirely by the one-time asset transfer, and although of course we cannot remove the value of the periodic consumption support and savings group promotion from the treatment effect, we posit that this likely supports the argument that the implementer's coaching is effective (whether done as a group or individually).

We contribute to the literature on improving the effectiveness of training and education programs by comparing the relative costs and benefits of group versus dyadic training modalities. There is some experimental evidence from behavioral sciences on the effects of group versus dyadic coaching modalities on helping subjects reduce procrastination (Mühlberger and Traut-Mattausch, 2015) and their respective pedagogic effectiveness on reading instruction from the education literature (Miles et al., 2022; Ross and Begeny, 2011). However, these studies are limited to high-income economies, have small sample sizes, and yield inconsistent results. In the context of development programs, the literature on firms indicates that peer interactions, fostered through group meetings, may have positive effects on entrepreneurial performance (Cai and Szeidl, 2018). In contrast, an RCT comparing mentors (dyadic) to classroom training (group) by Brooks et al. (2018) shows better outcomes for entrepreneurs who receive one-on-one mentoring, although treatment arms vary both the modality and instructional content.

We also advance our understanding of how best to deliver multifaceted economic inclusion programs to maximize impact and cost-effectiveness—a critical question as governments worldwide look to build and scale similar programs. Coaching, being both expensive and reliant on scarce human resources, represents a major obstacle to scaling (Al-Ubaydli et al., 2017). Motivated by this challenge, a growing body of evidence unpacks the role of particular program elements or modifies components to maximize the potential to scale. One-on-one personalized coaching and mentoring is a cornerstone of the early graduation model, and its value has been demonstrated across a range of contexts. For example, Banerjee et al. (2022) finds that asset provision alone is insufficient to generate the long-lasting impacts realized by pairing assets with additional coaching support and training. Similarly, Leight et al. (2023) finds that a lighter-touch graduation program without coaching in comparison with a control group yields only modest impacts on assets and income, and it has no detectable impacts on consumption or food security. Related work by Sedlmayr et al. (2020) and Burchi and Strupat (2018) arrive at similar conclusions, although none of these studies experimentally tests variations in coaching modalities or intensities.

2 Context and interventions

Ultra-poor graduation programs are multi-component interventions designed to help very poor households build sustainable livelihoods and "graduate" out of extreme poverty. They bundle several components to simultaneously relax multiple constraints. Participants receive a productive asset—such as livestock, business capital or cash in a lump sum—alongside training on managing that asset. In some settings, like the Philippines and Bangladesh in

our study, participants choose a desired livelihood from a menu of choices. For instance, in the Philippines, most participants chose swine fattening, while in Bangladesh almost all participants chose a livestock-rearing package comprising cows or goats. In Uganda, participants received a lump sum cash transfer as their "productive asset" after completing business trainings and submitting a business plan approved by the implementer. Coaches then provide regular follow-up coaching sessions on business practices, household financial management, health and education decisions, and the day-to-day challenges of livelihoods imlementation. The asset is intended to increase the recipient's ability to sustainably generate income, while coaching builds the skills, confidence, and behavioral routines needed to use it effectively. In some settings, participants also receive regular cash transfers as well as improved access to savings (usually through informal savings groups). These components have multiple goals: to facilitate smoother consumption, reduce incentives to sell productive assets, and increase the ability to save for larger investments.

We partnered with organizations experienced in graduation-style programming. In the Philippines and Bangladesh we partner with BRAC, a Bangladesh-based international NGO that pioneered the graduation approach in the early 2000s, implemented the programs in the Philippines and Bangladesh. The AVSI Foundation, an international NGO operating in Uganda since 1984, implemented the Uganda program. Aside from the coaching modality, households in T1–T3 received identical program elements within each study site. Program components are also similar across all three sites, comprising an asset transfer valued at PPP 700–1200, bi-monthly regular coaching sessions, skills training and livelihood support, and savings and financial inclusion activities.[1] Uganda and Bangladesh treated households also received monthly consumption support, while members of both control and treatment groups in the Philippines received monthly cash transfers independent of the tested interventions. Appendix C provides full intervention details, and Appendix Figure A.1 summarizes the key components for each site.

Households in each site were randomly assigned to one of two (or three) treatment arms or control:

T1 Graduation package, group-level coaching

T2 Graduation package, individual-level coaching

T3 Graduation package, light individual-level coaching (*Bangladesh only*)

C Control group with no program

[1]Details in Appendix B. We report all monetary values in 2024 PPP.

Coaching delivery lasted between 18 and 29 months across the three study settings (Appendix Figure A.2) and the coaching curriculum was identical across the individual and group treatment arms. Trained staff provided coaching, and recruitment methods were the same for both coaching modalities. Each implementer had a fixed curriculum that coaches would work through, and the curriculum did not differ between modalities. The content of the curriculum was broadly similar across the three studies and included topics like business development, asset management, financial literacy, and non-business topics such as health, education, and sanitation. However, in group coaching sessions, participants from the same implementation group met in a central location, whereas in individual coaching the coach instead visited participants at their homes.

Coaching session frequency also remained largely consistent between T1 and T2 within each study site. In Uganda, T1 held two-hour weekly group sessions alongside savings group meetings, while T2 began with weekly one-hour individual sessions before transitioning to fortnightly. In the Philippines, both T1 and T2 participants received fortnightly coaching sessions (group and individual, respectively). And in Bangladesh, both T1 and T2 delivered weekly coaching, with T1 alternating between group and individual sessions.

Some program elements, such as initial program sensitization and livelihood trainings, were delivered in groups of 10–25 participants for both coaching modalities. This means that participants receiving individual coaching (T2 and T3) experienced some group-based interactions that could generate peer ties and social learning typically associated with group coaching. The extent of this potential confound varies across our study sites. In Bangladesh and the Philippines, group-based activities outside coaching were limited: only one session on asset management was delivered in a group setting in Bangladesh, and only asset selection, asset training, and delivery were done in groups in the Philippines. These brief interactions are unlikely to generate the sustained peer ties and social learning that regular coaching sessions could provide. In Uganda, group-based activities were more common and included regular savings group meetings along with some livelihood trainings, which may somewhat limit our ability to isolate the effect of coaching modality in this context.

Beyond coaching modality, we note a few key differences in the other intervention components across the three sites: (1) in the Philippines and Bangladesh, the asset transfer was delivered as in-kind support, while Ugandan households received cash; (2) consumption support formed part of the intervention in Uganda and Bangladesh, while in the Philippines all study participants—treatment and control households—received consumption support through a pre-existing government program; (3) the Uganda study organized participating households into informal savings groups called Village Savings and Loan Associations (VSLAs) to promote financial inclusion; (4) in Bangladesh all treatment groups received some

individual coaching—in the group coaching arm, fortnightly individual coaching was delivered alongside fortnightly group-based coaching; and (5) in the Philippines, the choice of livelihoods was not balanced across treatment arms, with T1 members less likely than T2 members to select swine fattening (37% vs. 77%) (details in Appendix C).

3 Experimental Design

3.1 Sample and randomization

All three study sites targeted poor or ultra-poor households in primarily rural areas, using eligibility criteria developed by each implementing partner. In the Ugandan site, the program was implemented in refugee and host communities in Kamwenge District. We collected data from 10,509 recipients across 114 village clusters. The Philippines site includes 2,339 program-eligible households across 29 *barangays* (the smallest local administrative units) in northern Negros Occidental. The Bangladesh sample comprises 8,468 recipient households from 88 branches in 11 districts.

All studies use a cluster-randomized design. Randomization clusters in the Philippines are based on dividing each of 29 barangays into four geographically proximate quadrants, with 116 clusters in total. In Bangladesh, we randomized at the branch level (88 clusters), the smallest unit in BRAC's administrative structure. The Uganda study used a two-level design: first randomizing 114 village clusters to receive any treatment or serve as pure control then randomizing the type of graduation program at the household level within treatment village clusters. We used a fixed re-randomization procedure in Uganda and the Philippines to ensure baseline balance.[2]

In all three sites, additional experimental variations were nested within the main experimental design, and we control for random assignment to these nested arms in our primary specifications.[3]

[2]As Appendix D details, we select the set of village-level treatment assignments that yields the best balance on our target covariates: household size, poverty scores, and female headship in Uganda and household size, number of adult household members, value of durable assets, and value of livestock in the Philippines. In Uganda, household-level randomization in treatment village clusters was conducted publicly and not re-randomized.

[3]In Bangladesh, errors in record-keeping prevent us from confidently identifying treatment status in a community-level mobilization arm nested within the group coaching arm (T1). Appendix Table A.11 reports results that control for this compromised treatment assignment variable.

3.2 Primary outcomes

Our main pre-specified outcomes of interest are consumption, food security, subjective well-being, productive asset values, and household income.[4] We also report results on an index of all outcome variables, which was not pre-specified. We report monthly consumption per adult equivalent using 2024 PPP-adjusted values based on 2020 conversion factors (World Bank, 2022). We use a food security index that averages the Food Consumption Score (FCS) and the Household Food Insecurity Scale (HFIAS); reverse-coded so that higher values indicate greater food security), normalized to the control group in each site. We build a subjective well-being index, normalized to the control group. In Uganda and the Philippines, we normalize the average of a reverse-coded 6-item Kessler score and a Cantril's ladder score from four questions about current and future life satisfaction. In Bangladesh, we employ the same methodology, except that the 10-item Center of Epidemiological Studies Depression Scale is used instead of the Kessler score. Productive asset values and total monthly household income are also reported in 2024 PPP dollars.[5] Appendix Table B.1 describes each primary outcome in detail.

We also estimate impacts on an overall index that aggregates these five primary outcomes based on their standardized inverse-covariance weighted average (Anderson, 2008).

3.3 Data collection

We collected baseline data approximately six months before randomization and program roll-out in all three sites. Asset transfers took place approximately one year after baseline in Uganda and the Philippines, and six months after baseline in Bangladesh. In Uganda and Bangladesh, asset deliveries were completed within two months, while in the Philippines they continued for 12 months after program commencement. In all settings coaching started within six months of baseline.

In Uganda and the Philippines, the endline was conducted roughly two years after the asset transfer.[6] In Bangladesh, the endline was conducted approximately six years after the asset transfer. Appendix Figure A.2 shows the study timeline.

[4]In Bangladesh, only consumption and productive asset values were pre-specified as primary outcomes, while food security, subjective well-being, and household income were specified as secondary. In the Philippines and Uganda, some variables pre-specified as primary outcome variables are not reported among the primary outcomes in this paper for lack of overlap across sites. In Uganda, these are saving and borrowing behaviors, and in the Philippines, these are productive time use, physical health, financial behavior, social capital, and women's empowerment.

[5]Asset and income sample sizes in Uganda are smaller because these modules were administered to a randomly selected subset of respondents

[6]In Uganda, the endline stalled for two months mid-implementation due to COVID restrictions.

3.4 Program take-up and participation

In Uganda, nearly all households assigned to treatment (95%) participated in the program, with balance across treatment arms (see Appendix Figure A.3). Initial participation rates were lower in the Philippines, at 80%, but were balanced across treatment arms.[7] By the end of the program, however, the difference in participation rates between individual and group coaching arms is statistically significant, at 75% for individual arms and 70% for group arms. In Bangladesh, 91% of selected beneficiaries received assistance from BRAC, while 6% of control households also reported receiving an asset from BRAC. Participation rates are balanced across treatment arms.

3.5 Balance and attrition

Appendix Tables A.1, A.2, and A.3 show no systematic imbalance across treatment arms for nearly all covariates. In all cases except for the group treatment arm in the Philippines, we fail to reject the null hypothesis of equality. Our results are robust to controlling for covariates using a post-double-selection lasso procedure (Appendix Table A.5.)

At endline, we surveyed 98% of households in the Philippines, 94% in Uganda, and 87% (after six years) in Bangladesh. Appendix Table A.4 demonstrates no evidence of differential attrition by treatment arm when comparing individual to group coaching arms or when comparing treatment arms to the control group.

3.6 Empirical specification

We estimate intent-to-treat effects using the following estimation equation:

$$y_{ij} = \beta_0 + \beta_1 GroupCoach_{ij} + \beta_2 IndCoach_{ij} + \beta_3 IndLightCoach_{ij} + \rho_j + X'\gamma + T'\phi + \epsilon_{ij} \quad (1)$$

where y_{ij} is outcome y measured for household i in cluster j. Our main pre-specified outcomes are consumption, food security, productive asset values, income, and well-being. The $IndLightCoach$ indicator is only included in the Bangladesh specification. We control for stratification-cell fixed effects, ρ_j. X is a vector of re-randomization covariates used in

[7]This lower rate reflects weaker initial targeting, as 11% of initially screened households were later deemed ineligible by BRAC. Among those who did not participate, roughly half (9% overall) declined the initial invitation, most citing lack of time.

Uganda and the Philippines.[8] T is a vector of assignment to nested treatment arms. [9]

In Uganda, we cluster standard errors at the "village cluster" level, as we randomly selected village clusters to receive any intervention or to be part of a pure control group. When we test for differences in impacts between group and individual coaching, we do not cluster our standard errors, reflecting household randomization to coaching modality *within* treatment village clusters.[10] In the Philippines, we cluster standard errors at the "quadrant," or "sub-barangay" level. In Bangladesh, we cluster standard errors at the branch office level. Our results are robust to multiple hypothesis test corrections — see Appendix Table A.6, where we report Anderson's sharpened q-values in brackets below the estimated standard errors (Anderson, 2008).

While each study was designed as a stand-alone contribution and pre-analysis plans were written independently, the empirical strategy followed in this multi-site paper follows the common strategy laid out in each respective pre-analysis plan (Beam et al., 2020; Das et al., 2022; Brune et al., 2024).[11] Deviations to pre-specified primary outcomes to ensure consistency across the three studies have been noted in footnote 3. We also test for heterogeneous treatment effects across subgroups using the generic machine learning approach of Chernozhukov et al. (2025) , noting that this was not specified in our pre-analysis plans.

4 Results

4.1 Average effects

The graduation programs have positive effects on nearly all primary outcome indicators relative to control. We find no evidence of differential impacts by the type of coaching received in all three sites.

Figure 1 presents the point estimates on each coefficient, with a 95% confidence interval, relative to a standardized control group mean across the key pre-specified outcome variables. Table 1 presents the coefficient estimates for the aggregate index (Column 1) and each subsequent outcome, along with the p-value from testing whether the group and individual

[8]Re-randomization covariates are pre-treatment village-cluster characteristics used to select the treatment assignment that best balanced these covariates across repeated random draws. See Appendix D.

[9]As Appendix D describes, this is assignment to the no-asset and the spillover control arms within treatment villages in Uganda, such that we compare treated households to households in pure control villages. In the Philippines, this is assignment to a group livelihood arm.

[10]Appendix Table A.10 demonstrates that our results are unaffected if we instead exclude pure control villages.

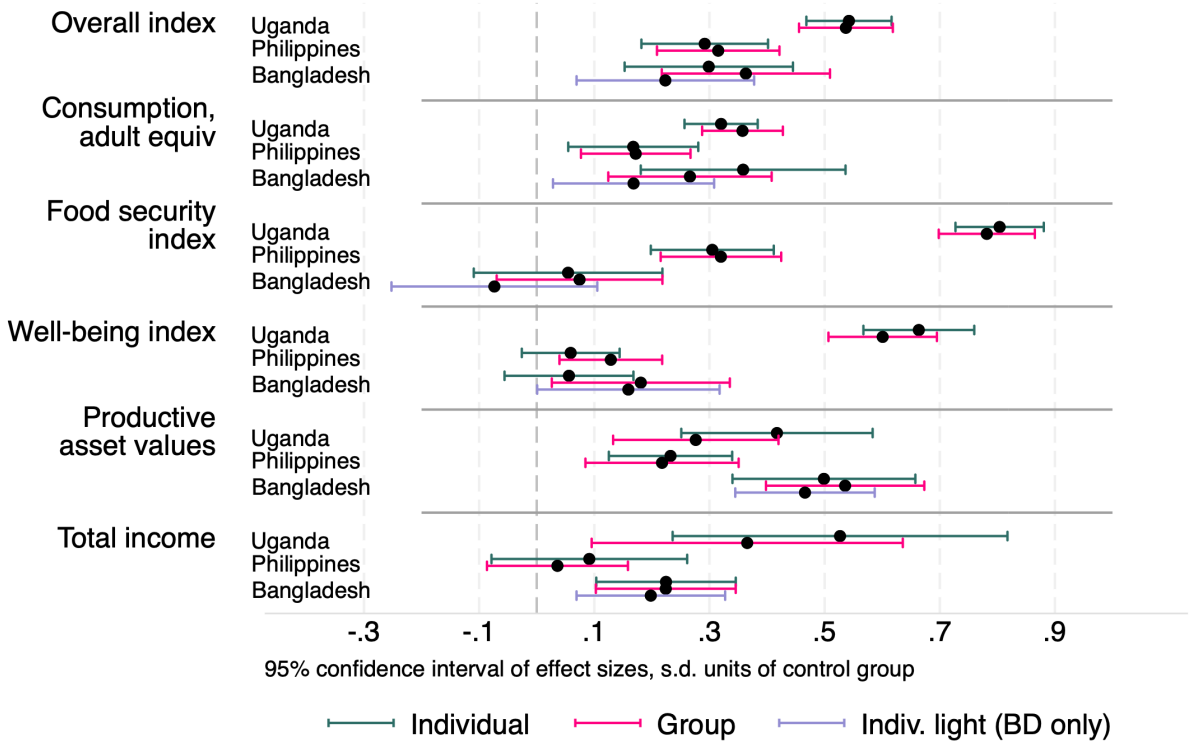
[11]Final program reports for Uganda and the Philippines sites are available in IPA (2022a) and IPA (2022b). Zizzamia (2023) reports the full set of Bangladesh results.

coaching estimates are statistically different.[12] The first three panels of Table 1 do this separately for the three sites. Panel D presents results from a fixed-effects meta analysis, pooling results across the three sites.[13] We weight the estimate from each site by the inverse of the sampling variance of its estimated treatment effect.

[12]Results are robust to estimation controlling for baseline covariates selected using double lasso regression (Urminsky et al., 2016) (Appendix Table A.5).

[13]The light-touch individual treatment arm is included separately but not reported.

Figure 1: Standardized program impacts on primary household outcomes



Notes: Points indicate treatment effects estimated using Equation (1) relative to the endline control group mean, with 95% confidence intervals included. Reported coefficients standardized to the control-group mean of each site. Aggregate index based on inverse-covariance weighted average of all five normalized primary outcomes. Monthly consumption includes food, non-durable and durable goods, and 10% of value of household durable assets, reported per adult equivalent. Food security index is normalized average of FCS and negative HFIAS. Subjective well-being index is normalized average of negative Kessler score and a Cantril's ladder score based on averaging four questions about current and future life satisfaction. 95% confidence intervals based on standard errors clustered at the unit of randomization. UG: village-level; PH: quadrant (sub-barangay)-level; BD: branch-level. All estimates include stratification cell fixed effects and, in UG/PH, baseline covariates targeted during re-randomization.

All three programs yield substantial improvements in household economic well-being, particularly when considering our aggregate outcome index. These results are significant at the 1% level after adjusting for multiple hypothesis testing (See Appendix Table A.6 for results alongside Anderson q-values). Food security improves by 0.79–0.80 s.d. in Uganda and by 0.31–0.32 s.d. in the Philippines, also statistically significant at the MHT-adjusted 1% level. Reflecting the additional cash-transfer component, impacts on consumption and food security are higher in Uganda than in the Philippines, increasing by 28% in Uganda and 9% in the Philippines. In Bangladesh, for which we measure longer-run impacts, consumption rises by 8% (light-intensity individual coaching), 13% (group coaching), and 17% (individual coaching), while there is no detectable impact on food security. These more muted effects are possibly due to general improvements in consumption between baseline and endline, including for the control group, which experienced an inflation-adjusted 25% increase in the value of household consumption.

Impacts on asset holdings are much greater in Bangladesh, at around 0.5 standard deviations relative to control. The absolute value of productive assets increases by PPP 543–675 in Uganda, PPP 228–243 in the Philippines, and PPP 425–489 in Bangladesh. Monthly household income increases for program participants in all sites, but while we estimate statistically significant average impacts of PPP 66–68 per month in Uganda and PPP 54–61 in Bangladesh, the estimated effect in the Philippines of PPP 62–157 per month is imprecisely estimated and not statistically significant. Additionally, subjective well-being increases in all sites, with the largest effects in Uganda (0.61–0.67 s.d.) versus more modest increases in the Philippines (0.06–0.13 s.d.) and Bangladesh (0.06–0.18 s.d.).

Figure 1 shows that the impacts from individual and group coaching are close in magnitude with overlapping confidence intervals. The bottom rows of each panel in Table 1 show that no tests for equality between group and individual arms are statistically significant at the 5% level. Impacts on consumption are 12% and 3% higher among group coaching recipients in Uganda and the Philippines, respectively, and 26% lower in Bangladesh. However, these differences are not statistically significant ($p = 0.29$ in Uganda, $p = 0.94$ in the Philippines, $p = 0.34$ in Bangladesh). The results suggest that the well-being index differs in all contexts, although not in a consistent direction: group coaching yields larger improvements in the Philippines (0.13 s.d. vs 0.06 s.d., p = 0.137) and Bangladesh (0.18 s.d. vs 0.06 s.d., p = 0.10), while individual coaching generates greater well-being improvements in Uganda (0.67 vs. 0.61 s.d., p = 0.12).

At the bottom of each panel of Table 1, we report the group impact per PPP 1,000 spent and the difference between group impact per PPP 1,000 and individual impact per PPP 1,000. Group coaching is more cost effective across most outcomes and settings—reflecting

that impacts are approximately equivalent while costs are lower. Because each treatment effect is estimated with noise, we also report the 95% confidence interval of this difference in the ratios of each treatment effect to its cost and report this in the bottom row. The lower bound thus reports the threshold below which we can rule out (at the 5% level) the individual implementation model leading to an improvement relative to the group approach. In Uganda and the Philippines, we find that group coaching is more cost-effective in our aggregate index, while in Bangladesh, it is marginally insignificant at the 5% level [–0.01, 0.09] but is significant at the 10% level.

We also test whether halving individual coaching intensity affects program impacts in Bangladesh. We find no statistically significant differences in effects when comparing weekly individual coaching to fortnightly individual coaching, apart from a difference in consumption (PPP 20 for weekly coaching vs. PPP 10 for fortnightly coaching, $p = 0.05$). In general, we do not find evidence that light individual coaching is more cost-effective, as the difference in impacts per $1,000 spent are negative for consumption and food security but positive for the other outcomes, and we cannot reject a difference of zero for any individual or aggregate measures.

Appendix Table A.8 reports graduation impacts on a broader range of economic outcomes, including land ownership, land use, and household labor supply. We observe increased land use and value, particularly in Uganda and Bangladesh, and increased household hours worked, particularly in livestock. However, we find no evidence that coaching modality affects these outcomes.

The final rows of Table 1 present the estimated impact per $1,000 (in 2024 PPP) spent. Compared to individual coaching, group coaching is more cost-effective, yielding 0.01–0.07 additional standard deviations in our aggregate index per $1,000 across sites, which is statistically significant at the 5% level. Using the pooled sample, we estimate a 95% confidence interval of the group vs. individual difference in cost-effectiveness of [0.005, 0.023]. We also reject a null of equal cost-effectiveness in the pooled analysis (Panel D) for consumption, food security, and well-being, although we cannot always reject at the 5% level when considering outcomes at the site level (Panels A–C).

4.2 Channels of impact

Individual- and group-based coaching may differentially impact program effectiveness through several channels. Individual coaching may increase engagement and accountability through reduced travel frictions, greater personalization, and direct observation of participants' living conditions and the condition of the transferred asset. Conversely, group

Table 1: Impact of graduation programs on primary outcomes

	(1)	(2)	(3)	(4)	(5)	(6)
Panel A. Uganda						
	Index of all primary outcomes	Monthly adult equiv consumption (2024 PPP)	Food security index	Wellbeing index	Productive asset values (2024 PPP)	Monthly HH income (2024 PPP)
	(1)	(2)	(3)	(4)	(5)	(6)
Group coaching	0.693***	30.339***	0.785***	0.611***	543.055***	65.630***
	(0.050)	(3.068)	(0.044)	(0.048)	(80.096)	(10.243)
Individual coaching	0.710***	27.061***	0.804***	0.673***	675.301***	67.645***
	(0.048)	(2.743)	(0.043)	(0.051)	(103.364)	(9.574)
Control mean	-0.00	100.84	-0.00	0.00	513.68	126.08
Control s.d.	1.00	84.76	1.00	1.00	1584.25	154.06
Observations	10263	10509	10509	10453	10430	10337
p-value, group = ind.	0.880	0.287	0.761	0.115	0.324	0.938
Group impact per 1k cost	0.09	3.75	0.10	0.08	67.05	8.10
Group impact per 1k cost - Ind impact per 1k cost	0.01	0.77	0.01	0.00	-7.33	0.65
CI: Group impact per 1k cost - Ind impact per 1k cost	[0.00, 0.02]	[0.18, 1.35]	[0.00, 0.02]	[-0.01, 0.01]	[-29.27, 14.62]	[-1.82, 3.13]
Panel B. Philippines						
Group coaching	0.315***	18.574***	0.320***	0.129***	227.630***	61.780
	(0.054)	(5.188)	(0.053)	(0.045)	(70.263)	(106.056)
Individual coaching	0.292***	18.104***	0.305***	0.059	243.047***	156.678
	(0.056)	(6.158)	(0.054)	(0.043)	(56.563)	(147.208)
Control mean	0.00	216.39	-0.00	0.00	560.88	715.37
Control s.d.	1.00	108.02	1.00	1.00	1046.01	1715.74
Observations	2288	2287	2288	2288	2288	2288
p-value, group = ind.	0.685	0.939	0.774	0.137	0.824	0.478
Group impact per 1k cost	0.23	13.34	0.23	0.09	163.53	44.38
Group impact per 1k cost - Ind impact per 1k cost	0.07	3.82	0.07	0.06	35.67	-38.04
CI: Group impact per 1k cost - Ind impact per 1k cost	[0.00, 0.15]	[-3.58, 11.22]	[0.00, 0.14]	[0.00, 0.12]	[-56.72, 128.07]	[-192.31, 116.24]
Panel C. Bangladesh						
Group coaching	0.363***	15.174***	0.074	0.181**	488.939***	61.221***
	(0.073)	(4.070)	(0.072)	(0.078)	(63.158)	(16.698)
Individual coaching	0.299***	20.437***	0.054	0.056	455.296***	61.342***
	(0.074)	(5.101)	(0.082)	(0.056)	(72.933)	(16.646)
Individual coaching, light	0.223***	9.588**	-0.074	0.159**	425.304***	54.135***
	(0.078)	(4.014)	(0.090)	(0.080)	(55.603)	(17.742)
Control mean	0.00	117.57	0.01	-0.00	458.70	224.74
Control s.d.	1.00	57.02	0.97	0.99	912.84	273.18
Observations	7565	7534	7534	7534	7447	7445
p-value, group = ind.	0.437	0.337	0.797	0.099	0.619	0.994
Group impact per 1k cost	0.12	4.82	0.02	0.06	155.27	19.44
Group impact per 1k cost - Ind impact per 1k cost	0.04	-0.45	0.01	0.04	37.92	3.63
CI: Group impact per 1k cost - Ind impact per 1k cost	[-0.01, 0.09]	[-3.49, 2.60]	[-0.03, 0.05]	[0.00, 0.09]	[-0.05, 75.90]	[-6.46, 13.72]
p-value, light = ind.	0.376	0.046	0.182	0.169	0.629	0.691
Light impact per 1k cost	0.08	3.51	-0.03	0.06	155.73	19.82
Light impact per 1k cost - Ind impact per 1k cost	0.00	-1.76	-0.04	0.04	38.39	4.01
CI: Light impact per 1k cost - Ind impact per 1k cost	[-0.05, 0.06]	[-4.97, 1.46]	[-0.10, 0.02]	[-0.01, 0.10]	[1.96, 74.81]	[-7.86, 15.88]
Panel D. Pooled, meta-analysis						
Group coaching	0.429***	23.706***	0.486***	0.293***	402.762***	64.356***
	(0.030)	(2.208)	(0.031)	(0.030)	(40.094)	(8.697)
Individual coaching	0.427***	23.035***	0.510***	0.243***	382.357***	66.334***
	(0.029)	(2.178)	(0.031)	(0.028)	(40.929)	(8.286)
Observations	3	3	3	3	3	3
p-value, group = ind.	0.935	0.747	0.360	0.086	0.629	0.826
Group impact per 1k cost	0.082	4.382	0.099	0.068	105.120	9.592
Group impact per 1k cost - Ind impact per 1k cost	0.014	1.143	0.013	0.015	0.591	1.052
CI: Group impact per 1k cost - Ind impact per 1k cost	[0.005, 0.023]	[0.457, 1.829]	[0.003, 0.023]	[0.002, 0.028]	[-21.537, 22.720]	[-1.434, 3.539]

Notes: All dollar values and costs reported in 2024 PPP. Aggregate index based on inverse-covariance weighted average of all five normalized primary outcomes. Monthly consumption includes food, non-durable and durable goods, and 10% of value of household durable assets, reported per adult equivalent. Food security index is normalized average of FCS and negative HFIAS. Subjective well-being index is normalized average of negative Kessler score and a Cantril's ladder score based on averaging four questions about current and future life satisfaction. All indices normalized to endline control group mean. UG asset and income based on imputation of sub-components for respondents randomly selected to answer abridged questionnaires (see Online Appendix for more details). Panel D meta-analysis weights sites by their inverse sampling variance. Robust standard errors are shown in parentheses and clustered at the unit of randomization. UG: village-level randomization for comparisons with control group, with individual-level randomization when testing the equality of group vs. individual coaching impacts. PH: quadrant (sub-barangay)-level randomization. BD: branch-level randomization. Group and individuals impact per 1k cost based on dividing the respective treatment coefficient by the site-specific treatment arm cost (in thousands). All estimates include stratification cell fixed effects and in UG/PH, baseline covariates targeted during re-randomization. Confidence interval reflects a 95% confidence interval around the Group impact per 1k cost minus Individual impact per 1k cost shown in the row above. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

coaching may strengthen social networks and facilitate peer learning.

Administrative and self-reported survey data provide little evidence of systematic differences in these mechanisms between coaching modalities, based on a set of proxy measures. Attendance data from Uganda and the Philippines (Appendix Table A.12), reveal no clear pattern between coaching modality and attendance. The small and insignificant differences we observe are insufficient to result in meaningful changes in economic outcomes discussed above.

Regarding personalization of content, there are only modest differences in whether participants asked about or discussed a range of topics in Uganda (Appendix Figure A.4). In Bangladesh, confidence in performing skills taught during coaching did not differ by modality (Appendix Table A.13).

We also examine whether group coaching strengthens social connections, measured by a social capital index capturing households' perceptions about giving and receiving social assistance and trust indices (Appendix Table A.7).[14] We find suggestive evidence that group coaching modestly improves social capital relative to individual coaching, with the difference across coaching modalities only statistically significant in the Philippines (an increase of 0.12 vs. 0.03 standard deviations, $p = 0.03$) and marginally significant in Bangladesh (group coaching reduces social capital by 0.14 s.d. instead of reducing it by 0.27 s.d., $p = 0.09$). Notably, these are the two study sites where individual coaching participants had minimal exposure to other group activities, unlike in Uganda where all participants attended group savings meetings, as discussed in Section 2. Results on trust are modest and statistically insignificant across all study sites.

4.3 Heterogeneous effects

While we find no differential impacts of group versus individual coaching on average, group coaching could have had heterogeneous and opposing effects relative to individual coaching for different groups of participants, effectively canceling each other out. To examine this possibility, we apply Chernozhukov et al. (2025)'s generic machine learning approach to estimate group average treatment effects by terciles for each of our five primary outcomes, separately for each site. We do not find evidence that the difference between group and individual coaching impacts varies across the broad set of covariates we consider. Appendix E contains additional details.

Overall, we fail to reject the null hypothesis of equal treatment effects between the top and

[14]We only measure the trust index in Uganda and Bangladesh. Definitions of each index are provided in Appendix B.2.1.

bottom terciles across outcomes at a conservative 10% significance level. Results are shown in Appendix Figures A.5, A.6, and A.7.[15] Similarly, we see no evidence of heterogeneous impacts of treatment intensity variation in Bangladesh (see Appendix Figure A.8). One exception is that we find evidence of differential impacts of group coaching on household income (Appendix Figure A.7e). Classification analysis indicates that households with larger assets and higher income at baseline see greater improvements in income with group coaching than individual coaching, while individual coaching leads to larger improvements among poorer households.

While this analysis cannot entirely rule out the possibility of heterogeneity along other unobserved characteristics, such heterogeneity would have to be based on characteristics uncorrelated with the broad range of potential predictors we consider. Additionally, it indicates that among subgroups that could be readily identified and targeted by policymakers—such as those based on household demographics, initial household economic activity, or asset holdings—there is no evidence of disproportionate harm or benefit from shifting from individual to group coaching.

4.4 Benefit-cost analysis

Having found that group coaching is generally more cost-effective than individual coaching across most measures, we now estimate the overall benefits relative to total program costs. We use consumption as our well-being proxy, allowing comparability with Banerjee et al. (2015).

We calculate program implementation costs by summing reported expenditures on all treatments and staffing along with administrative costs, described in Appendix F, and estimate the cost per household *offered* the program, reflecting that program impacts are measured as intent-to-treat (ITT) effects.[16]

Overall, group coaching yields per-participant cost savings of 21% (PPP 523) in the Philippines and 11% in Uganda (PPP 980) (Appendix Table F.1). The hybrid group/individual coaching arm in Bangladesh saves 19% (PPP 731) compared to the individual coaching arm, while the lighter intensity individual coaching arm saves 30% (PPP 1149).

Following Banerjee et al. (2015), we calculate program benefits as the total increase in annual household consumption, using the point estimates from Table 1. We assume an annual

[15]Further disaggregation to quartiles or quintiles demonstrates a similar absence of detectable treatment heterogeneity.

[16]We use cost per participant in Uganda because take-up is near universal. Costs are especially high for Uganda due to administrative expenses that comprise 43–47% of overall costs (see Appendix Table F.1, as well as the cost of one-year pre-intervention refinement activities.

discount rate of 5% and report the benefit-cost ratio assuming a 100% persistence rateAs a benchmark, Bandiera et al. (2017) and Banerjee et al. (2021a) find that impacts increased over time, while Barker et al. (2024) find that effects of a graduation program in Ethiopia fade but remain positive by year seven, with effects on consumption and asset holdings roughly 70% and 45% of the year-three impacts, respectively. Our results in Bangladesh show, at a minimum, that effects remain substantial for six years and in line with the shorter-run impacts in Uganda and the Philippines. Because the persistence rate is unknown, we also report the break-even persistence rate, the rate above which benefits exceed costs.

All three programs yield positive benefit-cost ratios under reasonable assumptions about persistence. In Uganda and Bangladesh, the benefit-cost ratio exceeds one for group coaching if consumption impacts decay by no more than 22% per year.[17] In the Philippines, total benefits exceed costs within the time frame of the program itself. Reflecting equivalent consumption impacts but substantial cost savings, group coaching is consistently more cost-effective than individual coaching in Uganda and the Philippines. In Bangladesh, individual coaching is marginally more cost-effective than the hybrid group coaching arm or lighter-intensity individual arms due to larger estimated impacts. However, the estimated consumption benefits for Bangladesh are noisy, and the difference in benefits between treatment arms is not statistically significant.

Table 2: Site-level benefit-cost analysis

	Uganda		Philippines		Bangladesh		
	Individual (1)	Group (2)	Individual (3)	Group (4)	Individual (5)	Group (6)	Ind. Light (7)
Panel A: Estimated Program Costs							
(1) Costs as incurred (Year 1)	\$9,079	\$8,099	\$1,901	\$1,392	\$3,880	\$3,149	\$2,731
(2) Costs inflated to endline year with annual discount rate	\$10,010	\$8,929	\$2,096	\$1,535	\$5,014	\$4,070	\$3,530
Panel B: Estimated Program Benefits							
(3) Monthly consumption per adult equiv	\$27	\$30	\$18	\$19	\$20	\$15	\$10
	[\$22, \$32]	[\$24, \$36]	[\$6, \$30]	[\$8, \$29]	[\$10, \$30]	[\$7, \$23]	[\$2, \$17]
(4) Annual total household consumption (endline)	\$1,348	\$1,512	\$861	\$884	\$620	\$461	\$291
	[\$1080, \$1616]	[\$1212, \$1811]	[\$287, \$1435]	[\$400, \$1367]	[\$317, \$924]	[\$218, \$703]	[\$52, \$530]
(5) Total benefits though endline	\$4,045	\$4,535	\$1,722	\$1,767	\$3,723	\$2,764	\$1,747
(6) NPV post-endline benefits (100% persistence)	\$26,966	\$30,232	\$17,223	\$17,671	\$12,409	\$9,214	\$5,822
Panel C: Cost Benefit Ratios							
(7) Benefit-cost ratio (at 100% persistence; 5% discounting)	3.1	3.9	9.0	12.7	3.2	2.9	2.1
	[2.5, 3.7]	[3.1, 4.7]	[3.0, 15.1]	[5.7, 19.6]	[1.6, 4.8]	[1.4, 4.5]	[0.4, 3.9]
(8) Internal rate of return (at 100% persistence)	23%	34%	230%	n/a	48%	35%	16%
(9) Break-even persistence rate (at 5% discounting)	86%	78%	32%	$\leqslant 0\%$	71%	78%	90%

Notes: All values converted to 2024 PPP. Calculations based on a social discount rate of 5%. Year 1 and 2 (and in Bangladesh's case also 3, 4, and 5) benefits are extrapolated based on Year 3 (in Bangladesh's case, Year 6) estimates, with no Year 1 benefit assumed in the Philippines because participants received the asset transfer and primary program supports in late Year 1 or early Year 2. Costs are inflated to year 3 in Uganda and the Philippines, and to year 6 in Bangladesh using the social discount rate. 95% confidence intervals for program impacts and benefit/cost ratios in square brackets. Break-even persistence rate reflects the annual consumption benefit persistnece rate above which benefits exceed costs. When accumulated benefits at endline exceed total costs, the break-even persistence rate is less than 0% and the internal rate of return is undefined (n/a).

[17]We note that this is particularly conservative for Bangladesh, as dissipation between implementation and the Year 6 endline would also lead us to understate accrued benefits.

5 Conclusion

We conduct randomized controlled trials in Uganda, the Philippines, and Bangladesh to measure the impact of providing group versus individual coaching within multifaceted social protection programs. All three programs show large improvements in key economic outcomes: consumption, food security, subjective well-being, and asset holdings. We find that benefits exceed program costs under reasonable assumptions about persistence. In the Philippines, where costs are the lowest of the three sites, the benefits accrued in years 2 and 3 alone already exceeded implementation costs.

Additionally, we find that delivering coaching in groups yields substantial cost reductions (11–21%) without negative impacts on program effectiveness on average. Given the reduced costs of implementing group coaching, we find that the benefit per dollar spent is greater for group than individual coaching in aggregate and specifically for increasing consumption, food security, and subjective well-being. Reducing coaching intensity, which we test in Bangladesh, reduces costs by 30% but shows mixed effects on cost-effectiveness: confidence intervals for the difference in benefit per $1,000 spent are wide and generally include zero, although point estimates tend to favor regular intensity coaching. Furthermore, we do not find any evidence of differences in impacts among most- or least-benefited groups across any primary outcome, indicating that shifting from individual to group coaching is unlikely to harm, for example, the poorest or least socially connected households.

As governments look to existing evidence to develop and scale similar programs, our results indicate that programs can improve cost-effectiveness by shifting to group coaching modalities, which in turn may also reduce logistical constraints, such as reducing the total number of coaches that need to be found, trained, and monitored.

References

Ahmad, N., F. Ahmed, D. M. Jolliffe, M. A. Khan, M. A. Mahbub, l. Sharif, V. Swaroop, N. Yoshida, S. Zaidi, and J. Zutt (2010). Poverty maps of Bangladesh. (90487).

Al-Ubaydli, O., J. A. List, and D. L. Suskind (2017). What can we learn from experiments? Understanding the threats to the scalability of experimental results. *American Economic Review 107*(5), 282–286.

Anderson, M. L. (2008). Multiple inference and gender differences in the effects of early intervention: A reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects. *Journal of the American Statistical Association 108*(484), 1481–1495.

Andrews, C., A. de Montesquiou, I. Arevalo Sanchez, P. V. Dutta, B. V. Paul, S. Samaranayake, J. Heisey, T. Clay, and S. Chaudhary (2021, January). *The State of Economic Inclusion Report 2021: The Potential to Scale*. The World Bank.

Athey, S. and S. Wager (2019). Estimating treatment effects with causal forests: An application. *Observational Studies 5*(2), 37–51.

Balboni, C., O. Bandiera, R. Burgess, M. Ghatak, and A. Heil (2022). Why do people stay poor? *The Quarterly Journal of Economics 137*(2), 785–844.

Bandiera, O., R. Burgess, N. Das, S. Gulesci, I. Rasul, and M. Sulaiman (2017, May). Labor markets and poverty in village economies. *The Quarterly Journal of Economics 132*(2), 811–870.

Banerjee, A., E. Duflo, N. Goldberg, D. Karlan, R. Osei, W. Parienté, J. Shapiro, B. Thuysbaert, and C. Udry (2015). A multifaceted program causes lasting progress for the very poor: Evidence from six countries. *Science 348*(6236), 1260799.

Banerjee, A., E. Duflo, and G. Sharma (2021a). Long-term effects of the targeting the ultra poor program. *American Economic Review: Insights 3*(4), 471–86.

Banerjee, A., E. Duflo, and G. Sharma (2021b, December). Long-term effects of the targeting the ultra poor program. *American Economic Review: Insights 3*(4), 471–86.

Banerjee, A., R. Hanna, B. A. Olken, and D. S. Lisker (2024). Social protection in the developing world. *Journal of Economic Literature 62*(4), 1349–1421.

Banerjee, A., D. Karlan, R. Osei, H. Trachtman, and C. Udry (2022). Unpacking a multifaceted program to build sustainable income for the very poor. *Journal of Development Economics 155*, 102781.

Bank, T. W. (2025). PPP conversion factor, private consumption (LCU per international $). https://data.worldbank.org/indicator/PA.NUS.PRVT.PP?locations=UG-PH-BD. Accessed on 12 March, 2025.

Barker, N., D. Karlan, C. Udry, and K. Wright (2024). The fading treatment effects of a mul-

tifaceted asset-transfer program in Ethiopia. *American Economic Review: Insights 6*(2), 277–294.

Beam, E., L. Brune, D. Karlan, Y. Sawada, and C. Udry (2020). Ultra poor graduation pilot in Philippines. AEA RCT Registry.

Blattman, C., E. P. Green, J. Jamison, M. C. Lehmann, and J. Annan (2016, April). The returns to microenterprise support among the ultrapoor: A field experiment in postwar Uganda. *American Economic Journal: Applied Economics 8*(2), 35–64.

Bossuroy, T., M. Goldstein, B. Karimou, D. Karlan, H. Kazianga, W. Parienté, P. Premand, C. C. Thomas, C. Udry, J. Vaillant, et al. (2022). Tackling psychosocial and capital constraints to alleviate poverty. *Nature 605*(7909), 291–297.

Brooks, W., K. Donovan, and T. R. Johnson (2018). Mentors or teachers? Microenterprise training in Kenya. *American Economic Journal: Applied Economics 10*(4), 196–221.

Brune, L., N. Goldberg, D. Karlan, D. Parkerson, and C. Udry (2024). Randomized impact evaluation of a graduation program on livelihoods in refugee and host communities in Uganda. AEA RCT Registry.

Burchi, F. and C. Strupat (2018). Unbundling the impacts of economic empowerment programmes: Evidence from Malawi.

Cai, J. and A. Szeidl (2018). Interfirm relationships and business performance. *The Quarterly Journal of Economics 133*(3), 1229–1282.

Chernozhukov, V., M. Demirer, E. Duflo, and I. Fernández-Val (2025). Fisher–schultz lecture: Generic machine learning inference on heterogeneous treatment effects in randomized experiments, with an application to immunization in india. *Econometrica 93*(4), 1121–1164.

Das, N., S. Dercon, I. Matin, A. Pople, N. Sulaiman, and R. Zizzamia (2022). The role of coaching within bundled ultra-poor graduation programmes: Experimental evidence from Bangladesh. AEA RCT Registry.

Innovations for Poverty Action (2022a). Endline report of the resilience food security activity Graduating to Resilience in Uganda, cohort 1. Technical report, The TOPS Program, Washington, DC. Prepared by Innovations for Poverty Action.

Innovations for Poverty Action (2022b, April). Impact evaluation of the Graduation of the Ultra-Poor Pilot in the Philippines: Report — endline survey. Technical report, Innovations for Poverty Action, Manila, Philippines. Prepared by IPA Philippines.

Leight, J., H. Alderman, D. Gilligan, M. Hidrobo, and M. Mulford (2023). Can a light-touch graduation model enhance livelihood outcomes? Evidence from Ethiopia. Technical report, International Food Policy Research Institute, Washington, DC.

McKenzie, D., C. Woodruff, K. Bjorvatn, M. Bruhn, J. Cai, J. Gonzalez-Uribe, S. Quinn, T. Sonobe, and M. Valdivia (2023, September). Training entrepreneurs. *VoxDevLit 1*(3).

Miles, K. P., K. E. McFadden, D. Colenbrander, and L. C. Ehri (2022). Maximising access to reading intervention: Comparing small group and one-to-one protocols of Reading Rescue. *Journal of Research in Reading 45*(3), 299–323.

Mühlberger, M. D. and E. Traut-Mattausch (2015). Leading to effectiveness: Comparing dyadic coaching and group coaching. *The Journal of Applied Behavioral Science 51*(2), 198–230.

Roelen, K. and S. Devereux (2019). Money and the message: The role of training and coaching in graduation programming. *The Journal of Development Studies 55*(6), 1121–1139.

Ross, S. G. and J. C. Begeny (2011). Improving Latino, English language learners' reading fluency: The effects of small-group and one-on-one intervention. *Psychology in the Schools 48*(6), 604–618.

Sedlmayr, R., A. Shah, and M. Sulaiman (2020). Cash-plus: Poverty impacts of alternative transfer-based approaches. *Journal of Development Economics 144*, 102418.

Urminsky, O., C. Hansen, and V. Chernozhukov (2016). Using double-lasso regression for principled variable selection. *Available at SSRN 2733374*.

U.S. Bureau of Labor Statistics (2025). Consumer price index for all urban consumers (cpi-u). U.S. Bureau of Labor Statistics. Series Id: CUUR0000SA0, Not Seasonally Adjusted, All items in U.S. city average, all urban consumers, Base Period: 1982-84=100.

Welz, M., A. Alfons, M. Demirer, and V. Chernozhukov (2022). *GenericML: Generic Machine Learning Inference*. R package version 0.2.3.

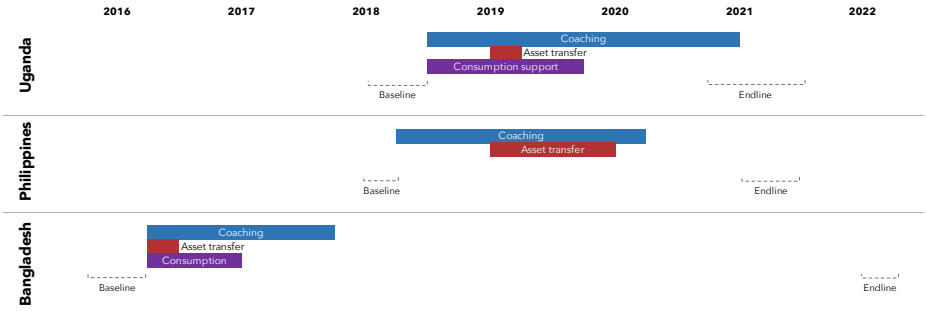
Zizzamia, R. (2023). *Essays on labour markets and social protection in development economics*. PhD thesis, University of Oxford.

A Appendix Figures and Tables

Figure A.1: Intervention components

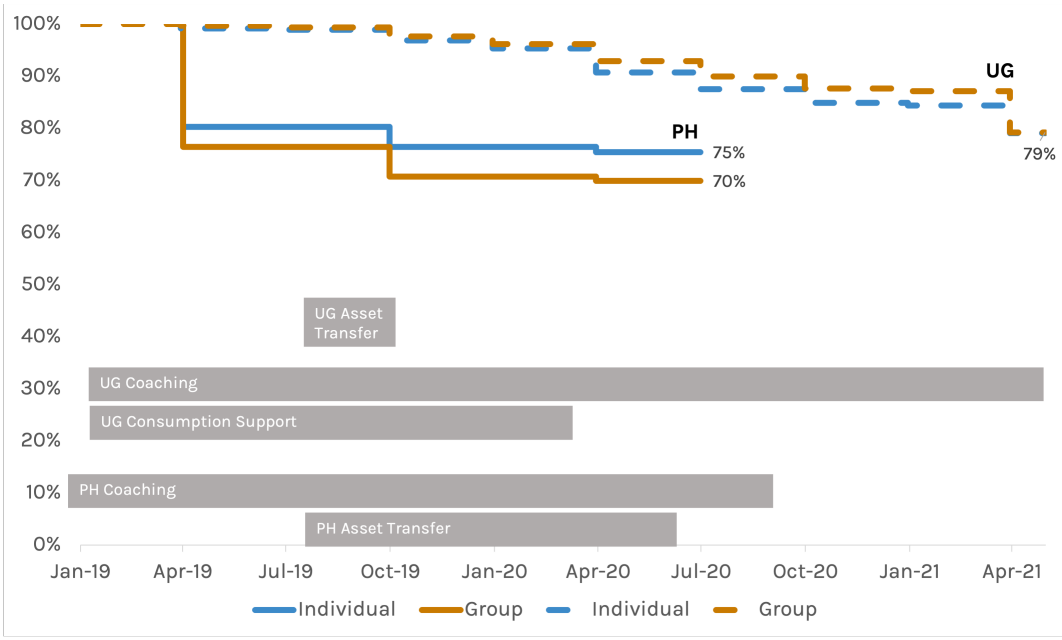
		Consumption support	Asset transfer	Coaching	Coaching frequency
Uganda	Control				
	Individual	X	Cash, PPP 850	Individual	2x/month
	Group	X	Cash, PPP 850	Group	2x/month
Philippines	Control	X			
	Individual	X	In-kind, PPP 720	Individual	2x/month
	Group	X	In-kind, PPP 720	Group	2x/month
Bangladesh	Control				
	Individual	X	In-kind, PPP 1180	Individual	Weekly
	Group	X	In-kind, PPP 1180	Ind. + Group	Weekly
	Individual Light	X	In-kind, PPP 1180	Individual	2x/month

Figure A.2: Project and data collection timeline



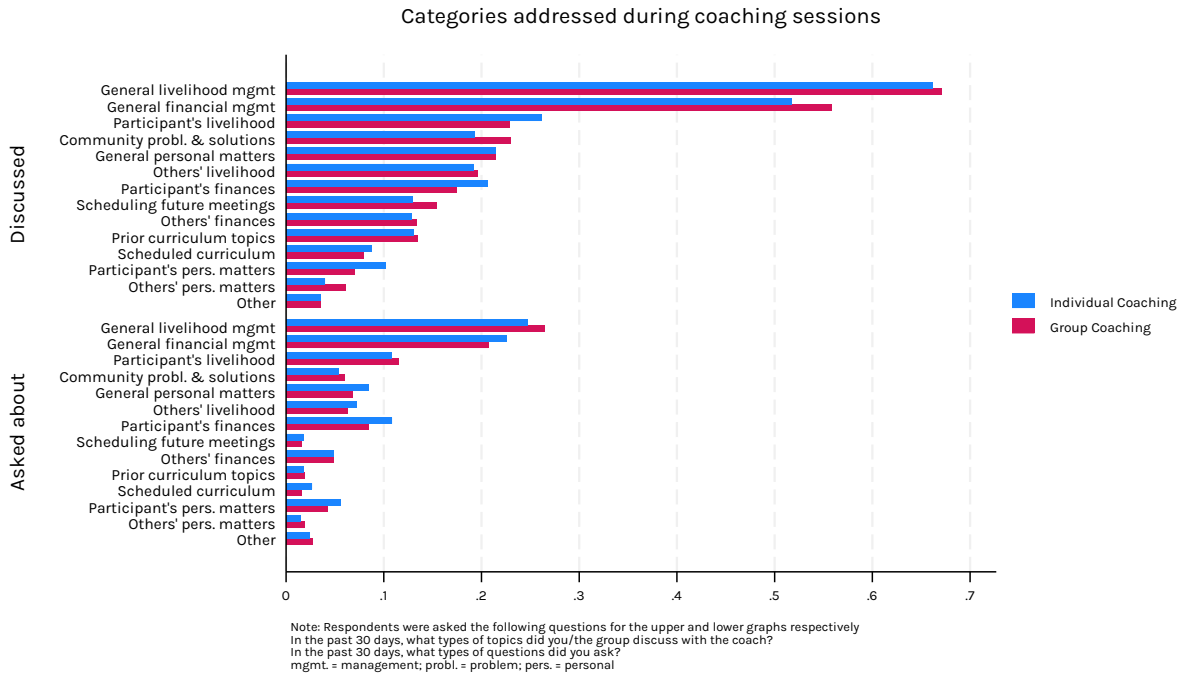
Notes: Because consumption support provided to Philippines participants in both treatment and control communities as part of the ongoing 4Ps program, we do not list it as part of the intervention.

Figure A.3: Program participation over time



Notes: Share of participants actively participating in program, based on partner administrative records.

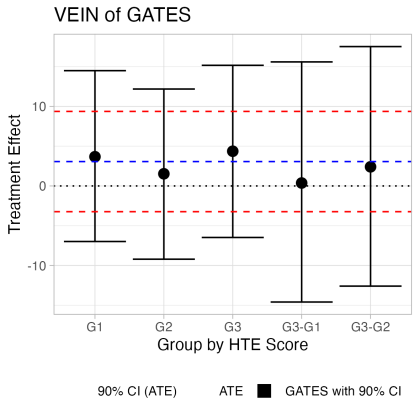
Figure A.4: Topics discussed and asked about, Uganda



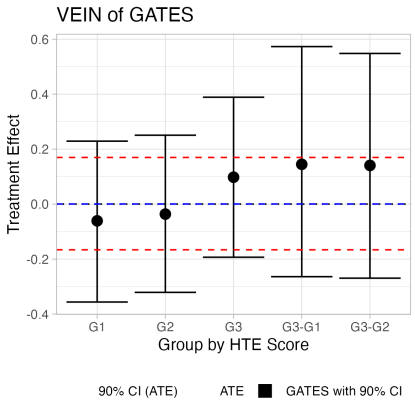
Note: Respondents were asked the following questions for the upper and lower graphs respectively
In the past 30 days, what types of topics did you/the group discuss with the coach?
In the past 30 days, what types of questions did you ask?
mgmt. = management; probl. = problem; pers. = personal

Figure A.5: Heterogeneous treatment effects by tercile, primary outcomes, Uganda

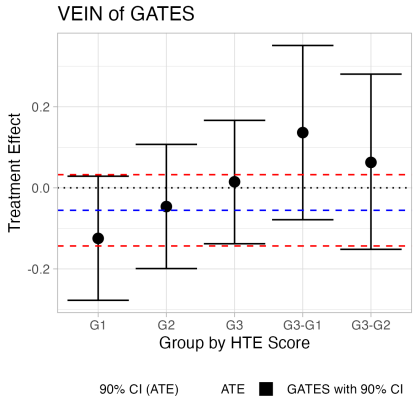
(a) Monthly adult equiv consumption



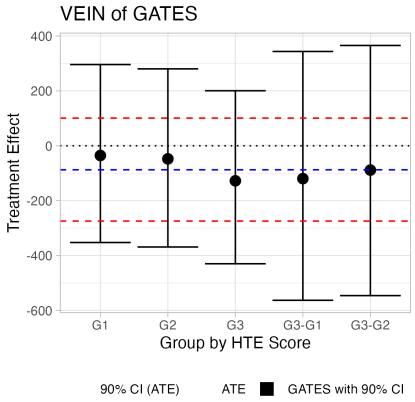
(b) Food security index



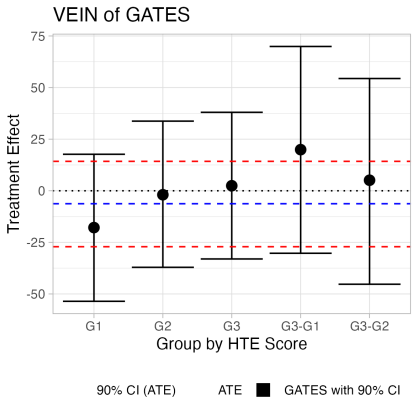
(c) Well-being index



(d) Productive asset values



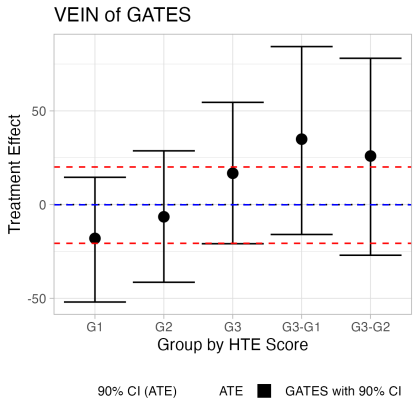
(e) Monthly household income



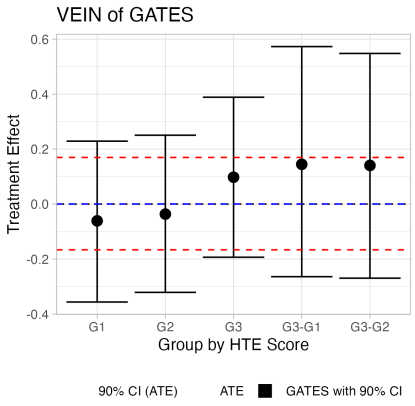
Notes: All values reported in 2024 PPP. Group Average Treatment Effects estimated using the Generic Machine Learning (GenericML) framework, comparing outcomes under the group-based treatment arm versus individual treatment arm. The sample is restricted to households assigned to these two treatment arms. Households are partitioned into terciles groups based on predicted differences in treatment effects between the group and individual arms, generated from a causal forest estimated on baseline covariates. The figure plots estimated average group–individual treatment differences within each group, ordered from least to most affected. GATES are estimated using repeated sample splitting and cross-fitting with 250 splits. A positive value indicates that group coaching yields a larger impact relative to individual coaching. Error bars show 90% confidence intervals based on heteroskedasticity-robust standard errors, as randomization to group vs. individual arms took place at the household level.

Figure A.6: Heterogeneous treatment effects by tercile, primary outcomes, Philippines

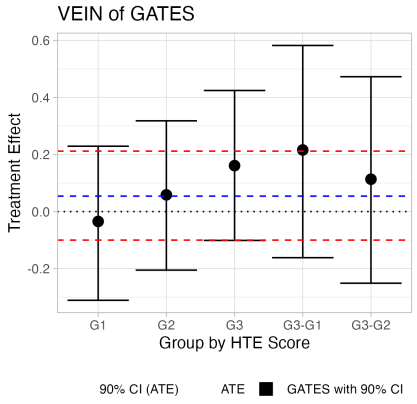
(a) Monthly adult equiv consumption



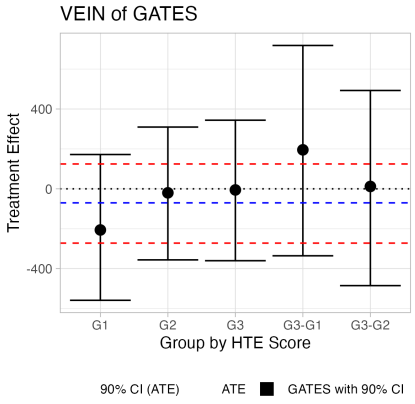
(b) Food security index



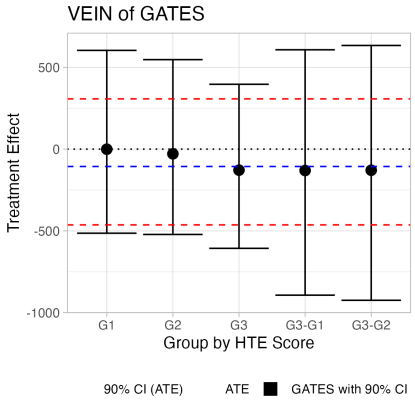
(c) Well-being index



(d) Productive asset values



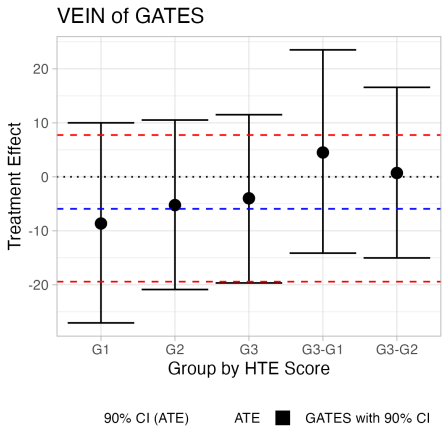
(e) Monthly household income



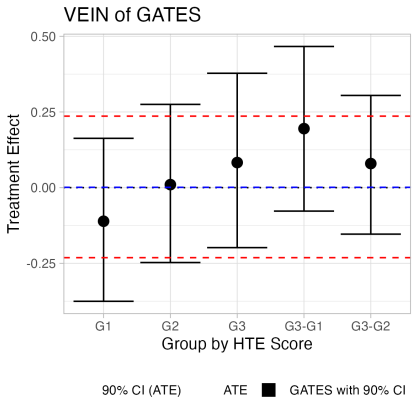
Notes: All monetary values reported in 2024 PPP. Group Average Treatment Effects estimated using the Generic Machine Learning (GenericML) framework, comparing outcomes under the group-based treatment arm versus individual treatment arm. The sample is restricted to households assigned to these two treatment arms. Households are partitioned into terciles groups based on predicted differences in treatment effects between the group and individual arms, generated from a causal forest estimated on baseline covariates. The figure plots estimated average group–individual treatment differences within each group, ordered from least to most affected. GATES are estimated using repeated sample splitting and cross-fitting with 250 splits. A positive value indicates that group coaching yields a larger impact relative to individual coaching. Error bars show 90% confidence intervals, with standard errors clustered at the quadrant (sub-barangay) level.

Figure A.7: Heterogeneous group vs. individual treatment effects by tercile, primary outcomes, Bangladesh,

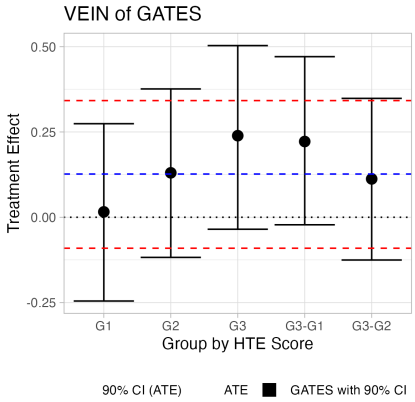
(a) Monthly adult equiv consumption



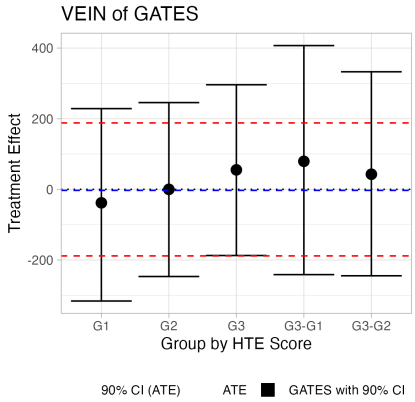
(b) Food security index



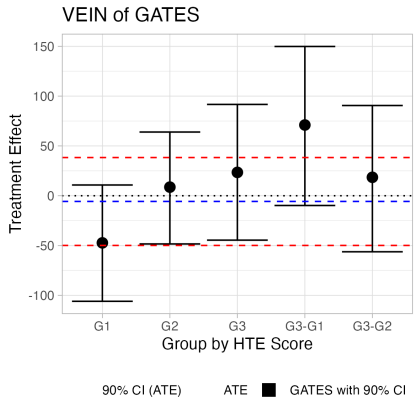
(c) Well-being index



(d) Productive asset values



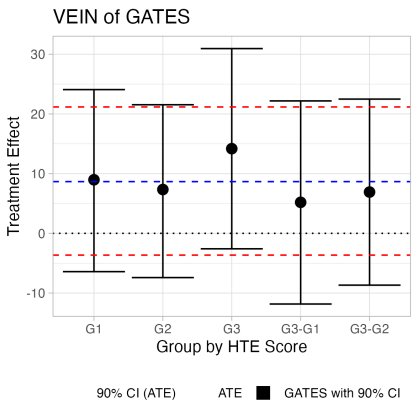
(e) Monthly household income



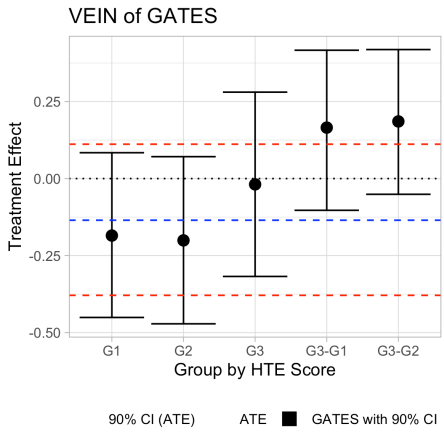
Notes: All monetary values reported in 2024 PPP. Group Average Treatment Effects estimated using the Generic Machine Learning (GenericML) framework, comparing outcomes under the group-based treatment arm versus individual treatment arm. The sample is restricted to households assigned to these two treatment arms. Households are partitioned into terciles groups based on predicted differences in treatment effects between the group and individual arms, generated from a causal forest estimated on baseline covariates. The figure plots estimated average group–individual treatment differences within each group, ordered from least to most affected. GATES are estimated using repeated sample splitting and cross-fitting with 250 splits. A positive value indicates that group coaching yields a larger impact relative to individual coaching. Error bars show 90% confidence intervals, with standard errors clustered at the branch level.

Figure A.8: Heterogeneous individual vs. individual light treatment effects by tercile, primary outcomes, Bangladesh,

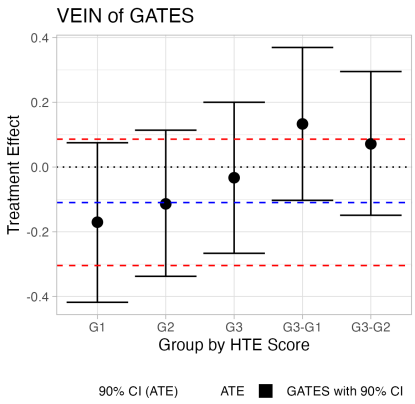
(a) Monthly adult equiv consumption



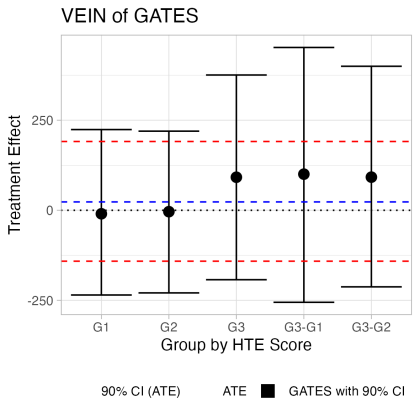
(b) Food security index



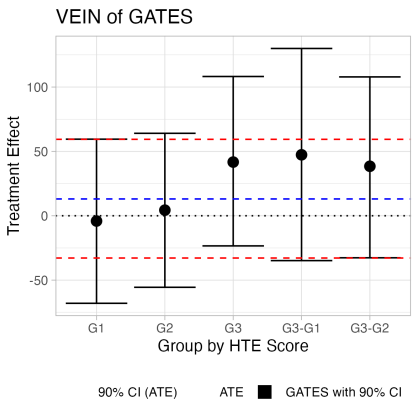
(c) Well-being index



(d) Productive asset values



(e) Monthly household income



Notes: All monetary values reported in 2024 PPP. Group Average Treatment Effects estimated using the Generic Machine Learning (GenericML) framework, comparing outcomes under the individual treatment arm versus the light individual treatment arm. The sample is restricted to households assigned to these two treatment arms. Households are partitioned into terciles groups based on predicted differences in treatment effects between the group and individual arms, generated from a causal forest estimated on baseline covariates. The figure plots estimated average individual–light individual treatment differences within each group, ordered from least to most affected. GATES are estimated using repeated sample splitting and cross-fitting with 250 splits. A positive value indicates that individual coaching yields a larger impact relative to light individual coaching. Error bars show 90% confidence intervals, with standard errors clustered at the branch level.

Table A.1: Balance table – Uganda

Variable	(1) Control Mean/(SE)	(2) Group Mean/(SE)	(3) Individual Mean/(SE)	F-test for balance across all groups F-stat/P-value
HH members	5.756	5.713	5.744	0.085
	(0.069)	(0.059)	(0.064)	0.919
Age HH head	39.489	39.190	39.785	0.795
	(0.658)	(0.552)	(0.611)	0.454
Female head	0.547	0.553	0.544	0.302
	(0.014)	(0.011)	(0.012)	0.740
HH head completed primary	0.175	0.167	0.179	0.609
	(0.022)	(0.015)	(0.010)	0.546
Children under 18 in HH	3.513	3.489	3.504	0.028
	(0.059)	(0.047)	(0.046)	0.973
Any income selling crops	0.867	0.899	0.897	0.523
	(0.040)	(0.016)	(0.012)	0.594
Skipped any meals	0.669	0.651	0.662	0.658
	(0.019)	(0.022)	(0.016)	0.520
F-test of joint significance (F-stat)		0.912	0.984	
Number of observations	2172	2113	2079	6364
Number of clusters	57	57	57	114

Notes: Sample includes all baseline respondents in treatment village clusters assigned to either the control group, group, or individual coaching arms. F-test statistic and respective p-value in Column 4 reflects a joint test of whether both treatment arms jointly predict the respective covariate. F-statistic row in third-to-last row reflects test of whether all covariates jointly predict treatment assignment, restricting to households in that respective arm and the control group. All tests include district (strata) fixed effects and use heteroskedasticity-robust standard errors, reflecting the unit of randomization.

Table A.2: Balance table – Philippines

Variable	(1) Control Mean/(SE)	(2) Group Mean/(SE)	(3) Individual Mean/(SE)	F-test for balance across all groups F-stat/P-value
HH members	5.939	5.842	5.974	1.023
	(0.084)	(0.101)	(0.104)	0.364
Age HH head	45.739	45.404	45.901	0.540
	(0.446)	(0.433)	(0.501)	0.584
Female head	0.217	0.189	0.218	1.154
	(0.019)	(0.018)	(0.022)	0.320
HH head completed primary	0.661	0.674	0.621	2.288
	(0.028)	(0.022)	(0.028)	0.108
Children under 18 in HH	3.019	2.989	3.044	0.286
	(0.059)	(0.073)	(0.067)	0.752
Any income selling crops	0.125	0.142	0.143	0.401
	(0.022)	(0.028)	(0.028)	0.671
Skipped any meals	0.063	0.093	0.093	3.572**
	(0.008)	(0.012)	(0.013)	0.032
F-test of joint significance (F-stat)		4.720***	1.367	
Number of observations	575	571	568	1714
Number of clusters	29	29	29	87

Notes: Sample includes all baseline respondents in quadrants assigned to either the control group, group, or individual coaching arms. F-test statistic and respective p-value in Column 4 reflects a joint test of whether both treatment arms jointly predict the respective covariate. F-statistic row in third-to-last row reflects test of whether all covariates jointly predict treatment assignment, restricting to households in that respective arm and the control group. All tests include barangay (strata) fixed effects and use standard errors clustered at the quadrant (sub-barangay) level.

Table A.3: Balance table – Bangladesh

Variable	(1) Control Mean/(SE)	(2) Group Mean/(SE)	(3) Individual Mean/(SE)	(4) Individual Intense Mean/(SE)	F-test for balance across all groups F-stat/P-value
HH members	3.128	3.297	3.259	3.102	3.844**
	(0.158)	(0.205)	(0.183)	(0.175)	0.012
Age HH head	47.304	46.899	46.507	47.638	1.029
	(1.253)	(1.153)	(0.912)	(0.980)	0.384
Female head	0.548	0.498	0.529	0.519	2.080
	(0.030)	(0.037)	(0.037)	(0.037)	0.109
HH head completed primary	0.044	0.048	0.048	0.060	0.969
	(0.009)	(0.007)	(0.008)	(0.013)	0.411
Children under 18 in HH	0.708	0.801	0.781	0.675	3.851**
	(0.082)	(0.113)	(0.085)	(0.074)	0.012
Skipped any meals	0.632	0.583	0.610	0.578	0.303
	(0.059)	(0.077)	(0.069)	(0.060)	0.823
F-test of joint significance (F-stat)		1.387	1.405	1.913	
Number of observations	1723	2328	2389	2115	8555
Number of clusters	22	22	22	22	88

Notes: Sample includes all baseline respondents. F-test statistic and respective p-value in Column 4 reflects a joint test of whether all three treatment arms jointly predict the respective covariate. F-statistic row in third-to-last row reflects test of whether all covariates jointly predict treatment assignment, restricting to households in that respective arm and the control group. All tests include district (strata) fixed effects and cluster standard errors at the branch level.

Notes: *p < .1; ** p < .05; *** p < .01.

Table A.4: Response rates by treatment assignment

	Uganda		**Philippines**		**Bangladesh**	
	(1)	(2)	(3)	(4)	(5)	(6)
	No covariates	Covariates	No covariates	Covariates	No covariates	Covariates
Group coaching	0.008	0.009	-0.002	-0.003	-0.000	-0.001
	(0.006)	(0.006)	(0.009)	(0.008)	(0.014)	(0.014)
Individual coaching	0.006	0.006	-0.007	-0.009	0.012	0.012
	(0.006)	(0.006)	(0.008)	(0.008)	(0.013)	(0.013)
Individual coaching, light					0.017	0.018
					(0.012)	(0.012)
Observations	11145	11145	2339	2285	8555	8555
Response rate, control						
P-val, group = ind	0.760	0.750	0.449	0.350	0.271	0.260
P-val, group = ind. light					0.096	0.081
P-val, ind. = ind. light					0.566	0.516
P-val, joint signif.	0.628	0.562	0.811	0.699	0.271	0.245

Notes: Sample includes all randomized baseline respondents. Standard errors are clustered at the unit of randomization. Stratification-cell fixed effects are included in all regressions. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.5: Impact of graduation programs on primary outcomes, post double selection lasso

	(1)	(2)	(3)	(4)	(5)	(6)
Panel A. Uganda						
	Index of all primary outcomes	Monthly adult equiv consumption (2024 PPP)	Food security index	Well-being index	Productive asset values (2024 PPP)	Monthly HH income (2024 PPP)
Group coaching	0.681***	29.970***	0.777***	0.600***	551.616***	64.219***
	(0.049)	(3.096)	(0.043)	(0.045)	(71.089)	(10.328)
Individual coaching	0.696***	27.170***	0.789***	0.657***	655.435***	64.967***
	(0.043)	(2.581)	(0.042)	(0.048)	(86.933)	(8.668)
Control mean	-0.00	100.84	-0.00	0.00	513.68	126.08
Control s.d.	1.00	84.76	1.00	1.00	1584.25	154.06
Observations	10263	10509	10509	10453	10430	10337
p-value, group = ind.	0.880	0.287	0.761	0.115	0.324	0.938
Group impact per 1k cost	0.08	3.70	0.10	0.07	68.11	7.93
Group impact per 1k cost - Ind impact per 1k cost	0.01	0.71	0.01	0.00	-4.08	0.77
CI: Group impact per 1k cost - Ind impact per 1k cost	[0.00, 0.02]	[0.11, 1.31]	[0.00, 0.02]	[-0.01, 0.01]	[-24.80, 16.63]	[-1.71, 3.26]
Panel B. Philippines						
Group coaching	0.318***	20.011***	0.319***	0.125***	230.619***	56.803
	(0.058)	(5.372)	(0.053)	(0.046)	(74.455)	(111.690)
Individual coaching	0.305***	17.737***	0.308***	0.059	280.213***	168.208
	(0.061)	(6.610)	(0.054)	(0.043)	(67.934)	(152.487)
Control mean	0.02	216.39	0.00	0.00	569.28	715.66
Control s.d.	1.00	108.02	1.00	1.00	1052.52	1717.22
Observations	2287	2287	2287	2287	2287	2287
p-value, group = ind.	0.841	0.734	0.834	0.170	0.520	0.412
Group impact per 1k cost	0.23	14.38	0.23	0.09	165.67	40.81
Group impact per 1k cost - Ind impact per 1k cost	0.07	5.05	0.07	0.06	18.27	-47.68
CI: Group impact per 1k cost - Ind impact per 1k cost	[-0.01, 0.15]	[-2.99, 13.08]	[0.00, 0.13]	[0.00, 0.12]	[-81.29, 117.83]	[-204.70, 109.34]
Panel C. Bangladesh						
Group coaching	0.362***	15.174***	0.074	0.181**	489.180***	61.221***
	(0.073)	(4.070)	(0.072)	(0.078)	(63.471)	(16.698)
Individual coaching	0.302***	20.437***	0.054	0.056	459.517***	61.342***
	(0.073)	(5.101)	(0.082)	(0.056)	(73.293)	(16.646)
Individual coaching, light	0.224***	9.588**	-0.074	0.159**	426.961***	54.135***
	(0.077)	(4.014)	(0.090)	(0.080)	(55.841)	(17.742)
Control mean	0.00	117.57	0.01	-0.00	460.24	224.74
Control s.d.	1.00	57.02	0.97	0.99	913.99	273.18
Observations	7534	7534	7534	7534	7416	7445
p-value, group = ind.	0.470	0.337	0.797	0.099	0.662	0.994
Group impact per 1k cost	0.12	4.82	0.02	0.06	155.34	19.44
Group impact per 1k cost - Ind impact per 1k cost	0.04	-0.45	0.01	0.04	36.91	3.63
CI: Group impact per 1k cost - Ind impact per 1k cost	[-0.01, 0.08]	[-3.49, 2.60]	[-0.03, 0.05]	[0.00, 0.09]	[-1.19, 75.02]	[-6.46, 13.72]
p-value, light = ind.	0.356	0.046	0.182	0.169	0.601	0.691
Light impact per 1k cost	0.08	3.51	-0.03	0.06	156.34	19.82
Light impact per 1k cost - Ind impact per 1k cost	0.00	-1.76	-0.04	0.04	37.91	4.01
CI: Light impact per 1k cost - Ind impact per 1k cost	[-0.05, 0.06]	[-4.97, 1.46]	[-0.10, 0.02]	[-0.01, 0.10]	[1.41, 74.40]	[-7.86, 15.88]
Panel D. Pooled, meta-analysis						
Group coaching	0.420***	21.882***	0.489***	0.305***	400.099***	55.767***
	(0.029)	(2.123)	(0.030)	(0.029)	(37.202)	(8.370)
Individual coaching	0.423***	22.340***	0.518***	0.254***	405.616***	62.211***
	(0.028)	(2.000)	(0.030)	(0.027)	(38.113)	(7.536)
Observations	3	3	3	3	3	3
p-value, group = ind.	0.896	0.821	0.267	0.067	0.888	0.460
Group impact per 1k cost	0.082	4.192	0.101	0.070	102.238	8.708
Group impact per 1k cost - Ind impact per 1k cost	0.016	0.991	0.014	0.015	0.002	0.742
CI: Group impact per 1k cost - Ind impact per 1k cost	[0.007, 0.025]	[0.293, 1.689]	[0.005, 0.024]	[0.002, 0.027]	[-19.513, 19.518]	[-1.757, 3.240]

Notes: All dollar values and costs reported in 2024 PPP. Aggregate index based on inverse-covariance weighted average of all five normalized primary outcomes. Monthly consumption includes food, non-durable and durable goods, and 10% of value of household durable assets, reported per adult equivalent. Food security index is normalized average of FCS and negative HFIAS. Subjective well-being index is normalized average of negative Kessler score and a Cantril's ladder score based on averaging four questions about current and future life satisfaction. All indices normalized to endline control group mean. UG asset and income based on imputation of sub-components for respondents randomly selected to answer abridged questionnaires (see Online Appendix for more details). Panel D meta-analysis weights sites by their inverse sampling variance. Robust standard errors are shown in parentheses and clustered at the unit of randomization. UG: village-level randomization for comparisons with control group, with individual-level randomization when testing the equality of group vs. individual coaching impacts. PH: quadrant (sub-barangay)-level randomization. BD: branch-level randomization. Group and individuals impact per 1k cost based on dividing the respective treatment coefficient by the site-specific treatment arm cost (in thousands). All estimates include stratification cell fixed effects. Additionalcontrols are selected via post–double-selection lasso with 5-fold cross-validation using standardized baseline covariates, stratification-cell fixed-effect dummies, and all pairwise interactions (with clustering at the randomization unit); the final regression includes the union of covariates selected from treatment- and outcome-lasso models. Confidence interval reflects a 95% confidence interval around the Group impact per 1k cost minus Individual impact per 1k cost shown in the row above. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.6: Impact of graduation programs on primary outcomes, multiple hypotheses adjustments

	(1)	(2)	(3)	(4)	(5)
Panel A. Uganda					
	Monthly adult equiv consumption (2024 PPP)	Food security index	Wellbeing index	Productive asset values (2024 PPP)	Monthly household income (2024 PPP)
Group coaching	30.934***	0.783***	0.605***	513.829***	73.974***
	(2.873)	(0.042)	(0.046)	(99.158)	(15.336)
	[0.001]	[0.001]	[0.001]	[0.001]	[0.001]
Individual coaching	28.091***	0.803***	0.669***	657.474***	67.658***
	(2.519)	(0.039)	(0.048)	(116.403)	(13.820)
	[0.001]	[0.001]	[0.001]	[0.001]	[0.001]
Control mean	100.84	-0.00	0.00	755.50	127.40
Control s.d.	84.76	1.00	1.00	2077.63	165.90
Observations	10509	10509	10453	5259	2217
p-value, joint test	0.00	0.00	0.00	0.00	0.00
p-value, group = ind.	0.29	0.76	0.12	0.39	0.71
95% CI, Group-Ind	[-7.2, 2.1]	[-0.05, 0.07]	[-0.01, 0.11]	[-131.5, 335.3]	[-41.4, 28.4]
Panel B. Philippines					
Group coaching	18.574***	0.320***	0.129***	227.630***	61.780
	(5.188)	(0.053)	(0.045)	(70.263)	(106.056)
	[0.001]	[0.001]	[0.004]	[0.002]	[0.203]
Individual coaching	18.104***	0.305***	0.059	243.047***	156.678
	(6.158)	(0.054)	(0.043)	(56.563)	(147.208)
	[0.004]	[0.001]	[0.069]	[0.001]	[0.107]
Control mean	216.39	-0.00	0.00	560.88	715.37
Control s.d.	108.02	1.00	1.00	1046.01	1715.74
Observations	2287	2288	2288	2288	2288
p-value, joint test	0.00	0.00	0.02	0.00	0.57
p-value, group = ind.	0.94	0.77	0.14	0.82	0.48
95% CI, Group-Ind	[-12.6, 11.6]	[-0.12, 0.09]	[-0.16, 0.02]	[-121.9, 152.8]	[-168.9, 358.7]
Panel C. Bangladesh					
Group coaching	15.721***	0.083	0.184**	528.022***	67.457***
	(4.343)	(0.076)	(0.079)	(72.410)	(17.750)
	[0.001]	[0.101]	[0.016]	[0.001]	[0.001]
Individual coaching	22.021***	0.060	0.058	500.362***	68.463***
	(5.425)	(0.086)	(0.058)	(80.258)	(18.147)
	[0.001]	[0.151]	[0.108]	[0.001]	[0.001]
Individual coaching, light	11.006**	-0.065	0.165**	442.112***	76.689***
	(4.549)	(0.093)	(0.082)	(57.614)	(24.508)
	[0.014]	[0.151]	[0.031]	[0.001]	[0.003]
Control mean	118.04	-0.00	-0.00	461.98	226.61
Control s.d.	61.68	1.00	1.00	936.24	287.07
Observations	7534	7534	7534	7447	7445
p-value, joint test	0.00	0.34	0.07	0.00	0.00
p-value, group = ind.	0.28	0.76	0.11	0.74	0.96
p-value, ind. = ind. light	0.07	0.20	0.17	0.42	0.74

Notes: All dollar values and costs reported in 2024 PPP. Aggregate index based on inverse-covariance weighted average of all five normalized primary outcomes. Monthly consumption includes food, non-durable and durable goods, and 10% of value of household durable assets, reported per adult equivalent. Food security index is normalized average of FCS and negative HFIAS. Subjective well-being index is normalized average of negative Kessler score and a Cantril's ladder score based on averaging four questions about current and future life satisfaction. All indices normalized to endline control group mean. UG asset and income based on imputation of sub-components for respondents randomly selected to answer abridged questionnaires (see Online Appendix for more details). Panel D meta-analysis weights sites by their inverse sampling variance. Robust standard errors are shown in parentheses and clustered at the unit of randomization. UG: village-level randomization for comparisons with control group, with individual-level randomization when testing the equality of group vs. individual coaching impacts. PH: quadrant (sub-barangay)-level randomization. BD: branch-level randomization. We control for the false discovery rate using Anderson sharpened q-values, which are reported in brackets below the standard errors. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.7: Impact of graduation on social capital and trust

	(1)	(2)
Panel A. Uganda		
	Social capital index	Trust index
Group coaching	0.090*	0.101***
	(0.051)	(0.037)
Individual coaching	0.032	0.075*
	(0.054)	(0.039)
Control mean	-0.00	0.00
Control s.d.	1.00	1.00
Observations	5256	10509
p-value, joint test	0.19	0.02
p-value, group = ind.	0.15	0.32
Panel B. Philippines		
	Social capital index	
Group coaching	0.120**	
	(0.051)	
Individual coaching	0.026	
	(0.052)	
Control mean	0.00	
Control s.d.	1.00	
Observations	2288	
p-value, joint test	0.03	
p-value, group = ind.	0.03	
Panel C. Bangladesh		
	Social capital index	Trust index
Group coaching	-0.142*	-0.067
	(0.083)	(0.096)
Individual coaching	-0.266***	-0.165**
	(0.088)	(0.074)
Individual coaching, light	-0.212**	-0.131
	(0.098)	(0.081)
Control mean	0.00	0.00
Control s.d.	1.00	1.00
Observations	7534	7534
p-value, joint test	0.03	0.14
p-value, group = ind.	0.09	0.24
p-value, ind. = ind. light	0.54	0.62

Notes: Social capital index: Site-specific index capturing access to and engagement in informal support and civic/community structures: in Uganda it is the mean of bonding and bridging support networks (who respondents could turn to or would help inside vs. outside the village across kin and non-kin ties), while in the Philippines and Bangladesh it is the sum of binary indicators on recent giving/receiving assistance (PH) and involvement/voice in local dispute-resolution and civic participation (BD). Trust index: Site-specific index of perceived interpersonal trust and mutual support: in Uganda it is the sum of five Likert items on community trust/helpfulness (with jealousy items reverse-coded), and in Bangladesh it is the sum of binary measures of trust in others' fairness/cooperation (including land sharing, general trust, and likelihood of money being returned). All estimates include stratification cell fixed effects and in UG/PH, baseline covariates targeted during re-randomization. Robust standard errors are shown in parentheses and clustered at the unit of randomization. UG: village-level randomization for comparisons with control group, with individual-level randomization when testing the equality of group vs. individual coaching impacts. PH: quadrant (sub-barangay)-level randomization. BD: branch-level randomization * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.8: Impact of graduation on secondary economic outcomes

	(1)	(2)	(3)	(4)	(5)
Panel A. Uganda					
	Own/use land	Plot value	Plot size	Hours HH worked, past 7 days	Hours HH worked in livestock, past 7 days
Group coaching	0.030***	1625.067***	1456.945***	6.014*	5.340***
	(0.010)	(208.478)	(211.565)	(3.353)	(1.030)
Individual coaching	0.028***	1568.185***	1378.453***	6.503**	4.538***
	(0.010)	(229.898)	(219.206)	(3.152)	(1.034)
Control mean	0.90	3455.55	3830.07	41.89	9.36
Control s.d.	0.30	4447.67	5604.58	43.69	20.14
Observations	10095	10509	10505	2046	5252
p-value, joint test	0.01	0.00	0.00	0.09	0.00
p-value, group = ind.	0.67	0.32	0.31	0.82	0.44
Panel B. Philippines					
	Use land		Plot size	Hours HH worked, past 7 days	Hours HH worked in livestock, past 7 days
Group coaching	0.056**		1623.659	1.404	1.913***
	(0.027)		(1098.676)	(2.717)	(0.604)
Individual coaching	0.036		1171.421	3.484	0.778
	(0.029)		(1301.189)	(2.699)	(0.552)
Control mean	0.31		2272.10	62.20	7.75
Control s.d.	0.46		18593.20	45.97	11.12
Observations	2288		2178	2288	2288
p-value, joint test	0.11		0.31	0.43	0.01
p-value, group = ind.	0.50		0.74	0.43	0.06
Panel C. Bangladesh					
	Own/use land	Plot value		Hours HH worked, past 7 days	Hours HH worked in livestock, past 7 days
Group coaching	0.089**	3376.793***		4.703***	3.702***
	(0.037)	(942.162)		(1.552)	(0.582)
Individual coaching	0.120***	2718.066**		3.772**	2.898***
	(0.036)	(1065.083)		(1.657)	(0.473)
Individual coaching, light	0.091***	2024.724**		3.550**	3.445***
	(0.035)	(850.443)		(1.348)	(0.541)
Control mean	0.57	6355.50		33.70	4.87
Control s.d.	0.50	11577.81		19.03	7.07
Observations	7534	7416		7445	7445
p-value, joint test	0.01	0.00		0.01	0.00
p-value, group = ind.	0.33	0.54		0.60	0.18
p-value, ind. = ind. light	0.34	0.49		0.89	0.33

Notes: All dollar values and costs reported in 2024 PPP. All estimates include stratification cell fixed effects and in UG/PH, baseline covariates targeted during re-randomization. Robust standard errors are shown in parentheses and clustered at the unit of randomization. UG: village-level randomization for comparisons with control group, with individual-level randomization when testing the equality of group vs. individual coaching impacts. PH: quadrant (sub-barangay)-level randomization. BD: branch-level randomization * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.9: Impact of graduation on financial inclusion

	(1)	(2)	(3)	(4)	(5)	(6)
Panel A. Uganda						
	Any savings	Savings value (2024 PPP)	Any borrowing	Outstanding loans (2024 PPP)		
Group coaching	0.090***	192.974***	0.024	43497.813***		
	(0.008)	(24.596)	(0.018)	(10422.595)		
Individual coaching	0.095***	216.700***	0.038**	38799.579***		
	(0.008)	(23.379)	(0.017)	(10435.654)		
Control mean	0.89	153.61	0.71	162020.85		
Control s.d.	0.32	370.69	0.45	273837.29		
Observations	10509	10407	10509	10509		
p-value, joint test	0.00	0.00	0.09	0.00		
p-value, group = ind.	0.58	0.42	0.39	0.29		
Panel B. Philippines						
	Any savings	Savings value (2024 PPP)	Any borrowing	Outstanding loans (2024 PPP)	Any loans given	Outstanding loans given (2024 PPP)
Group coaching	0.113***	12.882***	-0.017	10.980	0.036***	5.165
	(0.024)	(4.650)	(0.024)	(25.880)	(0.012)	(5.454)
Individual coaching	0.152***	23.999***	-0.055**	48.901	0.014	20.722*
	(0.027)	(5.472)	(0.023)	(30.969)	(0.014)	(10.731)
Control mean	0.26	23.05	0.61	250.25	0.06	7.06
Control s.d.	0.44	71.21	0.49	515.58	0.25	55.44
Observations	2288	2268	2288	2288	2286	2288
p-value, joint test	0.00	0.00	0.05	0.28	0.01	0.13
p-value, group = ind.	0.10	0.06	0.08	0.23	0.13	0.16
Panel C. Bangladesh						
	Any savings	Savings value (2024 PPP)	Any borrowing	Outstanding loans (2024 PPP)	Any loans given	Outstanding loans given (2024 PPP)
Group coaching	0.146***	67.709***	0.028**	13.710	0.062***	213.031***
	(0.034)	(19.087)	(0.014)	(8.327)	(0.023)	(79.660)
Individual coaching	0.066**	44.763**	0.022	4.032	0.014	150.489**
	(0.032)	(18.401)	(0.014)	(7.204)	(0.023)	(69.567)
Individual coaching, light	0.055*	19.980	0.007	5.396	0.045*	160.049**
	(0.029)	(15.718)	(0.015)	(8.858)	(0.023)	(77.502)
Control mean	0.36	96.84	0.06	24.12	0.57	502.62
Control s.d.	0.48	283.56	0.24	144.73	0.50	1212.06
Observations	7421	8529	7534	8529	7534	8529
p-value, joint test	0.00	0.00	0.18	0.43	0.03	0.05
p-value, group = ind.	0.02	0.26	0.65	0.22	0.03	0.39
p-value, ind. = ind. light	0.72	0.14	0.33	0.87	0.15	0.89

Notes: All dollar values and costs reported in 2024 PPP. PH: Whether given or received loans in the past 12 months, with current outstanding balances reported. UG: Savings based on any savings in past 3 months. BD: Savings and loans based on current balances. All estimates include stratification cell fixed effects and in UG/PH, baseline covariates targeted during re-randomization. Robust standard errors are shown in parentheses and clustered at the unit of randomization. UG: village-level randomization for comparisons with control group, with individual-level randomization when testing the equality of group vs. individual coaching impacts. PH: quadrant (sub-barangay)-level randomization. BD: branch-level randomization * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.10: Impact of group vs. individual coaching in Uganda, full sample and coaching only

	All treatment arms		Group and individual coaching only	
	Group coaching (1)	Obs. (2)	Group coaching (3)	Obs. (4)
Monthly adult equiv consumption (2024 PPP)	3.3	10509	-2.5	8337
	(2.5)		(2.4)	
Food security index	0.01	10509	0.01	8337
	(0.03)		(0.03)	
Wellbeing index	0.05	10453	0.05	8295
	(0.03)		(0.03)	
Productive asset values (2024 PPP)	-132.2	10430	79.8	8270
	(96.9)		(80.9)	
Monthly HH income (2024 PPP)	-2.0	10337	-0.6	8191
	(10.6)		(8.2)	

Notes: Columns 1 and 2 include all households, while Columns 3 and 4 restrict to only households assigned to individual or group coaching. All dollar values and costs reported in 2024 PPP. Aggregate index based on inverse-covariance weighted average of all five normalized primary outcomes. Monthly consumption includes food, non-durable and durable goods, and 10% of value of household durable assets, reported per adult equivalent. Food security index is normalized average of FCS and negative HFIAS. Subjective well-being index is normalized average of negative Kessler score and a Cantril's ladder score based on averaging four questions about current and future life satisfaction. All indices normalized to endline control group mean. UG asset and income based on imputation of sub-components for respondents randomly selected to answer abridged questionnaires (see Online Appendix for more details). Robust standard errors are shown in parentheses and clustered at the unit of randomization. UG: village-level randomization for comparisons with control group, with individual-level randomization when testing the equality of group vs. individual coaching impacts. All estimates include stratification cell fixed effects and in UG/PH, baseline covariates targeted during re-randomization. Confidence interval reflects a 95% confidence interval around the Group impact per 1k cost minus Individual impact per 1k cost shown in the row above. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.11: Impact of graduation programs in Bangladesh on primary and social outcomes, controlling for village assistance committee randomization

	Primary Outcomes					Social Outcomes	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Bangladesh						
	Monthly adult equiv consumption (2024 PPP)	Food security index	Productive asset values (2024 PPP)	Monthly household income (2024 PPP)	Wellbeing index	Social capital index	Trust index
Group coaching	11.866**	-0.007	460.979***	63.248***	0.172**	-0.101	-0.034
	(5.049)	(0.098)	(108.393)	(19.445)	(0.074)	(0.101)	(0.103)
Individual coaching	22.023***	0.060	500.333***	68.465***	0.058	-0.266***	-0.165**
	(5.408)	(0.086)	(79.766)	(18.097)	(0.058)	(0.088)	(0.074)
Individual coaching, light	11.039**	-0.064	442.729***	76.723***	0.165**	-0.212**	-0.132
	(4.531)	(0.093)	(58.049)	(24.505)	(0.082)	(0.098)	(0.081)
Control mean	118.04	-0.00	461.98	226.61	-0.00	0.00	0.00
Control s.d.	61.68	1.00	936.24	287.07	1.00	1.00	1.00
Observations	7534	7534	7447	7445	7534	7534	7534
p-value, joint test	0.001	0.641	0.000	0.000	0.075	0.019	0.124
p-value, group = ind.	0.114	0.500	0.737	0.792	0.103	0.075	0.163
p-value, ind. = ind. light	0.066	0.201	0.424	0.743	0.174	0.542	0.627

Notes: All dollar values and costs reported in 2024 PPP. Aggregate index based on inverse-covariance weighted average of all five normalized primary outcomes. Monthly consumption includes food, non-durable and durable goods, and 10% of value of household durable assets, reported per adult equivalent. Food security index is normalized average of FCS and negative HFIAS. Subjective well-being index is normalized average of negative Kessler score and a Cantril's ladder score based on averaging four questions about current and future life satisfaction. All indices normalized to endline control group mean. UG asset and income based on imputation of sub-components for respondents randomly selected to answer abridged questionnaires (see Online Appendix for more details). Panel D meta-analysis weights sites by their inverse sampling variance. Robust standard errors are shown in parentheses and clustered at the branch level, the unit of randomization. All estimates include stratification cell fixed effects and control for assignment to Village Assistance Committees. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.12: Coaching attendance by modality, Uganda and Philippines

	Uganda			Philippines		
	Group	Individual	Difference	Group	Individual	Difference
	Last three months			*Out of nine sessions*		
Any attendance	0.87	0.83	0.04***	0.73	0.79	-0.05**
Avg. attendance	2.74	2.29	0.45***	3.52	5.13	-1.61***
Observations	2,002	1,959	3,961	583	583	1,166

Notes: Uganda: Attendance based on self-reports over the past 3 months, while implementation was ongoing. Philippines: Attendance based on administrative data collected by graduation coaching facilitators across 9 coaching sessions and merged with survey data.

Table A.13: Confidence in ability to perform various tasks, Bangladesh only

	(1) Collecting debts	(2) Educating children	(3) Growing herd	(4) Accessing medical care	(5) Starting business	(6) Protecting business	(7) Resolving dispute	(8) Running business
Group coaching	-0.027	-0.010	0.039	0.013	-0.055	-0.001	-0.054	-0.077
	(0.075)	(0.074)	(0.054)	(0.079)	(0.059)	(0.072)	(0.078)	(0.063)
Individual coaching	0.010	0.006	0.048	0.041	0.001	0.055	-0.008	-0.016
	(0.079)	(0.075)	(0.061)	(0.084)	(0.064)	(0.074)	(0.088)	(0.066)
Individual coaching, light	0.060	0.092	0.090*	0.095	0.034	0.071	0.069	0.009
	(0.074)	(0.075)	(0.054)	(0.078)	(0.066)	(0.074)	(0.081)	(0.067)
Control mean at endline	0.58	0.52	0.70	0.56	0.36	0.54	0.51	0.46
P-val, group = ind.	0.61	0.81	0.87	0.71	0.29	0.39	0.56	0.27
P-val, group = ind. light	0.19	0.13	0.28	0.24	0.11	0.27	0.08	0.13
P-val, ind. = ind. light	0.47	0.20	0.45	0.47	0.59	0.81	0.34	0.67
Adjusted R-squared	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
Number of observations	7534	7534	7534	7534	7534	7534	7534	7534

Notes: Outcome variables are dummy variables indicating whether a respondent considers herself capable of performing a particular task. From left to right, these are: Collecting debts, furthering the education of children in one's household, increasing the number of cows one's household owns, accessing medical care, finding new business opportunities, protecting businesses from harm/loss by someone else, resolving an interpersonal dispute, and managing one's own business. Respondents could answer "No, cannot," "Maybe," or "Definitely." We recoded variables so that "No" = 0, and "Maybe" or "Definitely" = 1. Standard errors clustered at the branch level are shown in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

B Variable Definitions

All variables expressed in monetary values are reported in 2024 PPP, converting from local currency units to PPP dollars as of the year of endline outcome measurement: 2021 for Uganda and the Philippines and 2022 for Bangladesh (Bank, 2025), and then adjusting to 2024 using the US CPI (U.S. Bureau of Labor Statistics, 2025). One 2021 PPP dollar equals 1315.27 Ugandan shillings, 20.75 Philippine Pesos, and 0.89 2024 PPP dollars. One 2022 PPP dollar equals 29.13 Bangladesh taka and 0.96 2024 PPP dollars.

B.1 Primary outcomes

Table B.1: Definition of Primary Outcomes

Primary Outcome	Description	Calculation
Consumption	Monthly adult equiv consumption (PPP 2024)	Per capita household consumption is calculated by summing the following: (1) food consumption over the past seven days (including consumption out of purchases, consumption out of home produce, and consumption received in-kind/free.); (2) non-durable good consumption and frequently purchased services during the last 7 days and 30 days; (3) and semi-durable goods and durable goods and services during the last 12 months; and (4) 10% of the value of household durable assets. Adjusted to adult equivalents and scaled to monthly values (excluding durable assets). Adjusted to 2024 PPP dollars.
Food security	Food security index	In Uganda and the Philippines, normalized average of FCS and negative HFIAS standardized to the control group. In Bangladesh, a 6-item scale with substantial overlap with the HFIAS is used, also as a normalized average standardized to the control group.
Subjective well-being	Subjective well-being index (Cantril and Kessler)	Z-score index of Negative Kessler (K6) score (in Uganda and the Philippines) or the Center of Epidemiologic Studies Depression Scale (CES-D 10) (in Bangladesh) and Cantril's Ladder (in all three countries), standardized to the mean of control villages. Negative Kessler Score: Kessler Psychological Distress Scale (K6) sums responses to six mental health questions and ranges from 0 to 24. Reverse-coded so that higher scores indicate less psychological distress and normalized to the control group. Center of Epidemiologic Studies Depression Scale: The CES-D 10 scale is a 10-item scale designed to measure depressive symptomatology in the general population. Scores range from 0–30 and are reverse-coded where appropriate. Cantril's Ladder: Average of four (three in the case of Bangladesh) questions, each ranging from 1 (low) to 10 (high), normalized to the control group.
Assets	Value of productive assets (PPP 2024)	Sum of (1) value of non-fixed durable productive assets, (2) value of livestock fixed assets, (3) total value of livestock owned, (4) total value of business inventory. Excludes imputed values in Uganda. Data on business inventories were not collected in Bangladesh. Adjusted to 2024 PPP dollars.
Income	Total monthly household income (PPP 2024)	Sum of paid labor income, business profit, net income from livestock, and value of crop production (net of costs) across all household members. Data on crop production was not collected in Bangladesh. Adjusted to 2024 PPP dollars

Table B.2: Secondary economic outcomes

Outcome	Description
Own/use land	Whether own land
Use land	Whether use land
Plot value	Value of land owned or rented, in PPP 2024 dollars
Plot size	Size of land cultivated, in square meters
Hours HH worked, past 7 days	Total hours in past 7 days dedicated by the respondent or any other household member to all remunerative work, including casual hours, livestock hours, and other hours
Hours HH worked in livestock, past 7 days	Total hours in past 7 days dedicated by the respondent or any other household member to livestock rearing and to other forms of agricultural self-employment.

B.2 Secondary outcomes

B.2.1 Social outcomes

- **Social capital index**
  - **Uganda:**
    - The indicator is calculated as the mean of bonding and bridging social capital, each ranging from 0–6. Respondents choose whether they could turn to or help each of the following: relatives, non-relatives in my ethnic group/clan, and non-relatives in other ethnic group/clan.
    - **Question wording:**
      - Bridging Social Capital (0–3): "If your household had a problem and needed help urgently (e.g., food, money, labor, transport, etc.), who OUTSIDE THIS VILLAGE could you turn to for help?"
      - Bridging Social Capital (0–3): "Who OUTSIDE THIS VILLAGE would you help if they needed help urgently (e.g., food, money, labor, transport, etc.)?"
      - Bonding Social Capital Index (0–3): "If your household had a problem and needed help urgently (e.g., food, money, labor, transport, etc.), who IN THIS VILLAGE could you turn to for help?"
      - Bonding Social Capital Index (0–3): "Who INSIDE THIS VILLAGE would you help if they needed help urgently (e.g., food, money, labor, transport, etc.)?"
  - **Philippines:**
    - Sum of binary responses to four questions.

    - **Question wording:**
        - "Has your household provided social assistance in the past 12 months?"
        - "Did you receive assistance from community groups in the past 12 months?"
        - "Did you receive assistance from family outside your household in the past 12 months?"
        - "Did you receive assistance from neighbors and friends in the past 12 months?"
  - **Bangladesh:**
    - Sum of binary responses to four questions about involvement in the shalish, or village courts.
    - **Question Wording:**
        - "Do you think you will get help from shalish?"
        - "Can you participate in shalish willingly?"
        - "Can you give an opinion in shalish?"
        - "Did you vote in the last national election?"
        - "Are you treated as equal by villagers richer than you?"

- **Trust index**
  - **Uganda:**
    - Sum of 5 variables (5-item Likert scale).
    - **Question Wording:**
        - "People in this community trust one another to borrow and loan items."
        - REVERSE CODE: "I am jealous of others in my community."
        - REVERSE CODE: "Others in the community are jealous of me."
        - "If I have a problem, there is always someone to help me."
        - "Most people in this community are willing to help if you need it."
  - **Bangladesh:**
    - Sum of 3 binary variables.
    - **Question Wording:**
        - "Do you think other villagers will share/lease/mortgage land with you?"
        - "Do you think most people can be trusted?"
        - "If you lost Tk 15,000 and a person around you found it, how likely is it to be returned?"

C Program details

The Uganda program was implemented by the AVSI Foundation as part of a USAID Resilience Food Security Activity (RFSA). The 2019–2021 program targeted households in the Rwamwanja refugee settlement and surrounding host communities. AVSI conducted a community-led poverty mapping and household visits to determine eligibility in the study's 114 village clusters (42 in the settlement and 72 in the surrounding host community). Roughly 70% of households met eligibility criteria: being "extremely poor" or "poor" with a woman or youth as household members who are—or could be—economically active.

The Philippines program was developed and implemented by BRAC, a large Bangladeshi NGO that played a pioneering role in the development of the graduation approach, and the Philippine Department of Labor and Employment (DOLE). DOLE provided financial support for group livelihoods as an extension of its existing DOLE Integrated Livelihood and Emergency Program (DILEEP). BRAC recruited households from a list of those who had begun receiving the government's conditional cash transfer program ("4Ps") in 2015–2017. It then organized participants into groups, procured asset packages, conducted livelihood training, and provided regular coaching. Assets were transferred in 2019. From the 4Ps list, 80 households per barangay were selected for the survey, using a randomly ranked backup list of up to 30 additional households.

The Bangladesh program was also developed and implemented by BRAC. The 2016 BRAC Ultra-Poor Graduation Program (UPGP) targeted the poorest households in selected rural upazilas (subdistricts). The 2016 UPGP was implemented in 126 BRAC branches across 26 districts. We work only in those 11 districts where a previous version of the program had never been implemented. Households were selected as follows: First, the poorest rural *upazilas* (subdistricts) in which BRAC operates were identified using World Food Program poverty mapping tools (Ahmad et al., 2010). Within these *upazilas*, BRAC field staff identified communities with the highest concentration of poverty. They then worked with local communities to list the poorest households via participatory wealth ranking (PWR), which was validated against eligibility criteria using a BRAC survey.

C.1 Uganda

The graduation model in Uganda comprised six components: consumption support, a cash lump sum "asset" transfer, coaching, core training and skills, savings, and business linkages. In addition to the two program versions studied here—group coaching and individual coaching arms—another set of randomly selected households in treatment villages received a package identical to the individual coaching arm but *without* the asset transfer.

1. **Asset transfer** Participant households received an asset transfer beginning in July 2019, which was delivered on a staggered basis. Participant households received 1,100,000 UGX to use as start-up capital after submitting business plans approved by their coaches. Households did not need to invest the start-up capital according to their business plans.

2. **Coaching**

   (a) Individual Coaching: Individual coaches, who most closely resemble social workers in their job responsibilities, met one-on-one with individual participant households. These coaches guided participant households on AVSI's graduation pathway, helping them to identify, manage, and reach their personal goals. Each individual coach was assigned 23–27 households, meeting biweekly for nearly 30 months. Each meeting lasted, on average, 60 minutes. Topics of discussion include nutrition, water, health and sanitation, savings, business, and preventive health strategies.

   (b) **Group Coaching**: Group coaches, drawn from the same pool of coaches as those providing individual coaching, met with participant households in groups ranging between 23–27 households. Rather than one-on-one meetings, group coaches delivered their program in group settings. Each group coach managed three groups in their portfolio. Group coaches did not conduct home visits except in exceptional cases. Group coaches meet with their groups weekly for two hours.

3. **Business Linkages**: In May 2019, AVSI, in partnership with NGOs and the local government, held a three-day workshop to connect beneficiary households with private sector vendors and market-linked NGOS. During the first two days, the organizers allowed only Graduating to Resilience beneficiary households to attend the event. During the third and final day, organizers opened the event to all surrounding households. The organizers invited seven main categories of exhibitors to attend the event: (1) big agribusiness firms, (2) agricultural tool and equipment suppliers, (3) agro-processing, post-harvest handling and value addition technology suppliers, (4) green energy solution service providers, (5) produce buyers, (6) providers of market information and extension, and (7) financial service providers. Overall, this market event served as the beginning of AVSI's work to connect private-sector agricultural input suppliers with participant households. The objective of this component was to help participant households access quality seeds and agricultural inputs that AVSI determined were not

readily available in the local market context. Over the course of the program, AVSI worked to engage with these external service providers.

4. **Consumption support**: Households received a monthly cash stipend for the first 12 months of the program. The stipend was valued at 18,000 UGX per household member per month for refugee households, and 15,000 UGX per household member per month for host households.

5. **Core Training and Skills**: Throughout the program lifecycle, AVSI delivered a series of trainings to improve household livelihood management, separately from coaching. These include selection, planning and management training designed to help households form sustainable businesses, agronomic training to improve household farming practices by introducing new farming techniques (e.g. use of organic manure, fertilizer), and business coaching to help participant households manage their businesses.

6. **Savings**: AVSI formed Village Savings and Loan Associations (VSLAs) with the members of each implementation group. A VSLA is a member-run financing mechanism that helps participant households save weekly.

C.2 The Philippines

BRAC USA implemented the graduation pilot in partnership with the Philippine government's Department of Labor and Employment (DOLE), with financial support from the Asian Development Bank. Prior to program launch, BRAC's conducted a market analysis to identify potential livelihoods that were likely to be feasible and profitable in this setting. They used the results of this assessment to construct livelihood packages from which participants could choose.

At the beginning of the program, BRAC Graduation Coaching Facilitators (GCFs) used a market assessment tool to determine which livelihoods would be appropriate for each community. This tool had a list of requirements for each livelihood to help GCFs cross-check. For example, livelihoods like swine, working carabao, and chicken egg-laying require abundant space, which beneficiaries in populated lowland areas did not have. As Table C.1 shows, most beneficiaries (54%) chose swine fattening across all treatment groups. This was followed by business cart or NegoKart (15%), meat processing (13%), and free-range chicken (8%).

We observe a key difference between group and individual coaching arms regarding their choice of livelihoods, as detailed in Table C.1. BRAC coaches highlighted different livelihood options through the workshops where these selections were made. Since participants in the

individual coaching arm (T2) made their selections just after those in the group coaching arm (T1), their choices differed slightly. T1 members were less likely than T2 members to select swine fattening (37% vs. 77%), with a larger share participating in a prepared food stand (21% vs. 9%) and free-range chicken raising (22% vs. 0%). As a result, the comparison between group and individual coaching in the Philippines reflects not only differences in coaching modalities, but also differences in livelihood selection.

The graduation program was built upon the existing DOLE Kabuhayan (Livelihood) program, which targets poor and vulnerable households and had been previously implemented with "pre-formed" applicants, who applied as a group for seed capital. In this study, individuals formed groups within their barangay quadrant and applied for group or individual livelihoods, depending on their quadrant assignment.

Besides the working capital for selected projects, beneficiaries also received training on setting up and running the livelihood project as well as technical and business advisory services to support the sustainability of the business.

In addition to the livelihood training and asset transfer, participants received regular coaching from BRAC GCFs delivered fortnightly in individuals or groups. After facilitating the asset transfer and technical and business management training for households, GCFs provided life-skills training on topics ranging from household financial management to water, sanitation, and hygiene, health, domestic violence, and child nutrition, tailored to the most prominent issues among the poor in the community. Additionally, coaches provided access to savings and financial inclusion mechanisms as well as troubleshooting business development concerns for households to grow and improve upon their enterprises. Finally, coaches promoted social integration through supporting community-led activities and linking households to community associations or committees, including local government officials and initiatives.

C.3 Bangladesh

C.3.1 Program eligiblity

The program has five inclusion and two exclusion criteria. To be eligible, a household must fulfill at least three of the inclusion criteria and none of the exclusion criteria. *The inclusion criteria*: 1. The household depends on female day labor for income; 2. It has less than or equal to ten decimal land[18]; 3. It has no productive assets; 4. The household has children of school-going age (6–14 years) who have to work; and 5. The household has

[18]A unit of measurement used in rural Bangladesh, a decimal is one hundredth of an acre of land, or 40.47 square meters.

Table C.1: Philippines: Program livelihood asset chosen by arm

Livelihood	Total households	Group	Individual
Swine Fattening	545	77%	37%
NegoKart Business	170	9%	21%
Meat Processing	133	9%	15%
Free Range Chicken	99	0%	22%
Working Carabao	46	0%	0%
Chicken Egg Production	32	0%	0%
Backyard Vegetable Farming	24	2%	3%
Salted Eggs Processing	16	2%	2%
Cosmetology/ Massage	4	0%	1%
Fish Vending	1	0%	0%
Total	1134	100%	100%

no adult male members. *The exclusion criteria*: 1. The household has no female members capable of working; and 2. The household has one or multiple members that are participants in other government/non-government development projects.

C.3.2 Program components

We describe below the key components of the program. Where treatment is at the individual level, the recipient is always the main woman in the household. Unless explicitly stated otherwise, the treatment elements described below apply to the interventions received by all treatment groups in the study. On average, approximately five percent of households in treated villages received the intervention, although this proportion was somewhat variable.

1. *Asset transfers*:

   At the start of the program, the main woman in the household is presented with a menu of assets, each of which can be used productively in an income-generating activity. These include, for instance, livestock, land, and capital equipment. Nearly all recipients (97%) chose a package with either cows and/or goats, and 76% chose a package containing cows. Fewer than 3% received a package that included land or non-agricultural capital goods.

   Prior to the asset transfers, BRAC staff provide recipients with classroom training on how to manage the asset for income-generating activities. Assets were transferred

within a month of this training session, and BRAC encouraged beneficiaries not to liquidate the transferred assets for at least two years after receipt, although there were no hard sanctions if the asset was liquidated prematurely.

Table C.2: Bangladesh: Program livelihood asset chosen by arm

Asset package	**Total households**	**Group**	**Individual**	**Individual Light**
Cow	4255	79%	77%	77%
Goats	1030	18%	19%	20%
Other	178	3%	4%	3%
Total	5463	100%	100%	100%

Notes: The 'Cow' asset package is defined as an asset package that consists of one cow and either one or no goats. The 'Goats' asset package is defined as an asset package that consists of four or five goats and no cows. The 'Other' package is any other combination of assets. The average value of the 'Cow' asset package is \$500, the 'Goats' asset package is \$370 and the 'Other' package is \$240 (prices in 2020 \$PPP).

2. *Coaching*

All beneficiaries receive regular one-on-one coaching through home visits by BRAC program staff. As described above, BRAC introduced experimental variation in whether this coaching was received on a weekly or a fortnightly basis, as well as whether a group coaching element was included. For all treatment groups, these coaching sessions included life skills training, such as guidance on health and education-related issues, enterprise management skills, and guidance on how to care for the asset transferred through the program, among other things. Health topics included family planning, prevention and treatment of waterborne diseases, non-transmittable diseases and hookworm, vaccinations, and nutrition. Coverage of social issues included, for instance, guidance on how to further the education of children. The participants also received coaching meant to boost confidence, were prompted to set goals, were taught about their rights and responsibilities, and were encouraged to save and invest in simple businesses such as retail trading. They were also informed about different government and non-government services, such as access to credit and eligibility for support through various government rural development schemes.

An important part of the home visits was to follow up on the status of the asset. For

instance, coaches observed whether the asset was being used for income generation and discussed challenges that the beneficiary may have experienced in putting the asset to productive use. If the asset was livestock, coaches also helped ensure that it was healthy and receiving appropriate care. For instance, coaches typically provided training on animal husbandry best practices, such as insemination processes, vaccination, feeding, milking, and so on.

In summary, coaching comprised several elements serving different purposes and potentially operating through different mechanisms. Training was provided to transmit hard skills along with soft skill development, beneficiaries were exposed to mentors who might serve as role models, beneficiaries could access other social services through the intermediation of the coach, and the regular visits served as a compliance check, ensuring that beneficiaries were using the transferred assets as intended.

Coaching sessions lasted on average approximately 30 minutes and continued for 18 months after the initial asset transfer. Coaches, also called BRAC program officers, were young and well-educated: they were required to have a masters' degree and be under the age of 35. Each branch had on average three program officers serving between 80–120 households.

3. *Group meetings (Group coaching only)*:

   For the group meetings, a BRAC coach convened a gathering of between 3–8 beneficiaries in a village. These group meetings were typically held on the same week as the fortnightly home visit in one of the beneficiary's houses. Beneficiaries of the T1 group variant received the subsistence allowance stipend once every two weeks during these group meetings instead of during home visits. Group meetings were not set up as rotating savings and credit associations.

   The topics covered during the group visits were the same in substance as those covered during the one-on-one coaching home visits. Group-based instruction and the moderation of these sessions was also undertaken by the same BRAC coaches who were responsible for the one-on-one sessions. However, while BRAC program officers also used the group meetings to provide instruction on various social and health-related matters, the group setting provided beneficiaries with a less structured forum in which to discuss concerns and raise questions. The group meeting was also a potential opportunity to create social networks between the beneficiaries. In addition, group-based coaching may be complementary to one-on-one house visits in other ways, for instance, representing a potentially more interactive and engaging approach to hard- and soft-skill acquisition, a more binding form of commitment device, and allowing recipients

to learn from the experiences of other recipients.

4. *Subsistence Allowance*:

   All beneficiaries receive regular consumption support in the form of small, unconditional cash grants. These cash grants were delivered in person by a BRAC officer during either weekly or fortnightly coaching sessions. [19] Consumption support in the form of this subsistence allowance lasted for the first 40 weeks of the program. Recipients were encouraged to save a small amount of the transfer, with a target set at approximately 10 percent.

5. *Health and Miscellaneous Social Services*:

   The beneficiaries were provided with access to medical support and encouraged to seek the assistance of a BRAC health worker if a family member was sick. BRAC health workers could provide basic medicines, give medical advice, and refer household members to local hospitals and health centers.

C.3.3 Nested experiment on community mobilization

In the 2016 cohort, BRAC also introduced experimental variation in community mobilization within the group coaching treatment arm (T1). Specifically, in all treatment villages apart from a subset within T1, informal institutions, called VACs or "Village Poverty Reduction Committees" were set up. VACs were *not* established in approximately half of those villages in T1, with randomization at the village level. Our records indicate that of 2328 households in T1, 1029 were in villages assigned to the establishment of VACs and 1299 were in villages in which VACs were not set up. However, errors in record-keeping prevent us from being confident of the treatment status for all villages, and so we do not include it as a covariate. Appendix Table A.11 reports results after controlling for this compromised treatment assignment variable.

These committees were created with the intention of inducing community mobilization, linking villagers across social strata, and creating mechanisms for better-off villagers to support the ultra-poor. The stated goal of these committee structures was to support the economic security of UPGP beneficiaries, create an enabling environment for the

[19] The total value received did not differ between beneficiaries who received weekly versus those who received fortnightly training—the latter simply received double the amount half as frequently. Since the amounts are small in absolute terms, we do not expect that the differences in the frequency of transfers would affect investment or have enduring impacts on consumption.

beneficiaries in the village, help recipients protect their assets, increase access to government services, and offer support in times of need by coordinating local community support. Committees consisted of 9–11 villagers and were comprised of UPGP recipients and community leaders—for instance wealthier land owners, teachers, imams, and local authorities. The VAC would meet monthly and discuss issues raised by the UPGP recipients. Committees were typically established within three months of the asset transfer.

D Experimental details

D.1 Uganda

D.1.1 Study area and listing

The study area comprises the entire refugee community of the Rwamwanja settlement and four subcounties contiguous to the settlement (Nkoma, Biguli, Bihanga, and Bwizi). Between July–September 2018, AVSI conducted an individual household poverty scorecard exercise and a community-based poverty ranking covering all households in the study area. AVSI reached 35,204 households and identified 25,104 households as eligible for program participation—households that were assessed as poor or extremely poor and included a woman or youth who could be economically active as part of the program. The program had an implementation target of 6,600 households evenly split by host and refugee communities. To accommodate implementation targets and the study design with four experimental groups in treatment clusters (three treatment arms and a spillover control) plus a "holdout" control for future implementation, each with a minimum target implementation group size of 25 members, four large villages were split and 14 small villages merged to create a total of 42 village clusters for the study in the refugee settlement and 72 village clusters in the host community. The study target sample consisted of a total of 11,000 households randomly sampled from the set of eligible households, split evenly across five experimental groups (three treatment groups plus a spillover control in treatment clusters and a pure control in control clusters), with a target of 2,200 households in each group evenly split between host and refugee communities. The final study sample consists of 11,145 program-eligible households interviewed for the baseline data collection, 10,514 of which were interviewed during the endline survey.

D.1.2 Randomization

We randomized village clusters into treatment and control, separately by host and refugee communities. In host clusters, random assignment was stratified by parish. In refugee clusters, strata were based on distance to the economic center of the settlement and on whether a cluster was above or below the median poverty score in its respective geographic grouping. For village-cluster assignment, we used a re-randomization procedure that repeatedly drew candidate treatment assignments and selected the draw that achieved the best balance on a list of covariates. Randomization was conducted separately by community type: refugees first, with host communities randomized later. We repeatedly assigned communities to treatment or control (100 candidate draws for refugees, 500 for hosts) within stratification cells

(geography × above/below-median village average vulnerability score for refugees; parish for hosts), merged each draw to implementer listing data, and assessed whether the balance covariates predicted treatment assignment. The balance covariates were vulnerability score, household size, female household headship, and eligible household count per village cluster. We selected the final assignment as the draw with the best balance (balance criterion for refugees: lowest joint F-statistic on the covariates; for hosts: highest minimum p-value across the single-covariate balance tests).

Within treatment clusters, households were then individually randomized with equal probabilities via public lotteries into one of four experimental conditions, which include a spillover control and a no-asset arm in addition to the two arms that vary only by coaching modality that are the focus of this paper.

D.1.3 Imputation

In Uganda, to limit the survey length, households were randomly assigned to complete a full ("long") or shorter versions of several survey modules. Randomization to short versus long survey modules was done separately by groups of modules, resulting in only a limited subset of households receiving all component modules needed to compute the productive assets and income aggregates. The depth of the information available in short modules varied across questions but typically included at least gateway questions for whether the household had any of a given activity/purchase/item category (e.g. any non-farm business, any purchase in each agricultural input category, any livestock structure for each livestock structure type) and sometimes included additional detail (e.g. business profits for the most important business; crop income from the most important crop, number of animals for all livestock types).

Appendix Tables D.1 and D.2 show the allocation to component modules for the asset and income aggregates.

Table D.1: Short Survey Patterns for Income Modules

Pattern	Employment (E)	Business (B)	Livestock (L)	Crops (C)	Frequency
`----`	0	0	0	0	683
`---C`	0	0	0	1	1,375
`-B--`	0	1	0	0	1,609
`-B-C`	0	1	0	1	1,589
`E-L-`	1	0	1	0	1,616
`E-LC`	1	0	1	1	1,586
`EBL-`	1	1	1	0	1,364
`EBLC`	1	1	1	1	692

Table D.2: Short Survey Patterns for Asset Modules

Pattern	Asset (A)	Business (B)	Livestock (L)	Frequency
`---`	0	0	0	2,058
`--L`	0	0	1	3,202
`AB-`	1	1	0	3,198
`ABL`	1	1	1	2,056

In our analysis, we conduct a multiple imputation procedure for these missing values in order construct aggregates for the full sample. Importantly, the imputations are done separately by treatment arms (and refugee status). We use Multiple Imputations by Chained Equations (MICE) to fill in missing values for each sub-component of the aggregates based on information collected in the short survey and the (imputed) value of the other sub-components. We specify a predictive mean matching procedure, which matches predictions from a linear model with the nearest neighbors in the distribution of the non-imputed values, to create 80 imputed data sets. The asset and income aggregates are recombined as the sum of the imputed components (where missing).

We use Stata's `mi estimate` to fit the treatment-effect regressions on the imputed aggregates. `mi estimate` combines point estimates and variances across imputations using Rubin's rules, so reported standard errors reflect between-imputation variability (i.e., imputation uncertainty). Since short survey modules were randomly assigned and we impute separately by experimental conditions, the imputed outcomes—and resulting treatment effect point estimates for the aggregates—match closely with the outcomes for observations where the full information is available.

For lasso selection of covariates, we use a randomly selected single-imputed dataset for our income and asset outcomes. We similarly use a single randomly selected imputed dataset when testing for treatment heterogeneity.

D.2 Philippines

D.2.1 Study area and listing

The original sample included poor households in 29 barangays across five municipalities in northern Negros Occidental who receive the government's conditional cash transfer program and were added to the program during the same two-year period between 2015 and 2017. The following describes how the final list of households eligible for inclusion in the pilot program was determined.

During the inception workshop that included representatives of ADB, BRAC, DOLE,

and IPA, it was agreed that the study sample would consist of 2,400 households across 30 barangays (80 households per barangay). In May 2018, IPA received a list of households that were all added to the government's conditional cash transfer program (4Ps) during the same two-year period. The original list consisted of approximately 3,200 households from 32 barangays that had 80 or more eligible 4Ps recipients. The number of households per barangay eligible for inclusion in the study varied: some barangays had more than 200 names, others just over 80. IPA defined the final master list of households as follows:

1. The number of sample barangays was reduced to 30 by dropping the two barangays with the fewest number of eligible households: barangay Alegria in the municipality of Murcia (81 eligible households) and barangay VI-A in the City of Victorias (81 eligible households).

2. IPA then randomly sampled 110 households per barangay from barangays with more than 110 eligible households. This reduced the likelihood that surveyors would skip respondents unavailable upon their first interview attempt, which has the potential of introducing selection bias, and left a sufficient number of "reserve" households in case the list proved outdated, or a significant fraction of respondents could not be reached during the 3–4 days IPA field staff spent in each barangay.

   After random sampling, the master list numbered 3,098 households eligible for inclusion in the study.

3. The list of up to 110 names per barangay was then randomly sorted. For each barangay, surveyor teams were provided with a hard copy list of up to 110 respondents and instructed to begin with the names at the top and work their way down until the team reached 80 interviews. They were granted permission to skip a name only if a respondent could not be reached in the 3–4 days of fieldwork allotted per barangay.

D.2.2 Randomization

Households were randomized in two stages:

1. Each barangay was divided into four clusters based on GPS coordinates collected during surveying. Households were assigned points based on their distance to the nearest cluster's center minus the distance to the farthest and ranked accordingly. Then each of the four clusters were filled with the highest-ranked households until 20 households were placed, at which point that cluster was taken out of the equation and the next cluster filled with the highest-ranking households.

This process was repeated until all 120 clusters of 20 households across the 30 sample barangays were filled. Once all households had been assigned to on of four quadrants in their barangay, we randomized assignment of these quadrants into one of three treatment arms or a control group.

We implemented a re-randomization procedure that repeatedly drew candidate treatment assignments and selected the one that yielded the best balance across treatment arms for our target covariates: household size, number of adult household members, total value of durable assets, and total value of livestock. For each of 200 replications, we randomly assigned quadrants to one of four treatment arms within each community. We then merged this assignment with household-level data and tested whether the target variables predict treatment assignment using a community-level fixed effects regression, comparing treatment arms in pairs and clustering standard errors at the quadrant level. For each replication, we record the maximum joint F-statistic from the set of pairwise balance tests, which reflects the worst imbalance across the set of comparisons. We then select the final randomization as the one that minimizes this worst-case F-statistic.

2. Following randomization, BRAC field staff conducted a validation survey of households in treatment areas. The team found several households that were recipients of DSWD's Sustainable Livelihood Program (SLP). BRAC identified 44 households that had or were participating in SLP. In one barangay (Barangay VI, Poblacion), more than half of study households were part of SLP, so BRAC and IPA agreed to exclude it, leaving 29 participating barangays.

D.3 Bangladesh

D.3.1 Study area and listing

This study draws a sample from the 2016 cohort of the BRAC UPGP intervention, a scaled version of BRAC's graduation program delivered to 80,000 households. A subsample of those identified as eligible for the 2016 intervention were selected for the experimental evaluation.

Initial selection into the program proceeded in the following steps: First, the poorest rural *upazilas* (subdistricts) in which BRAC operates were identified using poverty mapping tools developed by the World Food Programme (Ahmad et al., 2010). Within these *upazilas*, communities with the highest concentration of poverty were identified based on the knowledge of existing BRAC field staff already operating in these areas (for instance, in BRAC's

microfinance, health, and education programs). Next, BRAC officers produced a list of the poorest households in these communities using participatory wealth ranking (PWR) exercises in which all households within a community are ranked in terms of material well-being by villagers in an open meeting [20] The PWR serves to classify all households into categories such as very poor, poor, middle-class, and non-poor. Based on this classification, BRAC program staff then administer a questionnaire to those households in the village who were classified as "very poor" or "poor" to generate a preliminary list of ultra-poor households. This survey serves to verify whether particular households met the inclusion and exclusion criteria set by BRAC.[21] The primary female adult in those households who are verified as eligible through this survey were then registered as beneficiaries.

The 2016 UPGP was implemented in 126 branches from 26 districts. Of these, 11 districts where the program had never been implemented before were selected for inclusion in the experimental evaluation. The program then selected eight branches from each of these selected districts. These branches were then randomly assigned to each of the four treatment or control groups. In total, 8,468 women and their households completed baseline surveys, and they constitute the study sample.

D.3.2 Randomization

A baseline survey was conducted between April and August 2016, prior to randomization or the implementation of the intervention. Baseline surveys were completed by 8,468 recipients and their households, drawing a sample of between 50 and 200 eligible ultra-poor households from each branch, which is the smallest unit in BRAC's administrative structure.

The study covers 88 branches from 11 districts. These 88 branches were randomly assigned to three treatment groups and a control group of 22 branches each. Randomization

[20]The process starts with rapport building where a program staff walks around a village with a few villagers to decide whether there is a substantial concentration of poor households in the village. Once a village is chosen, the program staff engages with more villagers to identify a venue for the PWR. Villagers were invited for the PWR, which takes the form of a group discussion, usually on the following day. While the PWR may have differed from village to village, it usually included a group of about 40–50 villagers. A map of the village was drawn on the ground and villagers discussed the economic condition of each household as it was pointed out on the map. In the process of ascertaining economic conditions, the participants were asked to keep in mind issues such as the roofing material of the household, if any children were out of school, whether members had a steady income source, had any productive assets, et cetera. This was done for every household in the village.

[21]The program has five inclusion and two exclusion criteria. To be eligible, a household must fulfill three of the inclusion criteria and none of the exclusion criteria. *The inclusion criteria*: 1. The household depends on female day labor for income 2. It has less than or equal to ten decimal land 3. It has no productive assets 4. The household has children of school-going age (6–14 years) who have to work 5. The household has no adult male members. *The exclusion criteria*: 1. The household has no female members capable of working. 2. The household has one or multiple members that are participants in other government/non-government development projects.

was stratified by district.

E Heterogeneity

We follow the generic machine learning procedure outlined by Chernozhukov et al. (2025) to test for differential impacts by sub-groups, utilizing the `genericML` package by Welz et al. (2022) with the set of covariates that had above average influence for each outcome variable. Specifically, we do the following for each outcome, considering each program separately:

1. We restrict our sample to those receiving group or individual coaching, classifying group coaching as the "treatment group."

2. We identify the list of key baseline covariates. Note that all covariates are continuous or binary.

3. For each outcome variable, we train a random pilot forest using the training sample and the potential covariates as a pilot. We restrict the set of potential covariates to those with above-average importance, following Athey and Wager (2019).

4. Using the `genericML` package, we use three potential learners: lasso, random forest (with 1,000 trees), and a support vector machine.

    - To ensure results are robust to how the data are split between a "main" and "auxiliary" sample, for each of 250 splits, we train the machine learning algorithm on the auxiliary sample and use those estimates to predict outcomes in the main sample. We take the median of these results over 250 random sample splits.
    - We account for clustering by quadrant in the Philippines and by village in Bangladesh. Conditional on village-level assignment to treatment, randomization of coaching modality in Uganda was implemented at the household level, so we do not cluster standard errors.

5. We then estimate group average treatment effects (GATES) based on terciles.

F Cost-effectiveness analysis

We calculate the net present value of all benefits based on the estimated year 3 (year 6 in Bangladesh) impact on consumption and assumed social discount and persistence rates. We use that same social discount rate to inflate the incurred costs to year 3 (year 6 in Bangladesh). We follow (Banerjee et al., 2015) and assume a social discount rate of 5%. For this example, we take a persistence rate of 100%, although Table 2 reports the estimated benefit-cost ratio at persistence rates of 90% and 80%, following the same strategy.

We report program costs and benefits in PPP 2024 dollars (Bank, 2025). We assume that all program expenses were incurred during the first year of implementation. Expense data was collected in USD, which we inflation adjust to the endline survey year (2021 in Uganda and the Philippine and 2022 in Bangladesh) before converting to local currency units and then into PPP dollars, ensuring the LCU to PPP conversion occurs in the same year for both costs and benefits. We then use the US CPI to inflation-adjust to 2024 PPP dollars.

F.1 Program costs

In the Philippines, we received cost estimates from BRAC, which we reconciled with funding requests from the Asian Development Bank, which funded coaching support and the asset transfer for the individual-coaching treatment arm. This funding covered the time of both field and supervisory staff within BRAC, as well as staff time at higher management levels that was specifically dedicated to this project. The livelihood project is integrated into the Department of Labor and Employment's DILEEP (or Kabuhayan) program, and we account for the value of these assets and note that training was paid for by BRAC. However, we are not able to account for the value of staff time within DOLE to support the livelihood program, which includes reviewing proposals, approving and monitoring procurement, and liaising with BRAC. Consequently, we may underestimate the overall administrative program costs.

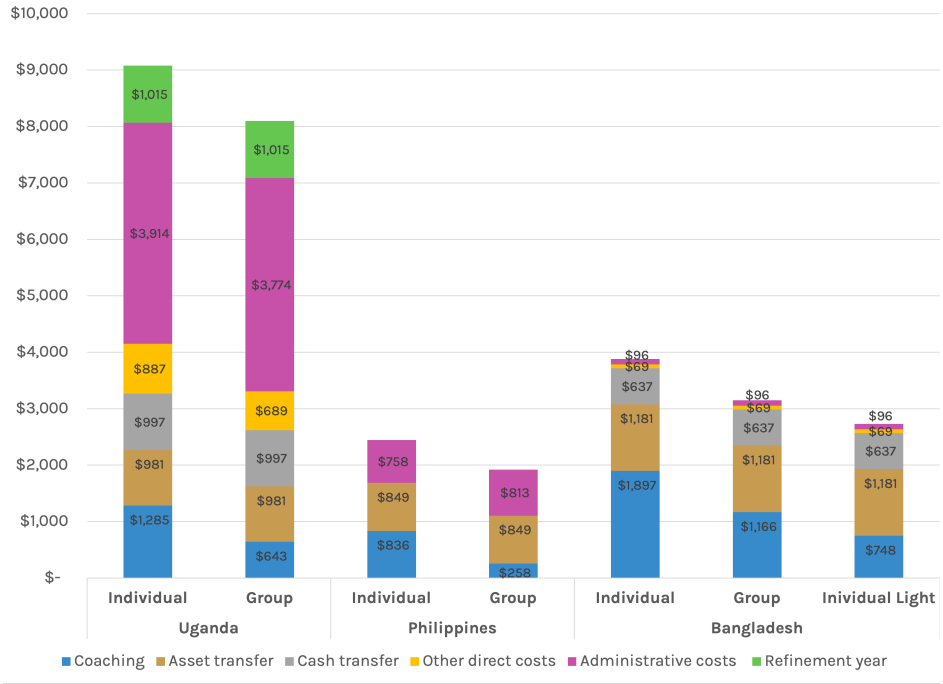
Program implementation costs in Uganda comprise the value of consumption support, asset transfers, coach and trainer salaries, other direct costs, and project management and monitoring and evaluation expenses. AVSI also spent one year refining the activity prior to launch. Because we anticipated that this refinement would benefit both this cohort and a second cohort, we ascribe half of the refinement cost as part of the overall program cost.

Cost estimates were harder to obtain in Bangladesh. BRAC either did not record or has been unable to provide detailed cost data disaggregated by treatment arm. They have only been able to supply total cost data for the entire 2016 cohort of 80,000 recipients—of which approximately 6,000 of which were included in the experiment. We therefore had to make

assumptions about the per-recipient cost and about the cost differential across treatment arms. We can break down the cost estimates provided between costs that are likely to remain fixed across treatment groups (i.e. the price of the asset, etc) and coaching-related costs that are likely to vary (coaching staff salaries, staff transport, etc). We estimate that 37% ($262) of per-participant program costs would be sensitive to the amount of coaching delivered, while the remainder would be fixed across participants ($496). We estimate that, compared to the fornightly-plus-group coaching arm, BRAC saves 40% of the coaching costs in the fortnightly-only arm, bringing coaching-related costs down to $187 for that arm. We then assume that doubling the amount of coaching delivered in the weekly treatment arm relative to the individual and group coaching arm increases coaching costs to BRAC by a factor of 1.7, to $474. This yields the cost figures in Table F.1 and Figure F.1.

Across all three sites, have not included the cost of beneficiary time or any resources that were expended to participate in the program, such as transportation to program activities. However, reasonable estimates of the value would not substantially change impacts.

Figure F.1: Cost per participating household



F.2 Program benefits

We calculate the estimated increase in monthly consumption per adult equivalent in Table 1 for both group and individual coaching arms. We adjust consumption to an annual household total by multiplying by the number of adult equivalents and by 12.

Table F.1: Costs per participating household

	Uganda		Philippines		Bangladesh		
	Group	**Individual**	**Group**	**Individual**	**Group**	**Individual**	**Individual light**
Coaching	643	1,285	258	836	1,166	1,897	748
Asset transfer	981	981	849	849	1,181	1,181	1,181
Cash transfer	997	997	–	–	637	637	637
Other direct costs	689	887	–	–	69	69	69
Administrative costs	3,774	3,914	813	758	96	96	96
Refinement year	1,015	1,015	–	–	–	–	–
Total	8,099	9,079	1,920	2,443	3,149	3,880	2,731

Notes: The Bangladesh program was nested within a scaled intervention of 80,000 recipients: economies of scale in delivery help explain why administrative and other fixed costs are lower in this program than in the Philippines and Uganda programs.

Consumption support and asset transfers were delivered in Year 1 in Uganda and Bangladesh, while in the Philippines the asset transfers were not completed until Year 2. For this reason, we inflate the Year 3 (Year 6) benefit by the discount rate to estimate a benefit in Year 1 and Year 2 (Years 1 to 5) in Uganda (Bangladesh). In the Philippines, we inflate the Year 3 benefit only to Year 2, and we assume there is no impact on consumption in Year 1.

We calculate the net present value of benefits from year 4 (year 7 in Bangladesh) onward assuming a persistence rate of 100% and discount rate of 5%.

We then divide the total benefit by the total inflated costs to estimate a benefit-cost ratio for each persistence rate. A ratio greater than 1 indicates that the total realized benefits exceed the costs, although we note that these only account for the value of consumption and ignore any non-pecuniary benefits of the program.