



Trinity College Dublin
Coláiste na Tríonóide, Baile Átha Cliath
The University of Dublin

What Aggregate Measures Conceal: Mixed-Methods Evidence on Women's Empowerment, Gender Targeting, and Couples Training in Malawi

Rebekah Avar, Tara Bedi, James Copestake, Michael King, and
Fiona Remnant



TEP Working Paper No. 1526

August 2026

What Aggregate Measures Conceal: Mixed-Methods Evidence on Women’s Empowerment, Gender Targeting, and Couples Training in Malawi

Rebekah Avar¹, Tara Bedi², James Copestake³, Michael King⁴, and Fiona Remnant⁵

^{2,4}Trinity College Dublin and TIME

^{1,5}Bath Social and Development Research

³University of Bath

30 August 2026

Abstract

Raising women’s empowerment through economic inclusion and social protection programmes has proved challenging. Women’s access to resources improves but across many studies measured decision-making agency barely moves. Part of the weakness in evidence may be measurement as composite indices may conceal change that occurs unevenly across domains. Within the context of a multi-faceted anti-poverty, or Graduation, programme in Malawi, we investigate the women’s empowerment impacts of gender targeting, and the additive effect of a 12-month couples economic cooperation and empowerment training. We combine the results from a randomised control trial with 2,934 women from poor married households with findings from qualitative research in which researchers collect unprompted accounts of change from the participants’ perspective to challenge and substantiate the quantitative findings. Female resources rise in all three treatment arms. The aggregate decision-making index moves only where the couples training is added, and only marginally, but beneath it financial decision-making rises in both female-targeted arms while male targeting reduces women’s say over productive decisions, the very activities the programme creates. Qualitatively, women across arms describe more collaborative decision-making over income generating activities, the essential realm of the Graduation programme, more than over budgeting, and budgeting moves mainly where the couples training addressed it directly; the driver they name unprompted is the programme’s gender messaging rather than the cash. We show that participant-led and unprompted qualitative evidence can identify which components of composite indices merit attention, independently of researcher priors.

JEL Codes: J16, D13, O12, C93

Keywords: Social Protection, Intra-household, Gender Empowerment, Measurement.

1 Introduction

Economic inclusion and social protection programmes are now a mainstay of anti-poverty policy, reaching hundreds of millions of poor households worldwide. Many of these programmes carry explicit goals for women’s empowerment, and many direct resources to the woman in the household (Fiszbein et al., 2009; Laszlo, 2019). The rationale rests on a long-standing paradigm, built on early quasi-experimental evidence, that resources controlled by women shift household spending toward food and children (Thomas, 1990; Hoddinott and Haddad, 1995; Lundberg et al., 1997), and on the expectation that targeting women advances their empowerment.

Recent experimental evidence has complicated our understanding of the relationship between these programmes and women’s empowerment. Multi-faceted ‘big push’ anti-poverty programmes explicitly targeting women find their decision-making power unchanged (Bandiera et al., 2017), improved in productive domains but not household decisions (Bossuroy et al., 2022), or, but for a notable exception in Afghanistan, improved overall (Bedoya et al., 2019). In Banerjee et al.’s six-country Graduation programme evaluation, gains in women’s decision-making either did not appear or faded over time (Banerjee et al., 2015).

The most direct evidence on targeting comes from cash transfer experiments in which recipient gender is itself randomised, and the findings are mixed. In India, randomly allocating weekly transfers to the wife rather than the husband empowered female recipients across a range of decision-making measures, despite little evidence of consumption shifting at the household level (Somville et al., 2020). In Kenya, by contrast, recipient gender made little difference to household outcomes (Haushofer and Shapiro, 2016). Only recently has the same test been applied to a full anti-poverty package: the companion paper to this study finds that women’s access to productive resources rises regardless of whichever spouse is targeted, with no difference between female and male targeting, and that targeting alone leaves women’s decision-making agency unchanged (Bedi et al., 2026c).

These disappointments have generated growing interest in engaging the couple directly in economic cooperation and empowerment training, which aims to improve spousal cooperation, shift gender norms within the household, and strengthen women’s role in decision-making. The evidence on couples trainings is promising but still emerging. Trials in Rwanda and Uganda find that such trainings can increase men’s participation in childcare and household tasks, improve communication, conflict management and power sharing within the couple, and reduce men’s dominance of household decision-making (Doyle et al., 2018; Dunkle et al., 2020; Boyer et al., 2022). When layered onto economic programming for the ultra-poor, however, the additive contribution of the training has not been separately identified: a Graduation plus family coaching programme in Burkina Faso improved women’s resource access and their participation in household decisions compared to the control arm, but found no differences between the individual treatment arms (Ismayilova et al., 2018), and gains in assets only partially translated into decision-making power (Karimli et al., 2021). The ran-

domised control trial behind this study provides that separation. Bedi et al. (2026c) finds that after 17 months the couples training raises women’s access to productive resources over and above female targeting alone, while decision-making agency scarcely moves. Across both targeting and training, then, the same pattern recurs. Women’s access to resources improves while measured decision-making agency barely moves.

Both targeting and training may simply be insufficient to shift decision-making power within the household, at least over the horizons studied. However, there is a second possibility, the instruments used to measure empowerment are not capturing the change that occurs. Empowerment is not directly observable, and the standard practice is to proxy it through survey modules on decision-making, resource control and mobility, aggregated into standardised composite indices (Kling et al., 2007; Bossuroy et al., 2022). Aggregation has real virtues, offering comparability across studies and disciplining the number of hypotheses tested (Anderson, 2008), but it comes at a cost. An index that averages across savings, purchases, production and fertility can register nothing even when women’s influence has shifted decisively in one domain, or when gains in one domain are offset by losses in another. Nor is aggregation across domains the only averaging at work: average treatment effects can likewise obscure significant heterogeneity across households, so that change concentrated among some women disappears in the mean. The underlying questions, typically of the form of who decides, reveal little about its underlying meaning where the same reported answer can reflect very different underlying dynamics between spouses (Bernard et al., 2020). Furthermore, measured decision-making power is only weakly connected to women’s revealed control over resources (Almås et al., 2018).

These are noteworthy weaknesses. Kabeer (1999b) conceives of empowerment as a process, in which resources are the pre-condition, agency the process, and achievement the outcome. As such, change should be expected to emerge unevenly, appearing first in the domains an intervention most directly touches. A composite index averages the domains where change begins with the domains it has not yet affected, and so is poorly suited to detecting a process of exactly this kind. Measurement must instead attend to which choices carry consequential significance in a given context (Kabeer, 1999a). Establishing that requires looking beneath the indices, and learning from the women themselves where they identify change has occurred in their lives.

In this paper we decompose our overall empowerment indices into their sub-components, and set the results against unprompted accounts of change from the women themselves. The setting is a Graduation programme implemented over 18 months by Concern Worldwide in Mangochi and Nsanje, the two districts of Malawi with the highest poverty index, evaluated through a cluster-randomised trial.¹ Our study involves 2,934 women from poor married households willing to engage in income generating activities, in which the programme was allocated to the female spouse, to the male spouse, or to the female spouse with the couple participating in Umodzi, a twelve-session economic cooperation and empowerment training for couples delivered over twelve months.²

This paper is one of three arising from the trial, and it is the one that takes short-

run empowerment effects as its focus.³ Looking beneath the indices, we decompose the empowerment measures into their sub-components, and compare our findings with the stories of change from programme participants gathered through a blindfolded qualitative research tool called the Qualitative Impact Protocol (QuIP).⁴ We focus on the short run, reporting quantitative results at 5 months post-intervention, the first window in which empowerment change can be observed in this trial, before some of these results fade as Bedi et al. (2026c) documents at 17 months. If empowerment is a process, as set out above, change should emerge first in this window or not at all. A single follow-up cannot establish the ordering of changes within it, but the earliest observable pattern is of importance in its own right.

Our mixed methods design follows the quant-led model of Copestake (2024). Alongside the RCT, we conducted 96 semi-structured interviews at 6 months post-intervention using the Qualitative Impact Protocol (QuIP), in which partially blindfolded local researchers, with no knowledge of the treatment arms, the hypotheses, or any respondent’s assignment, collect unprompted accounts of change and their perceived drivers. Respondents were drawn from the quantitative sample, as suggested in Copestake (2021), permitting cross-analysis between the two datasets, and causal maps derived from the coded interviews trace the pathways of change from the participants’ perspectives. Throughout, we report quantitative and qualitative results together, by empowerment outcome. Two features of this design discipline the disaggregated analysis. The index constructions were fixed before our endline data were analysed, following the Sahel Graduation evaluation (Bossuroy et al., 2022), so the menu of sub-components is not a researcher choice made after estimation (Kling et al., 2007; Anderson, 2008); and the qualitative evidence that directs attention to particular components was collected independently of the quantitative results. We return to what this combination implies for the use of pre-analysis plans in the conclusion (Olken, 2015).

What the combination reveals is a sharp asymmetry between the two factors of empowerment. This asymmetry runs through each of the questions this paper asks. Resources rise in all three treatment arms, by 0.52, 0.42 and 0.63 standard deviations for the female-targeted, male-targeted and female plus Umodzi arms respectively. Decision-making agency tells a different story. We find suggestive evidence that the overall decision-making index rises for the female plus Umodzi arm. This result is significant only at the 10 per cent level, and it does not hold up under multiple hypothesis testing.

On whether it matters which spouse is targeted, male or female, the aggregate index conceals more than it shows. Financial decision-making rises significantly for both female-targeted arms, while the male-targeted arm’s null index conceals offsetting movements. Women in that arm see a marginal gain in financial decisions against a significant loss of agency over productive decisions, the very activities the programme creates. Gender targeting, invisible at the index level, thus reappears beneath it. Who receives the programme shapes which domains of her agency move, and in which direction.

On whether a couples training adds to the standard Graduation bundle, the clearest gain is not in decision-making but in role sharing. Husbands’ engagement in chores within the female domain rises significantly, and robustly to multiple hypothesis testing, in the female

plus Umodzi arm alone. That gain in chore engagement comes without a significant reduction in women’s own load, and the qualitative interviews suggest these tasks are still seen as the woman’s responsibility, which may help explain why this effect fades by 17 months.

On what a mixed methods design can reveal about the pathways of change that a survey alone would miss, the qualitative evidence surfaces the pathways behind these numbers. In their interviews, women across arms describe more collaborative decision-making over income generating activities, the essential realm of the Graduation programme, more than over budgeting. The accounts place change in budgeting mainly where the couples training addressed it directly; the driver women name is the programme’s gender messaging rather than the cash itself. The share of interviewed women describing an unchanged arrangement of always consulting their husband, an answer a survey would record as joint decision-making, is far higher in the male-targeted arm than in either female-targeted arm.

Taken together, the changes that occur are real but domain-specific and early. This is the pattern Kabeer’s framework predicts. Resources move first, and they move everywhere. Agency moves only where the intervention reaches it most directly. The changes are visible beneath the indices, and in the women’s own accounts, but not at the level at which empowerment is conventionally measured.

2 Conceptual Framework

Our conceptual framework is built on the work of Kabeer (1999b). Kabeer views empowerment as a process with three inter-related factors, resources, agency and achievement. Resources are the pre-condition. Agency is the process. Achievement is the outcome. Access to material, social and human resources enables a woman to exercise choice. Agency is the ability to define goals and act on them. Achievement is the outcome of this process (Kabeer, 1999b). Each factor matters for a woman’s ability to make choices, but it is the process running through all three, not any one factor alone, that constitutes empowerment. Programmes such as Graduation, when targeted at women, are by design meant to address the first factor, resources. There is some evidence that resources feed agency. When conditional cash transfers are targeted at the woman in the household, her say in important expenditure decisions rises with her share of income (Attanasio and Lechene, 2002).

Two further features of this account matter for what follows. First, empowerment is about choices that matter. A woman having a say in areas no one contests tells us little about her power. The same indicator can be informative in one setting and empty in another. What counts is whether the choice carries consequential significance for women in that context (Kabeer, 1999a). Second, because empowerment is a process, change should not be expected to arrive everywhere at once. It should emerge unevenly, beginning in the domains an intervention most directly touches, and only later, if at all, spreading to the rest. An evaluation conducted mid-process should therefore expect movement in resources, early and partial movement in the agency domains closest to the intervention, and little or none elsewhere. Observing exactly this pattern is evidence of a process underway, not of failure.

Kabeer’s process view has direct implications for measurement. We operationalise the resource factor as women’s access to and control over financial and productive resources. This is a standardised index of seven components, covering her share of agricultural and livestock income, business ownership and control of its income, livestock ownership, wage income, and saving outside the household. We operationalise agency through three groups of outcomes (Donald et al., 2020). The first is decision-making, itself a standardised index of four domain indices covering financial, productive, household, and fertility and education decisions. The second is a two-component index covering freedom of movement and contact together with community and political involvement. The third is the sharing of household roles, measured through the male spouse’s participation in chores within the female domain (full definitions in Table A5). These sub-components do not measure the same thing three ways. They are distinct domains, and a woman’s agency can move in one and not another, or in different directions at once. A woman may gain a say over the loan she took for her business while losing ground over the livestock her husband now manages. A composite averaging the two records nothing. This is precisely why an aggregate can mask domain-specific change or net out offsetting movements, and why we analyse the components as well as the aggregates. The who-decides format of the underlying questions raises a further conceptual issue. It records the arrangement, not its meaning. Consulting a spouse may reflect joint deliberation or deference, and the same reported answer can reflect very different dynamics between spouses (Bernard et al., 2020). A final issue follows from Kabeer’s point that empowerment is about choices that matter. In a context where most women already move freely, a movement index has little room to register power, however carefully it is constructed. None of these problems can be resolved from within the survey itself. They call for a second kind of evidence, whose role we set out next.

The two kinds of evidence are complementary. The RCT delivers causal effects, but only on the outcomes as measured, and the problems described above are built into those measures. The QuIP narratives identify what the women themselves experienced as change, in which domains, and what they put it down to. The narratives have no counterfactual comparison group of their own, so differences in what arms report is suggestive rather than statistically identified; each narrative’s own causal claim relies instead on respondent attribution. Where a survey index is null, the narratives indicate whether nothing changed or whether change occurred that the index cannot see, and in which components. Where the narratives report change, the sub-indices indicate whether it survives comparison against a counterfactual. Convergence between the two, arrived at independently, is stronger evidence than either alone. Divergence is not a failure of one method. It points to a domain the survey misses, an arrangement the survey misreads, or change concentrated among some women that disappears in the average. The qualitative evidence was also collected by researchers blindfolded to treatment assignment and to the hypotheses, so each method disciplines the other without contaminating it.

The framework, together with the programme’s design, motivates three hypotheses. These follow from theory rather than from a pre-analysis plan. What was fixed in ad-

vance were the trial arms and the outcome families registered with the AEA RCT Registry. First, because all three treatment arms provide the same core package of transfers, productive support and access to savings mechanisms, we expect each Graduation treatment to increase women’s access to and control over productive and financial resources relative to the control group. Second, consistent with the conventional rationale for targeting economic programmes at women, we expect female targeting to produce greater improvements in women’s resource control and household decision-making agency than male targeting. Third, because Umodzi directly addresses communication, cooperation, decision-making, gender norms and the sharing of household responsibilities (Concern Worldwide, 2018), and the evidence from comparable trainings points to increased joint decision-making and better communication as the main mechanisms (Doyle et al., 2018), we expect adding Umodzi to female-targeted Graduation to produce greater improvements in women’s decision-making agency and role sharing than female-targeted Graduation alone. The analysis tests these hypotheses using the aggregate indices, and then examines their component measures to understand the pattern of any effects. Our design also specifies the pairwise comparisons between arms in advance, including male-targeted versus female plus Umodzi (Section 4). If empowerment is a process, change should appear first in the domains an intervention most directly touches, so partial or uneven support for a hypothesis at the aggregate level tells us where the process has and has not begun. The same logic motivates our short-run focus. Resources should move before agency, and the first follow-up at 5 months is where early, partial agency movement should be sought. The ordering itself is an assumption of the framework, not something a single follow-up can establish.

The framework leaves two questions open. The first is what male targeting does to a woman’s say over the productive activities the programme creates. Handing the programme to her husband could leave her say unchanged, or it could strengthen his position in exactly those domains, and nothing in the framework decides between the two. Nor does the finding in Bedi et al. (2026c) that household outcomes look the same whichever spouse is targeted settle it, because a household can pool its income while who decides shifts within it. That paper explains why the size of the pie does not depend on the recipient, but finds that targeting changes which spouse accumulates livestock and savings. Ours asks what the recipient’s identity does to her part in deciding over it. The second open question is persistence. Achievement, the third factor, lies beyond the horizon of this paper, and the medium-run trajectory of these outcomes is reported in Bedi et al. (2026c).

3 The Intervention

The Graduation programme is a multifaceted economic inclusion intervention combining temporary consumption support, productive capital, livelihood training and coaching, and access to savings mechanisms, designed to place ultra-poor households on a sustainable income path. The two study districts, Mangochi and Nsanje, have the highest poverty index in Malawi, and poor rural households there depend on rain-fed agriculture, casual labour

and small enterprises. Women have less access than men to productive inputs and financial services, carry most of the domestic and care work, and household finances sit under the husband’s control. The choices that carry consequential significance in this setting (Kabeer, 1999a) are therefore control over economic activities, financial decisions and the sharing of domestic responsibilities; movement is not one of them, as at baseline most women already report moving and contacting family freely. The fuller context is described in Bedi et al. (2026c).

Concern Worldwide implemented the programme, known locally as Tiwolohe. In addition to targeting households classified as poor or very poor via a community wealth ranking or a proxy-means test, Concern Worldwide’s Graduation programme consisted of four core components:

- Consumption support: Treated households received 12 monthly cash transfers of 15,000 MWK (\$57 PPP adjusted in USD 2018) to help them stabilise their consumption levels until they started earning from the income generating activities enhanced as part of the programme.
- Skills training and regular coaching: To enhance human capital, treated households received access to practical, short, hands-on trainings as well as routine coaching visits.⁵
- Savings and financial access: Treated households were encouraged to save through joining Village Saving and Loan (VSL) groups. In these groups, households saved on a regular basis for a defined period, after which they got a payout of their savings and could rejoin the group for another defined period. VSL groups aimed at building financial discipline and at familiarising households with financial service providers.
- Capital transfer: The capital transfer was created to help participants jump-start a sustainable and profitable economic activity. Households had to define a business plan, which once finalised was accompanied by a 96,000 MWK (\$368.88 PPP adjusted in USD 2018) transfer.

A third of treatment households targeted the female of the couple, and a second third targeted the male of the couple. The final third of treated households received an additional component, the gender transformative training. As gender inequality is regarded as a core driver of poverty, the couples’ economic cooperation and empowerment training was designed to address gender inequalities through a 12-module programme spread over a period of 12 months. The aim of the programme was to promote positive relationships between men and women, to prompt positive social norm changes within household and community and to ensure that women and men benefited equally from the programme and any adverse effects of the programme on women’s safety and household dynamics were minimised.

The intervention was delivered to two cohorts of households, separated by a year. The first cohort received Concern’s Graduation between September 2018–April 2020, and the second cohort received the intervention between September 2019–April 2021. Take-up of the training was high. Monitoring data show that at least one spouse attended more than 85

per cent of sessions, and both spouses attended more than 75 per cent (Bedi et al., 2026c). Umodzi added roughly 30 hours of contact with the programme over the Graduation-only arms, so our comparison of the female plus Umodzi and female-targeted arms captures the effect of the training as a whole, its content and its additional contact together. Because of the sensitivity of the subjects, sessions on power and violence were delivered separately to women and men by same-sex facilitators; the programme’s effects on intimate partner violence are analysed in Bedi et al. (2026b).

4 Research Design

Focused on a sample of low-income couples, we conduct a complementary mixed methods study to understand the impact of gender targeting and incorporating couples cooperation and gender empowerment training. Our mixed method approach consists of two main components, a quantitative randomised control trial (RCT) and a qualitative evaluation using the Qualitative Impact Protocol (QuIP). Our quantitative data was collected 5 months post programme completion, while our qualitative data was collected 6 months post intervention. The main objective of aligning both the quantitative and qualitative data collection rounds was to develop a deeper understanding of causal mechanisms connecting treatment and gender empowerment. In our results section we report on quantitative and qualitative results together by outcome variable.

The trial is registered in the AEA RCT Registry (AEARCTR-0007605), with ethical approval from the Research Ethics Committees of Trinity College Dublin and Innovations for Poverty Action. The registry specifies primary outcome families rather than item-level variable constructions; these families include empowerment and intra-household decision-making. Within those families, our measures were not constructed after seeing the data: the resource access and decision-making indices follow the constructions used in the Sahel Graduation evaluation of Bossuroy et al. (2022), and were fixed before our endline data were analysed. The mapping of measures to concepts follows Kabeer (1999b), as described in Section 2: the productive resource access factor captures the resource pre-condition of empowerment, while the decision-making, movement and community involvement, and role sharing indices capture agency. Two features of this design bear on the interpretation of our disaggregated results. First, the set of sub-indices we examine is not a researcher choice made after estimation; it is the component structure of the Bossuroy et al. (2022) indices, fixed *ex ante*. Second, the qualitative evidence that directs attention to particular components, budgeting and income generating activities, was collected by partially blindfolded researchers with no knowledge of the quantitative results.

As set out in the introduction, the mixed method design conforms to the “quant-led model” described in Copestake (2024). Selection of respondents for the QuIP study was informed by quantitative survey data (Copestake, 2021), permitting additional cross-analysis of data. Interpretation of findings was iterative through the two rounds of QuIP studies, with causal maps derived from qualitative coding used to explore possible mechanisms explaining

correlations identified between independent and dependent variables in the RCT.

4.1 Quantitative research design

The RCT involved four arms, with a control arm and three treatment arms where the Graduation programme is targeted at:

1. Women as the main recipient in the household
2. Men as the main recipient in the household
3. Women as the main recipient in the household but with an additional couples' economic cooperation and empowerment training called "Umodzi"

Each intervention was delivered to 600 households randomly selected in Mangochi and Nsanje, with 300 households in each district receiving each version of the proposed programme. In addition to this, there were 600 control households in pure control villages and an additional 900 control households in treatment villages. We implemented a village level cluster randomisation with 50 villages in each of the four experimental arms. Furthermore, each intervention was split between the two cohorts, where 300 households in each version of the programme were in cohort 1, and the remaining were in cohort 2. Control households were also split between cohorts for data collection.

Quantitative data was collected from the female spouse "before" and "after" the intervention. The two surveys were separated by 24 months, with the second survey conducted 5 months after all intervention activities finished.

In our empirical estimation, we examine the effects of gender targeting and gender training on women's empowerment. This design lets us directly compare the individual impact of each version of the Graduation programme, female-targeted, male-targeted and female plus Umodzi, compared to control households. In addition, we can look at how our effects vary when we target the Graduation to the male versus female in the house. We can also look at the additional value of layering the couples' economic cooperation and empowerment training on female targeting by comparing the female-targeted versus female plus Umodzi arm. Finally, we also can get greater insight on how involving the male spouse by either directly targeting him or by involving him in the couples' training has a differential impact on empowerment.

Our main specification (Equation 1) estimates the intent-to-treat effects using the analysis of covariance (ANCOVA) based on the following specification:

$$Y_i = \beta_1 + \beta_f T_h^f + \beta_m T_h^m + \beta_{f+} T_h^{f+} + \beta Y_{i,bl} + \delta d_{bl} + \gamma Z_{cd} + \mu_{1t} S1_{ht} + \mu_{2t} S2_{ht} + \mu_{3t} S3_{ht} + \epsilon_{iv} \quad (1)$$

where Y_i is our outcome of interest for the female spouse i , 5 months post programme. Our treatment status variables are T^f , T^m and T^{f+} , which take the value of 1 if household h in village v received the Graduation programme targeted to: the female spouse (T^f); the male spouse (T^m); and targeted to the female spouse and where the couple participates in the

couple’s empowerment training, Umodzi (T^{f+}). Coefficients β_f , β_m and β_{f+} are our main parameters of interest, as they capture treatment impact on our specified empowerment outcomes. We control for baseline values of outcome Y . For a portion of households where we are missing baseline values, we generate baseline values equal to the mean value of the outcome from the unit of randomisation (village) which we use as the replacement value, and also control for this by including a dummy that is equal to 1 if we are missing the baseline outcome (d_{bl}). We control for the district and cohort, Z_{cd} , the individual belongs to. $S1_{ht}$, $S2_{ht}$ and $S3_{ht}$ identify spillover status. Each equals 1 for a household that was eligible and included in the sample, but that the random allocation did not assign to treatment. In addition, ϵ_{iv} is our statistical error term, where the errors are clustered at the village level, which is our level of randomisation.

We also test and report on the equality of treatment effects between female-targeted and male-targeted Graduation, the added value of the couples’ training over the female-targeted Graduation, and between targeting the Graduation at the male spouse versus targeting the Graduation at the female spouse but involving the male spouse through the couples’ economic cooperation and empowerment training. For robustness, we correct our four headline indices for multiple hypothesis testing and report whether our results hold under this correction.

4.2 QuIP research design

In parallel, we conducted qualitative interviews using the Qualitative Impact Protocol (QuIP). The QuIP is a method for mapping causal drivers of change from the perspective of programme participants, where the qualitative researchers are partially “blindfolded” to reduce pro-project and confirmation bias. QuIP studies are designed to collect credible evidence on the household-level impacts of an intervention. This information is gathered directly from participants about their perceptions of what has changed in their lives over a set period of time and across a series of domains related to the interventions’ theory of change.

We conducted two rounds of QuIP, with a total of 96 semi-structured interviews and 16 focus group discussions 6 months after the completion of programme activities in order to triangulate the findings, provide evidence on causal mechanisms, and focus on specific research questions regarding empowerment.⁶ The semi-structured interviews were conducted by a team of local researchers who are fluent in local dialects. The researchers were partially ‘blindfolded’ in that while they knew about the Graduation programme, they had no knowledge of the targeting and empowerment training, the hypotheses being tested or what any of the interviewees had been exposed to. They worked completely independently of the Graduation programme team and Concern Worldwide, and did not reference Concern Worldwide or the programme in any of the interviews. These procedures are important as they help to reduce pro-project and confirmation bias and complement the open-ended nature of the questionnaire, which is consciously designed to increase the potential to uncover unintended intervention outcomes or unexpected drivers of change.⁷

Respondents were asked to reflect on changes over the last two years, given that the pro-

programme started in the prior two years for each cohort interviewed. In the semi-structured questions, women were asked about changes they experienced in the last two years related to household decision-making, women’s ownership of assets, her control over her resources, sharing of roles and responsibilities and her relationships, among other domains. Questions were purposively designed to be broad and open-ended to allow the respondents to speak freely about what they believed to be significant changes in relation to those outcome domains. Researchers were trained to use additional questions to probe further and establish what the perceived drivers of these changes were. Closed questions were also used at the end of each domain to capture overall perceptions of change in some specific areas.

The standard QuIP coding process involves allocating influence (driver) and consequence (outcome) factor labels to causal claims found in the respondents’ narratives. These factors are unique to each project and developed iteratively by the analyst based on what the respondents have reported. The factor labels are created based on what respondents say (unprompted) in the interviews, and re-used when narratives refer to a similar description; they are not suggested as options for the respondents to consider. Once the data was coded, the analyst looked for patterns and trends across the responses to understand which stories of change were frequently reported, and whether there were any differences between groups of respondents. For this paper, the analyst merged the data and factor labels from both QuIP rounds and focused on stories of change relating to empowerment outcomes, and differences between treatment arms.

The qualitative results allow us to trace and visualize the causal pathways reported by respondents to generate a high-level overview of changes linked to the interventions provided by Concern Worldwide, including the cash transfer and different forms of training in the form of a causal map. The causal maps are generated by presenting the number of times a link between two factors (influence and consequence) was cited by different respondents. The number in the box/label denotes the number of individuals who cited that factor (regardless of driver), whereas the number above the link denotes the number of individuals who made the link between two factors.⁸

While the causal maps in this analysis aggregate data across all treatment groups, we also use a ‘Surprise’ algorithm to check whether any specific treatment arm is identified more strongly as a driver of reported changes. Where arm comparisons are reported we provide the underlying counts; given the purposive, small and unequal arm samples (36, 24 and 36 respondents), chi-squared tests of independence are used only as descriptive diagnostics, are treated with caution given multiple comparisons, and are not interpreted as population treatment effects.

The sample size for the qualitative research is purposively selected to ensure confidence of thematic saturation within each group, but is not statistically representative of the wider population. Counts have been used to highlight trends in the data, but these should not be interpreted as being representative of a particular population other than those involved in the study. The programme-neutral approach to collecting stories of change in the in-depth interviews offers an opportunity to see whether similar drivers and outcomes considered in

the survey are brought up without prompting, and to compare these findings between groups as well as with the quantitative survey.

5 Data Selection, Description and Summary

5.1 RCT Sample Selection

Through the use of a poverty index, our implementing partner, Concern Worldwide, identified the two districts with the highest poverty score, Mangochi and Nsanje, for implementing the Graduation programme in. Within this, they selected sub-districts (VDCs) that they would implement the programme in. We identified 200 villages, 100 per district, based on each village having at least 20 eligible households in the village. We then randomised across villages, assigning 50 villages as pure control villages and the remaining as treatment villages. After our baseline, we did our next round of randomisation, which involved 3 steps. First, we randomised the treatment villages into the three treatment arms. Second, in treatment villages, we randomised 12 households to receive treatment, and 6 of the remaining eligible households were randomly selected as control households. Third, we randomised treatment and control villages into cohort 1 and cohort 2, with the first treatment cohort starting with the programme in September 2018, and cohort 2 starting in September 2019.

In order to be eligible, the households had to meet the following criteria:

- The household had to be classified as poor through a wealth ranking or a proxy means test.⁹
- The household head is married and his spouse lives within the household
- Both head and spouse are willing and able to participate in an Income Generating Activity.
- Both head and spouse reside at the household at least 60% of the year.

We find balance across baseline variables and treatment arms for our analysis sample of female spouses interviewed at baseline and follow-up, with no variable having a normalised difference above the rule of thumb value of 0.25 (Imbens and Rubin, 2015). This is in line with the balance results for the full trial sample in the companion paper (Bedi et al., 2026c). Balance results are shown in Table A1.

We find no differential attrition between control and any of the treatment arms, nor between treatment arms, in the sub-sample of households with the original female spouse living with her partner, the sample used for our quantitative analysis (Table A2). Given that divorce, re-partnering and migration could themselves be affected by the intervention, we also check that this holds for the broader household panel. We find no significant differences here either. This is in line with the companion paper, which finds no significant differences in relationship status across treatment arms at 5 months post intervention, the horizon of this paper (Bedi et al., 2026c).

5.2 Qualitative Sample Selection

Female spouses were drawn from each of the treatment arms, female, male and female plus Umodzi. A total of six households were interviewed from each treatment village that was included in the sample. As each treatment village had 12 households participating in Concern Worldwide’s Graduation programme, these 12 households were randomly ranked from 1-12. The local research team invited households ranked 1-6 to be interviewed for the QuIP. If the researchers could not reach a particular household in the interview time frame, they approached the next household on the list.

In each round of the QuIP, a total of 48 individual household interviews were conducted and 8 focus group discussions (FGDs), split equally between Nsanje and Mangochi districts.¹⁰ The QuIP sample for 2020 was chosen from cohort 1 of the RCT sample, while the 2021 sample was chosen from cohort 2. The QuIP data collection occurred in October 2020 and November 2021, following on directly from the quantitative data collection.

5.3 Data Description

At baseline, for our quantitative component we had a sample of 3,313 households, where the average household had 5.5 members, and the average age of the female spouse was 35.4 years old, while the male spouse was 42.5 years old. While men had on average 3.75 years of education, women only had 1.95 years of education. Overall, the sample of households had good access to borehole water, had quick access to a weekly market but had to walk for almost 4 hours to the closest public health clinic (Table A1).

For the analysis in this paper, our sample includes a panel of 2,934 female spouses in partnered relationships, successfully interviewed at baseline and follow-up 5 months post programme. Of this sample, over 91.5 per cent of households are still in the same original couple configuration as at baseline.¹¹ The remaining 8.49 per cent of households, either have a new male spouse (2.79%), or the male spouse has migrated (5.69%).¹²

Our sample is composed of households with the original female spouse, who remains partnered. We have 559 control households, 524 female-targeted, 534 male-targeted and 539 female plus Umodzi targeted households. The sample also includes 778 control households in treated villages, our spillover households.¹³

As the QuIP is designed to capture change from the respondents’ perspective and our QuIP sample is selected from our quantitative sample of households, we explore how aligned our two datasets are on defined concepts such as income and savings. In the QuIP interviews, female spouses were asked about changes to their household income and savings over the last two years. In the quantitative survey we measure changes in their household and personal income and savings over time. We run simple linear regressions where we regress our quantitative outcome measure on the related qualitative concepts. Our equation for this regression is: $Y_{1,i} = \beta_1 + \beta_2 QuIP_{1,i}^j + \beta_3 Y_{0,i} + \alpha_1 X_{1,ih} + \gamma Z_{1,cd} + \epsilon_{iv}$ Our outcome variable of interest is quantitative variable Y for individual i , post intervention in period 1. The variable $QuIP$ captures QuIP outcome j , and takes the value of 1 if individual i expressed a change

in the specific concept j in their semi-structured QuIP interview, post intervention and post period 1. As we control for the quantitative outcome at baseline ($Y_{0,i}$), we are in essence also looking at whether a respondent experiencing a change in their qualitative outcome is correlated to a change in their quantitative outcome. We include individual and household controls, including respondent's age, if they are literate, and household size in $X_{1,i}$. We also include district and cohort controls, Z_{cd} , as our sample was stratified by district and the cohorts were randomised. ϵ_{iv} is our statistical error term and errors are clustered at the level of cluster randomisation, the village unit. This helps us understand how changes identified through the QuIP interviews, such as an increase in income and savings, correlate with changes in our quantitative measures of income and savings.

Table 1: QuIP Outcomes & Quantitative Outcomes, Level (controlling for BL)

	Quantitative Outcomes					
	Total Hh Income	Total Female Income	Has VSL Group	Current Ttl Savings	Household has Business	Female has Business
Increase Hh Income	332.73***	141.39*	0.33***	33.09**	0.17	0.29**
Increased Savings	228.45**	116.14**	0.46***	36.15***	0.10	0.19
Part of VSL	107.91	38.39	0.16	36.60*	0.12*	0.17
Has New Business	82.80	-52.71	0.14**	23.14	0.09	0.11

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$

Monetary amounts are PPP-adjusted, CPI deflated, 2018 USD prices. 2018, 1 USD = 260.245 MWK PPP.

What we find in Table 1 is a positive correlation between households reporting increased income in their QuIP responses and an increase in household income over time from the quantitative data. We also find similar positive correlations for savings. This suggests that, at a minimum, the two methods are capturing similar concepts, when it comes to defined outcomes such as income and savings.

6 Results

Our analysis structure draws on Kabeer (1999b), where empowerment is viewed as a process involving three factors; resources, agency and achievement. In our results, we focus on the first two factors. We find evidence from our quantitative results that the programme increased women's resource factor, especially for women in the female plus Umodzi arm. We find that the Graduation programme helped women in treated households increase her resource factor, the pre-condition for empowerment. Our qualitative results suggest that the main drivers of increased resource factor is Concern Worldwide's cash and capital transfers, as well as their business training and gender messaging. We find suggestive evidence that this increased overall decision-making agency for women in the female plus Umodzi households, compared to women in the control arm. Both our quantitative and qualitative results point to decreased or no change in decision-making agency for women in the male-targeted group. We also find hints of positive change in relation to role sharing from both our quantitative and qualitative results. We start the analysis by looking at the results on women's resource factor, and then move to combined results on agency. Within our quantitative findings, we

first discuss the results between the control and treatment arms and then move to within treatment arm differences.

6.1 Resource Factor

We create a resource factor index, composed of women’s access to and control of financial and productive resources, as detailed in Table A5. We find that all women in treatment households saw an increase in her access to and control of financial and productive resources, compared to female spouses in the control arm (Table 2).

Table 2: Impact on Resource Factor Index, Female Spouse

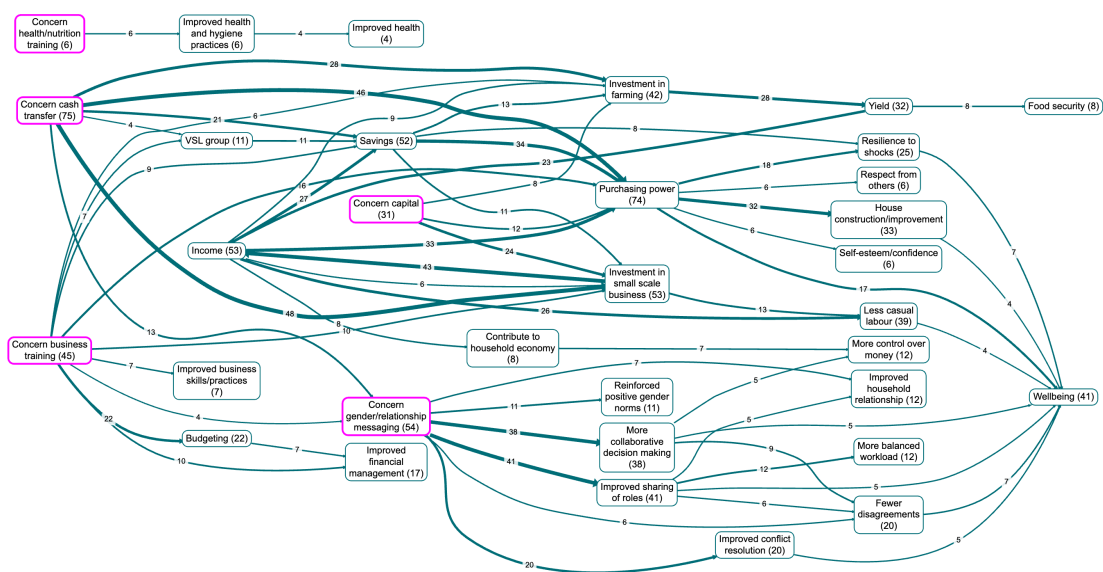
	(1) Resource Index
Female targeted	0.52*** (0.08)
Male targeted	0.42*** (0.07)
Female plus Umodzi	0.63*** (0.07)
Observations	2928
Control Mean	0.02
Control SE	1.01
p(T1=T2)	0.19
p(T1=T3)	0.15
p(T2=T3)	0.00

Standard Errors in parentheses, clustered at the village level

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$

This result suggests that women across all treatment arms strengthened their ability to engage in economic activities, including running a business, owning livestock and engaging in waged activities, and reap gains through income from these activities going to her, as well as increase her likelihood to save in channels outside of the household. The resource index increased by 0.52 standard deviations (SD) for women in the female-targeted arm, 0.42 for women in the male-targeted arm and 0.63 SD for women in the female plus Umodzi households, compared to control households, 5 months post intervention. As this index is standardised with mean zero and standard deviation 1, this equates to an increase of 52, 42 and 63 per cent of the standard deviation of the resource index, compared to control households. This result is critical, as resource factor is the pre-condition for women to get agency, which is seen as the process of empowerment. This suggests that Graduation creates, at minimum, the setting in which agency may be more likely to be realised. When we look at within treatment effects, we find that women in the female plus Umodzi arm made the largest gains. We also find that the difference in their gains compared to women in the

Figure 1: Links made to Concern Worldwide interventions in qualitative interviews, out of 96 total



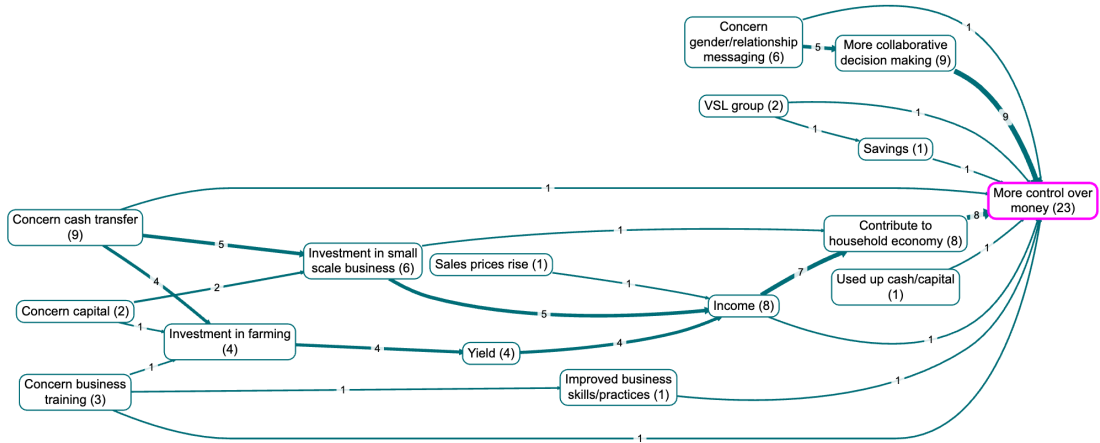
male-targeted arms is statistically significant. All these results hold when we correct for multiple hypothesis testing across our four headline indices (Table A4).

As our qualitative research was designed to let participants self-identify change, we don't have a direct comparison to all components of the quantitative index of women's access to and control over financial and productive resources. Figure 1, which presents a high-level overview of all reported changes from our qualitative analysis linked to the various components of Concern Worldwide's Graduation programme, has some evidence of a resulting impact on women's resource factor in the form of greater control over money. Respondents noted that they felt they had more of a say in financial matters because they contributed to the household economy. Through earning income, being "financially independent" or having their "own means", they had a greater say in how money was spent (IMBT-2, IMBT-3).¹⁴

Figure 2 shows all reported drivers of a change in control over money by women interviewed, with 20 women linking this to increased resources from the Concern Worldwide cash transfer, business training, capital and gender messaging. While women across all treatment groups mentioned increased control of money as an impact, we find that the primary drivers leading to this impact varied by treatment arms.

A disaggregated analysis of Figure 1 is presented in Figure 3, providing a summary of outcomes linked to the main drivers of change in the qualitative results, split by treatment group. Figure 3 highlights that the drivers behind increased control of money differ by treatment arm. In particular, women in the male-targeted arm identified Concern Worldwide's cash transfer as the driver behind their self-reported impact of increased control of money, whereas women in both female-targeted arms also mentioned Concern Worldwide's gender messaging as the main driver behind their increased control of money. This difference in drivers gives some insight into the mechanism behind the changes experienced in women's resource factor, as well as insight into different priorities by treatment arm, and how these

Figure 2: Drivers of more control over money



decisions had different impacts on the women’s lives.

In our qualitative analysis we also find that some of the unprompted references made by the women interviewed can act as proxies for resource factor, for example how the money received in the cash transfer was used. While it is not clear from the self-reported impacts who in the household had control over or used the money received in the cash transfer,¹⁵ patterns across the treatment groups are indicative of different priorities depending on who received the cash. While women in the male-targeted arm identified increased farming investment driven by the cash transfer, women in the female plus Umodzi group made strong references to using the cash to invest in small-scale business. Reports that income from small-scale businesses is used to contribute to savings, to reduce the need to take on casual labour and to purchasing items is accordingly higher amongst the female plus Umodzi group (clustered under a top level factor label called ‘Purchasing power’ in Figure 1). The link between the cash transfer and savings is also stronger in the female plus Umodzi group, echoing the findings in the quantitative data which showed increased control of savings and credit.

Another way we gain insight into women’s resource factor is how they exercised their purchasing power. Within the label ‘Purchasing power’ we are able to identify whether households purchased items classified under Assets, Household needs and Household goods. Figure 4, shows a somewhat different priority between treatment arms in how they used their purchasing power. Women in the female plus Umodzi were the most likely to report purchasing assets, including livestock, while women in the female-targeted were least likely to report this. Assets like livestock in rural settings reflects an important income diversification strategy, as well as a way to cope with shocks (Bedi et al., 2026a).

“This money boosted my income and I managed to buy goats, chickens, a bicycle and iron sheets. I now live in a better house and the livestock provide a steady source of income in times of shocks. As a woman, I own blankets and clothes but have recently acquired livestock and iron sheets due to the cash transfer from Concern Worldwide”

Figure 3: Proportion of respondents who cited outcomes linked to Concern Worldwide, where over 10 per cent

Starting driver of change	Reported outcomes					
	Female targeted		Male targeted		Female Plus Umodzi	
Concern cash transfer (excluding increased purchasing power)	Investment in small scale business	44%	Savings	63%	Investment in small scale business	69%
	Investment in farming	36%	Investment in farming	58%	Savings	58%
	Savings	33%	Investment in small scale business	42%	Less casual labour	53%
	Wellbeing	33%	Less casual labour	42%	House construction/improvement	42%
	Better yields	31%	House construction/improvement	42%	Wellbeing	42%
	Less casual labour	28%	Wellbeing	38%	Resilience to shocks	33%
	House construction/improvement	28%	Better yields	33%	Investment in farming	31%
	More collaborative decision making	25%	Resilience to shocks	29%	Better yields	17%
	Improved sharing of roles	19%	Respect from others	17%	More collaborative decision making	14%
	Resilience to shocks	14%	More control over money	13%	Improved household relationship	11%
	Improved household relationship	14%	Better self-esteem/confidence	13%	Aspirations to grow business	11%
	Fewer disagreements	14%	Improved food security	13%	VSL group	11%
	More control over money	11%	Contribute to household economy	13%		
Concern gender messaging	Better self-esteem/confidence	11%				
	Improved food security	11%				
	More collaborative decision making	36%	Improved sharing of roles	38%	Improved sharing of roles	58%
	Improved sharing of roles	31%	Improved conflict resolution	29%	More collaborative decision making	56%
	Fewer disagreements	22%	Reinforced positive gender norms	25%	Wellbeing	39%
	Wellbeing	14%	More balanced workload	25%	Improved conflict resolution	25%
	Improved household relationships	14%	More collaborative decision making	21%	Fewer disagreements	25%
	More control over money	14%	Improved household relationships	21%	Improved household relationships	14%
	Improved conflict resolution	11%	Wellbeing	17%	Reinforced positive gender norms	11%
	Reinforced positive gender norms	11%			More control over money	11%
	More balanced workload	11%				

Figure 4: Use of cash transfer by treatment group, from QuIP data, N = 96

Purchased more with cash transfer:	Female targeted	Male targeted	Female plus Umodzi
Assets: (including livestock, construction materials, bicycle, sewing machine, land)	67%	83%	92%
Household needs: (including food, clothes, school materials, hygiene items, medical expenses)	72%	67%	72%
Household goods: (Including kitchen items, blankets, furniture)	56%	67%	69%

(IFAY-2)

Overall, we find that the quantitative analysis shows strong evidence of Concern Worldwide's Graduation leading to increased resource factor for women across all treatment arms, highest for women in the female plus Umodzi arm. The qualitative results echo this finding, demonstrating the channels through which this is being achieved. Women in the female plus Umodzi arm were more likely to report that Concern Worldwide's cash transfers were used to invest in businesses and assets, leading to increased savings and purchasing power for the household, and in some cases increased personal control over money.

6.2 Agency Factor

While we find evidence for increased access to and control of resources from our quantitative analysis, we next consider whether this translates into women having greater agency within the household. To capture different dimensions of agency, we create three different groups of outcomes from our quantitative data; decision-making agency, women’s freedom of movement and involvement agency, and finally sharing of roles.¹⁶.

6.2.1 Decision-Making

We first focus on decision-making, to understand whether increased access to and control of resources is changing women’s voice in household decisions. We create an overall measure of decision-making, which is a standardised index of four indices that measure women’s decision-making related to financial, productive, household, and fertility and education decisions. We find suggestive evidence that women in the female plus Umodzi group have higher decision-making say in their households compared to women in the control group. This result is only significant at the 10 per cent level and does not hold up under multiple hypothesis testing. We find no differences for women in the male and female treatment arms on decision-making agency compared to the control group (Table 3). In addition, we find that women in the female plus Umodzi arm have greater decision-making agency than women in the male-targeted arm, which is robust to multiple hypothesis testing.

Table 3: Impact on Decision Agency, Female Spouse

	(1) Decisions Agency	(2) Financial Decisions	(3) Productive Decisions	(4) Household Exp Decisions	(5) Educ and Fert Decisions
Female targeted	0.06 (0.08)	0.14** (0.07)	-0.03 (0.08)	0.02 (0.09)	0.03 (0.07)
Male targeted	-0.06 (0.08)	0.13* (0.08)	-0.17** (0.08)	-0.07 (0.08)	-0.07 (0.08)
Female plus Umodzi	0.14* (0.07)	0.29*** (0.08)	-0.01 (0.07)	0.08 (0.08)	0.04 (0.07)
Observations	2928	2928	2926	2926	2918
Control Mean	0.00	-0.00	0.01	0.01	-0.00
Control SE	0.98	1.00	0.99	1.00	1.00
p(T1=T2)	0.19	0.95	0.10	0.32	0.19
p(T1=T3)	0.31	0.04	0.77	0.53	0.96
p(T2=T3)	0.02	0.06	0.04	0.08	0.14

Standard Errors in parentheses, clustered at the village level

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$

When we analyse by sub-indices, we find a slightly more nuanced picture emerging on financial and productive decisions. All women in treatment arms see greater agency in financial decisions than women in the control arm. In addition, women in the female plus

Umodzi arm have greater gains on this dimension than women in the male treatment arm.

In regard to productive agency, we find that female spouses in the male-targeted arm see lower agency related to productive decisions compared to women in the control arm. This suggests that directly targeting men with an economic inclusion programme, versus not targeting the household at all, may lower the women’s decision-making agency on economic inclusion activities. Again, we find that women in the female plus Umodzi arm have greater agency in productive decisions than women in the male-targeted arm.

We also find that women in the male-targeted arm had lower levels of household decision-making agency than women in the female plus Umodzi households. Overall, these findings on women in the male treatment arm having lower overall agency than women in the female plus Umodzi treatment arm (which is driven by three of the four decision-making indices) gives some insight into how best to engage the male spouse in economic inclusion activities that are also trying to increase women’s agency. Directly targeting the female spouse with the economic inclusion programme but having the husband involved through the gender training helps to maintain her say in financial, productive and household decisions, versus directly targeting him with the programme.

From our analysis of the qualitative data, we find that many women who received the cash transfer in both female-targeted groups felt more involved (to varying degrees) in household decisions, particularly around income generation and budgeting. Whilst the results do not tell us who made the decision about how to spend the money when investing in new assets, we can see some changes in agency in the map in Figure 1 - notably linked to the driver ‘Concern Worldwide gender/relationship messaging’. Gender messaging from both the Graduation programme and Umodzi were the main factors reported by respondents to be driving positive changes in decision-making in the home.

“Previously, it was my husband who was chiefly responsible for determining how to spend money but now he involves me because of the training on gender that Tiwoloke provided us with.” (IFNY-3)

There are references to gender messaging in Figure 5 as a driver across all treatment groups. This likely reflects gender-related content reaching all arms through the programme itself, via caseworkers who work across different treatment villages, or the business training where households were told that these new business activities were for the benefit of households, not just the individual. We find that few links are made directly to the cash transfer as the driver of increased agency and there is no other driver of these empowerment-related outcomes, as captured in Figure 5, which focuses in on the outcomes linked directly to Concern Worldwide’s gender messaging.¹⁷

Some respondents noted that they felt they had more of a say in financial matters because they contributed to the household economy. Women feeling more involved in decisions related to income generation may be explained by the positive quantitative impacts on productive agency, which measures a woman’s ability to work and save outside the house and the different sources of productive income she owns and receives.

Figure 5: Pathways of Change, Graduation Impact on Empowerment, QuIP

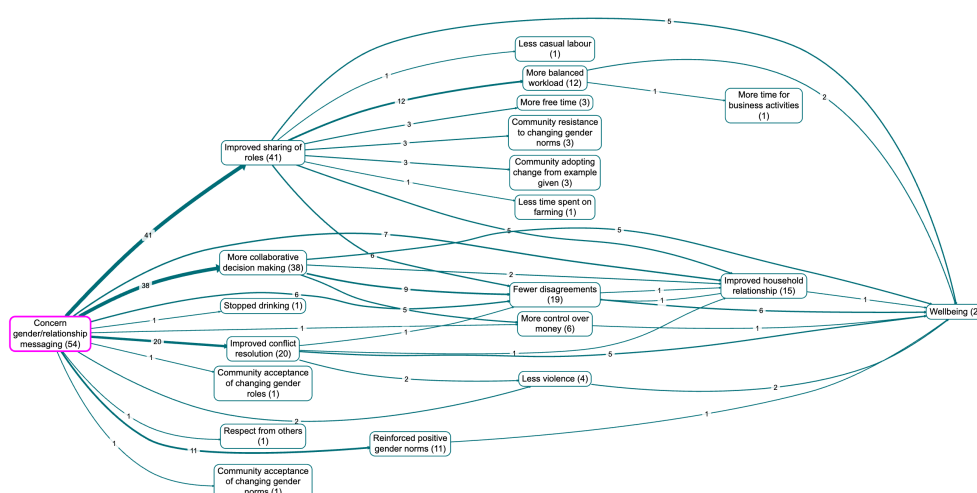
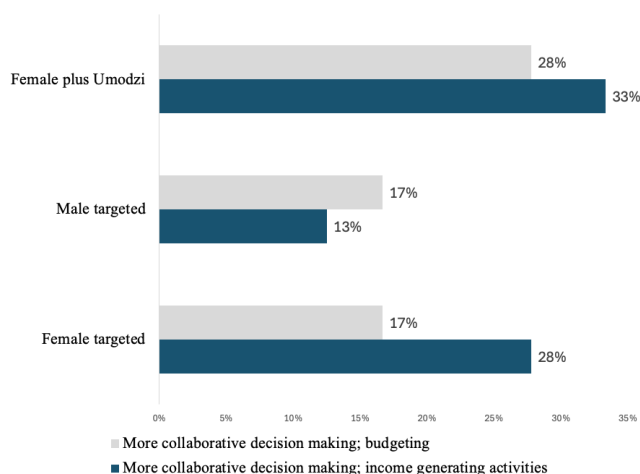


Figure 6: Changes in decision-making, QuIP by treatment arm



“I am able to generate income and contribute to the household budget. That means I am empowered and my decisions are taken on board. I am part of what we do and what we plan and this makes me feel that I am in control.” (IMAU-25)

While this impact was reported across all treatment arms, a higher proportion of Umodzi respondents overall reported becoming more involved, to some extent, in decision-making in the home (Figure 6). The difference between female-targeted versus female plus Umodzi treatment types was strongest for budgeting decisions.

“With Tiwoloke, we have learnt to consult each other [...] in making these decisions [...] this is a good change because tensions have reduced between us, and we are easily cooperating now. Tiwoloke has levelled the playing field because at first it was my husband who was mostly in control of so many things in the home including finances. I feel happy and empowered with the control I have on spending money.” (INAU-25)

Figure 7: No change, always consult husband on decisions, QuIP by treatment arm

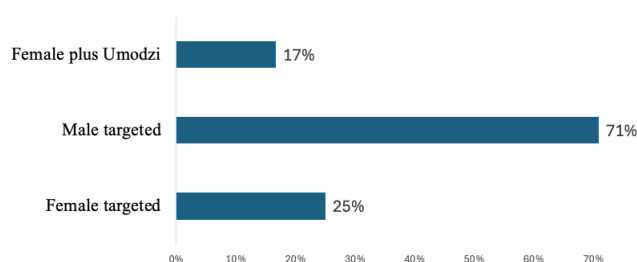


Figure 7 shows where women interviewed in the QuIP studies reported that there was no change in who made decisions, and specifically that they always consult their husband in any decision-making. The high proportion of women in the male-targeted households reporting no change is in marked contrast to the lower reports of no change in the female-targeted and Umodzi households. In a QuIP study, such strong evidence of ‘no change’ in a group which has received some sort of treatment would be considered a negative finding and is in line with quantitative results showing lower agency amongst women in the male-targeted arm related to productive decisions compared to women in the control arm. Taken together, targeting the male spouse increases women’s access to resources, while her null aggregate decision-making index conceals offsetting movements beneath it: a marginal gain in financial decisions alongside a significant loss of agency over productive decisions, the very activities the programme creates. Read at the index level, male targeting looks like no impact; read at the component level, it redistributes her agency across domains. If gender empowerment is a key objective of development, these findings highlight the importance of who in the household is targeted.

6.2.2 Freedom of Movement and Community Involvement

While decision-making captures one element of agency, we are also interested in whether we can find evidence in women’s behaviour that signals greater agency. For this we create a two-component index to capture her freedom of movement and ability to contact family and friends, as well as her ability to participate in and lead community and political activities. Our overall index while positive, is not significant (Table 4). We also find no significant differences between treatment arms. Again, when we analyse by sub-indices, we find that the intervention does have impact on her political and community agency. When women are directly targeted, they are more likely to have increased involvement in community and political activities, compared to women in households that do not receive the intervention. We find no significant effect for women in male-targeted households, compared to the control arm.

For our measure of freedom of movement and contact agency, we find no impact across all treatment arms. When we look at the individual components we see at baseline in four of the six indicators in this index, most women already report high levels of agency. Most women (94-97 per cent, Table A3) report that their spouse and/or his family have not prevented

Table 4: Impact on Movement and Involvement Agency, Female Spouse

	(1) Agency Evidence	(2) Movement Agency	(3) Community Agency
Female targeted	0.10 (0.07)	-0.00 (0.08)	0.14** (0.07)
Male targeted	0.04 (0.07)	-0.05 (0.08)	0.10 (0.06)
Female plus Umodzi	0.08 (0.07)	-0.04 (0.07)	0.16** (0.06)
Observations	2928	2927	2928
Control Mean	0.02	0.01	0.01
Control SE	0.97	0.95	1.00
p(T1=T2)	0.39	0.60	0.46
p(T1=T3)	0.78	0.65	0.81
p(T2=T3)	0.54	0.93	0.30

Standard Errors in parentheses, clustered at the village level

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$

them nor tried to prevent them from seeing/contacting their friends and family, nor from working outside the household. This is the ceiling anticipated in Section 2: where most women already move and contact family freely, a movement index has little room to register change, so the null is as much a property of the measure in this context as of the programme. We have no questions relating to freedom of movement or involvement in community and political activities in the QuIP interviews and no women gave any unprompted impacts in these areas.

6.2.3 Sharing of Roles Within the Household

Our final measure of agency is designed to capture whether there is any shift in sharing roles within the household. We look at this indicator as evidence of whether there is a change in the distribution of household chores, which often fall on the female. Our main measure is her reporting of her spouse's participation in chores that are deemed to be within her domain, including cooking, cleaning, collecting water and firewood and looking after children and elders. If we see greater involvement of the male spouse in these chores, we take this as evidence of more egalitarian role sharing in the household and therefore agency for the woman. We cross-check this against her own chore load, through two separate measures, to see whether any increase in his participation comes with a change in her own burden.

We find that women in the female plus Umodzi arm report higher engagement by their spouse in chores that are within her domain, compared to women in the control, targeted and male-targeted arms (Table 5). We find that the chore sharing index for the male spouse increased by 0.23 standard deviations (SD) reported by the female spouse in female plus

Table 5: Impact on women’s reporting of their spouses’ and their own chore load

	(1) Chore Spouse Load Index	(2) Chore Own Load Index	(3) Chores Last 24hr Load Index
Female targeted	0.06 (0.07)	-0.04 (0.07)	0.09 (0.07)
Male targeted	0.04 (0.07)	-0.09 (0.06)	0.13* (0.07)
Female plus Umodzi	0.23*** (0.08)	-0.10 (0.07)	-0.01 (0.07)
Observations	2926	2926	2928
Control Mean	0.01	0.01	0.03
Control SE	0.99	0.95	0.98
p(T1=T2)	0.75	0.49	0.55
p(T1=T3)	0.06	0.48	0.24
p(T2=T3)	0.03	0.89	0.09

Standard Errors in parentheses, clustered at the village level

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$

Umodzi households, compared to control households, 5 months post intervention. This change equates to a 23 per cent increase of the standard deviation of the chore equality index, compared to control households. The results comparing the female plus Umodzi versus control and the male-targeted arm are robust to multiple hypothesis testing.

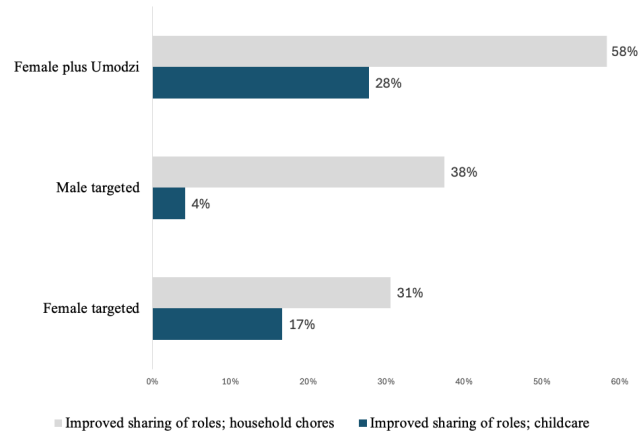
As a cross-check, we find that women in the female plus Umodzi arm have a chore load 0.10 SD lower and see no shift in their 24 hour load, though neither difference is significant. We also note that women in the male-targeted arm report an increase in chores done in the last 24 hours, and this is marginally higher than that reported by women in the female plus Umodzi households.¹⁸

Female spouses in Umodzi households also identified improved role sharing in households in the qualitative interviews. This is illustrated in Figure 8 which focuses on mentions of household chores (including cooking, cleaning, construction, gardening) and childcare by the women interviewed. The interviews also suggest that changes are most pronounced in but not limited to Umodzi households. While the quantitative results did not find this result from female spouses in the other treatment arms, it did find some evidence of female spouses in the female-targeted arm reporting greater involvement of their male spouses in these chores.

“Roles that were exclusively known to be for women in the household are also being performed by men and vice versa” (IMBT-1)

Despite these reported changes in involvement, the qualitative interviews revealed that

Figure 8: Changes in role sharing, QuIP by treatment arm



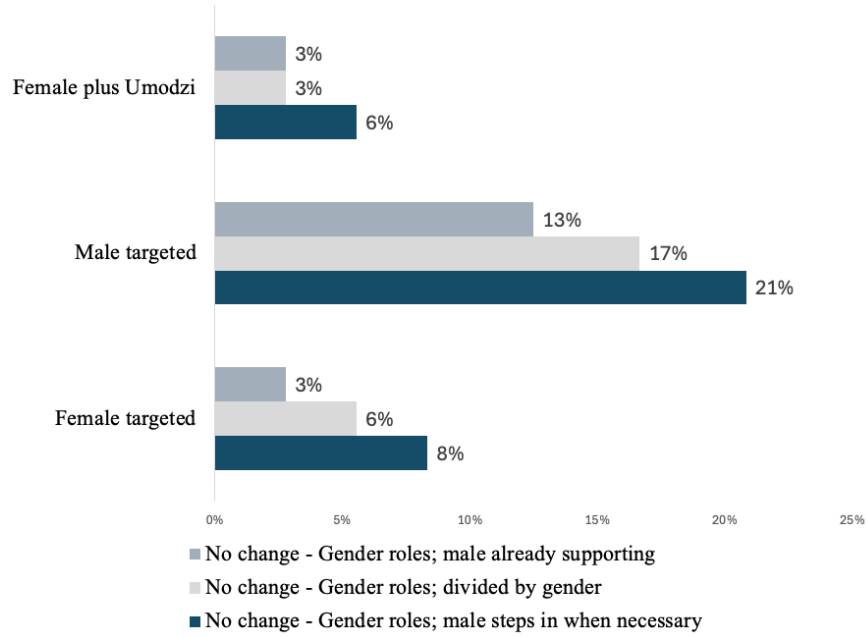
these activities are still largely perceived to be within the female domain, and it remains the norm for women to hold the primary responsibility for certain domestic jobs such as cooking, cleaning, and childcare. Whilst husbands would now “step in” to help (IMAU-5), this was often only if the woman was away, busy, or unwell.

This wife-reported difference does not persist: at 17 months, women in Umodzi households perceive no different levels of support from their husbands than women in the other treatment and control arms, though husbands’ own reports of engagement in these chores remain higher in both female-targeted arms, a result that holds over time (Bedi et al., 2026c). The QuIP interviews suggest why the change in wives’ perceptions does not stick. These tasks are still perceived as the woman’s responsibility, and husbands “step in” when she is away, busy or unwell: increased participation is experienced as help with her tasks, not as a transfer of responsibility. Since we did not interview men individually in the QuIP, we cannot examine the husbands’ perspective directly.

Overall, these combined findings suggest some level of change within the household on involvement in household chores within the female domain especially when the female spouse is targeted with an economic inclusion programme. It is perhaps therefore not surprising to see in Figure 9 that from the qualitative findings, the highest rate of reports of ‘no change’ in relation to role sharing was in male-targeted households. This chart presents the count of when female spouses reported that the dynamic over who does what in the household had not changed, and how they described the status quo. We can see that 51 per cent of women in the male-targeted arm say there is no change in how roles in the household are shared. For these women from the male-targeted arm, 21 per cent said that the husband has always stepped in to help when needed, 17 per cent said that roles continued to be divided by gender and the final 13 per cent said that their spouses had always been supporting on chores in the household.

An advantage of the causal qualitative coding used with the in-depth interviews is that it gives us greater insight into the mechanisms through which change happens from the participants’ perspectives. If we refer back to Figure 5, we see that as a result of improved role sharing, the female spouse said that they had a more balanced (less burdened) workload.

Figure 9: Where no change reported in role sharing, QuIP by treatment arm



Our quantitative results suggest that this more balanced workload is being driven by men increasing their chore responsibilities, versus women doing less. Nonetheless, in the qualitative results, women’s reporting of the results of husbands’ increased participation suggests that this is leading to important changes within the household, including improved household relationships and fewer disagreements. Twenty respondents from all three treatment arms reported improved conflict resolution with their spouse over the last two years, attributed to Concern Worldwide’s gender training. Bedi et al. (2026c) finds no differential effect of Umodzi on aggregate relationship quality measures, so these accounts are an example of the divergence described in Section 2: change that some women describe as meaningful but that a broad survey measure does not detect on average.

This was most evident in female-targeted and Umodzi households, with fewer reporting this in male-targeted households. Women shared that issues were addressed “through dialogue” (INAT-18) and “apologising to each other” (INAU-25).

Some women mentioned that because of these changes they were experiencing more harmony and unity in their marriage, and a small number spontaneously referred to changes in violence. We draw no conclusions from these accounts: violence in this trial is measured and analysed in the IPV paper, which finds no short-run increase under any treatment arm and no significant short-run differences across arms (Bedi et al., 2026b). The focus group discussions confirmed that the way partners resolve differences had changed because of Concern Worldwide’s gender training.

“Relationships have changed for the better. We solve our conflicts within the household through contact and dialogue. We do not fight. In the past my husband would slap me when offended. He stopped doing that. Tiwoloke encouraged us to stand up against

abusive partners by reporting gender-based violence to authorities. Men are now scared of beating their partners and marriages are more peaceful than before.” (INBT-23)

7 Conclusion

This paper presents findings on women’s empowerment from a mixed method evaluation of a gender focused, multifaceted anti-poverty, “Graduation”, programme, examining the impact of gender targeting and the additive effect of a couples cooperation and empowerment training. Women’s access to and control of resources, the pre-condition for empowerment, rises in all three treatment arms, with the largest gains where the couples training is added. The picture for agency is different. The overall decision-making index moves only for the female plus Umodzi arm, and only at the 10 per cent level, but beneath that aggregate financial decision-making rises significantly for both female-targeted arms, while for the male-targeted arm a null index conceals offsetting movements: a marginal gain in financial decisions alongside a significant loss of women’s agency over productive decisions, the very activities the programme creates. The qualitative evidence corroborates and extends this pattern. Women across arms report more collaborative decision-making over income generating activities more than over budgeting, and budgeting moves mainly where the couples training addressed it directly; the driver women name is the programme’s gender messaging rather than the cash; and the unchanged arrangement of always consulting the husband, an answer a survey would record as joint decision-making, is concentrated in the male-targeted arm.

The introduction posed two readings of the recurring pattern in which women’s resources improve while their measured decision-making agency barely moves: that targeting and training are insufficient to shift decision-making power, or that the instruments are not capturing the change that occurs. Our evidence distinguishes them by domain. Where change occurred, the indices concealed it: beneath the male arm’s null decision-making index, agency moved in offsetting directions, and beneath survey reports of joint decision-making, women described unchanged deference. Where the indices and the accounts agree that nothing moved, in freedom of movement and in fertility and education decisions, the absence is real rather than an artefact of measurement. The changes that occur are real but domain-specific and early: visible beneath the indices, and in the women’s own accounts, but not at the level at which empowerment is conventionally measured.

One noteworthy benefit of combining the RCT with the QuIP methodology is the pinpointing of precise aspects of change in people’s lives in a way that is independent from researcher priors, or even the intervention’s ex ante theory of change. This may allow for identification of specific, more narrowly defined aspects of larger concepts that would ordinarily be buried in composite indices. Pre-analysis plans were in large part motivated to avoid the cherry picking of results by researchers (Olken, 2015). Participant-led evidence, collected by blindfolded researchers and read against sub-components fixed before analysis, offers a disciplined route to the same end: researchers led to the changes taking place by the perspective of the participants themselves.

These findings carry limitations. Resources and agency are observed at the same follow-up, so their ordering within the interval is an assumption of the framework rather than something the data establish. The component estimates are secondary to the pre-specified aggregates and should be read as such. The QuIP samples are purposive and small, and counts should not be interpreted as representative of a wider population. And the companion paper shows that some of these early changes, including the aggregate decision-making result and wives' perceptions of chore sharing, attenuate by 17 months (Bedi et al., 2026c); the qualitative finding that domestic tasks are still perceived as the woman's responsibility suggests why.

If gender empowerment is a key objective of development, these findings highlight the importance of who in the household is targeted: the recipient shapes which domains of a woman's agency move, and in which direction, even where aggregate measures register little. The mixed method design, conforming to the quant-led model of Copestake (2024), allowed a deeper exploration of the mechanisms of change than either method alone, and we hope this paper inspires further research that integrates mixed methods to enrich our understanding of complicated stories of social change.

Notes

¹Our RCT is registered with the AEA RCT Registry (AEARCTR-0007605). Ethical approval was granted by the Research Ethics Committees of Trinity College Dublin and Innovations for Poverty Action.

²Umodzi, meaning ‘united’, engages couples on gender norms, power, decision-making, budgeting, violence, positive parenting and healthy relationships, through twelve sessions of two to three hours, some mixed sex and some single sex.

³Bedi et al. (2026c) reports the household-level welfare, income and asset outcomes together with the index-level empowerment results at both follow-ups, and develops a household production framework in which the couples training eases intra-household coordination frictions. Bedi et al. (2026b) asks whether the programme’s effect on intimate partner violence is better explained by bargaining, by the alleviation of poverty stress, or by backlash. The household overall welfare outcomes, broader mechanisms of change, spillovers and medium-run trajectories belong to those two papers.

⁴Blindfolded qualitative survey means that surveyors were not informed of the treatment arms, the hypotheses, any respondent’s assignment, the intervention of interest or the name of the implementing organisation.

⁵We converted to 2018 PPP-adjusted USD using the World Bank PPP conversion factor where 1 USD = 260.245 MWK, 2018

⁶In total 36 QuIP interviews were conducted for each of the two female-targeted treatment arms, and 24 QuIP interviews were conducted for the male-targeted treatment arm, with 48 interviews in each round.

⁷Consent was obtained from respondents prior to starting interviews, with an explanation of the reasons for holding an open-ended interview without any reference to a specific programme so that they felt able to share anything they wanted to.

⁸The causal maps use a feature called ‘thread tracing’ which ensures that every link is connected to the selected factor being filtered for (visible by a highlight around the box). This ensures that the factor label counts don’t include mentions which may have been influenced by other drivers not shown in this particular filter.

⁹For the proxy means test, ineligible household were characterized by having a cow or four or more goats OR a home constructed of iron sheets AND finished walls. An exception was made for households where the household head’s primary work was ganyu (day laborer), such households were classified as poor.

¹⁰The FGDs were split between district, sex and treatment arm, two with men and two with women in each district. Field researchers invited respondents from the household interviews to attend and encouraged them to bring one or two friends. The FGDs followed the same QuIP format but asked about changes in the community as a whole rather than in respondents’ own lives. They were not coded into the counts by treatment arm reported below, which are based on the 96 individual interviews, and we draw on them only where they corroborate the individual accounts.

¹¹In the case of 8 households (0.27%), we are missing the male survey, but the household is composed of the original couple living together

¹²Because of the criteria that this is the same female interviewed at baseline and follow-up who are in a relationship, we do not include 250 households that either have a new female spouse, that are divorced and or widowed. We are also not able to track 129 households between the two rounds.

¹³Our sample has 1,487 households from cohort 1 and 1,447 households from cohort 2. Data collection for cohort 1 ran from August to October 2020, and August to November 2021 for cohort 2. The QuIP was conducted for each cohort one year apart and a month after the quantitative survey.

¹⁴Codes in brackets after any quotes in this brief link the quote to one of the 96 anonymised respondents.

¹⁵The instrument did not capture who controlled the transfer; this detail has since been incorporated into similar questions in subsequent studies.

¹⁶Definitions of each outcome can be found in Table A5

¹⁷Note that although very similar to the relevant section of Figure 1, this map is restricted to only direct links to gender/relationship messaging; as an example there is one less count of the factor ‘Fewer disagreements’ since one of these was directly linked to the cash transfer and is therefore not included in this map.

¹⁸We include only chore spouse load in the multiple hypothesis testing correction, since the other two are cross-checks rather than separate outcomes.

References

- Almås, I., Armand, A., Attanasio, O., and Carneiro, P. (2018). Measuring and changing control: Women’s empowerment and targeted transfers. *The Economic Journal*, 128(612):F609–F639.
- Anderson, M. L. (2008). Multiple inference and gender differences in the effects of early intervention: A reevaluation of the abecedarian, perry preschool, and early training projects. *Journal of the American Statistical Association*, 103(484):1481–1495.
- Attanasio, O. and Lechene, V. (2002). Tests of income pooling in household decisions. *Review of Economic Dynamics*, 5(4):720–748.
- Bandiera, O., Burgess, R., Das, N., Gulesci, S., Rasul, I., and Sulaiman, M. (2017). Labor Markets and Poverty in Village Economies. *The Quarterly Journal of Economics*, 132(2):811–870.
- Banerjee, A., Duflo, E., Goldberg, N., Karlan, D., Osei, R., Pariente, W., Shapiro, J., Thuysbaert, B., and Udry, C. (2015). A multifaceted program causes lasting progress for the very poor: Evidence from six countries. *Science (New York, N.Y.)*, 348:1260799.
- Bedi, T., Jose, A., King, M., and McArdle, S. (2026a). Who recovers? Gender, coping strategies, and the psychological costs of Cyclone Idai in Malawi. TEP Working Paper 1726.
- Bedi, T., King, M., Ramos, A., and Vaillant, J. (2026b). Bargaining, poverty stress, or backlash? reducing violence among the ultra-poor. TEP Working Paper 1626, Department of Economics, Trinity College Dublin.
- Bedi, T., King, M., and Vaillant, J. (2026c). Household frictions and reaching the frontier: Experimental evidence from Malawi. Trinity Working paper.
- Bedoya, G., Coville, A., Haushofer, J., Isaqzadeh, M., and Shapiro, J. P. (2019). No household left behind: Afghanistan targeting the ultra poor impact evaluation. NBER Working Paper 25981, National Bureau of Economic Research, Cambridge, MA.
- Bernard, T., Doss, C., Hidrobo, M., Hoel, J., and Kieran, C. (2020). Ask me why: Patterns of intrahousehold decision-making. *World Development*, 125:104671.
- Bossuroy, T., Goldstein, M., Karlan, D., Kazianga, H., Parient, W., Premand, P., Udry, C., Vaillant, J., and Wright, K. A. (2022). Tackling psychosocial and capital constraints to alleviate poverty. *Nature*, 605(7909):291–297.
- Boyer, C., Paluck, E. L., Annan, J., Nevatia, T., Cooper, J., Namubiru, J., Heise, L., and Lehrer, R. (2022). Religious leaders can motivate men to cede power and reduce intimate partner violence: Experimental evidence from Uganda. *Proceedings of the National Academy of Sciences*.

- Concern Worldwide (2018). Mutu Umodzi Suzenza Denga: A Training Manual for Engaging Men for Gender Equality Within the Graduation Programme. Training manual, Concern Worldwide.
- Copestake, J. (2021). Case selection for robust generalisation: lessons from quip impact evaluation studies. *Development in Practice*, 31(2):150–160.
- Copestake, J. (2024). Mixed-methods impact evaluation in international development practice: distinguishing between quant-led and qual-led models. *Journal of Development Effectiveness*, 0(0):1–20.
- Donald, A., Koolwal, G., Annan, J., Falb, K., and Goldstein, M. (2020). Measuring women’s agency. *Feminist Economics*, 26(3):200–226.
- Doyle, K., Levto, R., Barker, G., Bastian, G., Bingenheimer, J., Kazimbaya, S., Nzabonimpa, A., Pulerwitz, J., Sayinzoga, F., Sharma, V., and Shattuck, D. (2018). Gender-transformative Bandedereho couples’ intervention to promote male engagement in reproductive and maternal health and violence prevention in Rwanda: Findings from a randomized controlled trial. *PLoS ONE*.
- Dunkle, K., Stern, E., Chatterji, S., and Heise, L. (2020). Effective prevention of intimate partner violence through couples training: a randomised controlled trial of Indashyikirwa in Rwanda. *BMJ Global Health*, 5(12).
- Fiszbein, A., Schady, N., Ferreira, F., Grosh, M., Keleher, N., Olinto, P., and Skoufias, E. (2009). *Conditional Cash Transfers: Reducing Present and Future Poverty*. The World Bank Group.
- Haushofer, J. and Shapiro, J. (2016). The short-term impact of unconditional cash transfers to the poor: Experimental evidence from Kenya. *The Quarterly Journal of Economics*, 131:qjw025.
- Hoddinott, J. and Haddad, L. (1995). Does female income share influence household expenditures? evidence from cote d’ivoire. *Oxford Bulletin of Economics and Statistics*, 57(1):77–96.
- Imbens, G. W. and Rubin, D. B. (2015). *Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction*. Cambridge University Press, New York.
- Ismayilova, L., Karimli, L., Gaveras, E., To-Camier, A., Sanson, J., Chaffin, J., and Nanema, R. (2018). An integrated approach to increasing women’s empowerment status and reducing domestic violence: Results of a cluster-randomized controlled trial in a West African country. *Psychology of Violence*, 8(4):448–459.
- Kabeer, N. (1999a). The conditions and consequences of choice: Reflections on the measurement of women’s empowerment. UNRISD Discussion Paper 108, United Nations Research Institute for Social Development.

- Kabeer, N. (1999b). Resources, agency, achievements: Reflections on the measurement of women's empowerment. *Development and Change*, 30(3):435–464.
- Karimli, L., Lecoutere, E., Wells, C. R., and Ismayilova, L. (2021). More assets, more decision-making power? mediation model in a cluster-randomized controlled trial evaluating the effect of the graduation program on women's empowerment in Burkina Faso. *World Development*, 137:105–159.
- Kling, J. R., Liebman, J. B., and Katz, L. F. (2007). Experimental analysis of neighborhood effects. *Econometrica*, 75(1):83–119.
- Laszlo, S. (2019). The gender transformative potential of graduation programs. (No. 25).
- Lundberg, S. J., Pollak, R. A., and Wales, T. J. (1997). Do husbands and wives pool their resources? evidence from the united kingdom child benefit. *The Journal of Human Resources*, 32(3):463–480.
- Olken, B. A. (2015). Promises and perils of pre-analysis plans. *Journal of Economic Perspectives*, 29(3):61–80.
- Somville, V., Almås, I., and Vandewalle, L. (2020). The effect of gender-targeted transfers: Experimental evidence from India. *NHH Dept. of Economics Discussion Paper*, (16).
- Thomas, D. (1990). Intra-household resource allocation: An inferential approach. *The Journal of Human Resources*, 25(4):635–664.

8 Appendix

8.1 Additional Tables

Table A1: Summary and Baseline Balance

	Total Sample Mean	Control Mean	T1 Female Targeted Mean	T2 Male Targeted Mean	T3 Female plus Umodzi Mean	Spillover Mean	C vs T1 p-value	C vs T2 p-value	C vs T3 p-value	C vs Sp p-value	T1 vs T2 p-value	T1 vs T3 p-value	T2 vs T3 p-value	Joint F Test	Pooled F Test	C-T1 Norm Diff	C-T2 Norm Diff	C-T3 Norm Diff	T1-T2 Norm Diff	T1-T3 Norm Diff	T2-T3 Norm Diff
Demographics																					
Female Spouse Age	35.38	36.01	35.23	34.93	35.80	35.07	0.30	0.13	0.78	0.90	0.69	0.44	0.21	0.40	0.24	0.06	0.09	0.02	0.02	-0.05	-0.07
Male Spouse Age	42.53	42.89	42.63	42.23	43.07	42.03	0.78	0.36	0.78	0.97	0.65	0.62	0.27	0.70	0.72	0.02	0.05	-0.01	0.03	-0.03	-0.06
Female Sp Educ	1.95	1.75	1.80	2.18	1.98	2.00	0.81	0.04	0.23	0.48	0.05	0.31	0.29	0.13	0.14	-0.02	-0.15	-0.08	-0.13	-0.06	0.07
Male Sp Educ	3.75	3.51	3.59	3.99	3.73	3.87	0.67	0.04	0.36	0.54	0.06	0.55	0.22	0.15	0.16	-0.02	-0.12	-0.06	-0.10	-0.04	0.07
HH Size	5.52	5.54	5.45	5.50	5.63	5.49	0.46	0.77	0.53	0.65	0.65	0.16	0.34	0.55	0.90	0.05	0.02	-0.05	-0.03	-0.10	-0.07
Household Assets																					
Total Type of Assets	15.57	15.23	15.48	15.95	15.64	15.55	0.56	0.07	0.28	0.39	0.18	0.60	0.37	0.30	0.18	-0.05	-0.16	-0.09	-0.11	-0.04	0.07
Total No. of Assets	46.49	45.87	43.67	49.39	46.41	46.94	0.28	0.12	0.72	0.93	0.01	0.16	0.21	0.11	0.66	0.08	-0.10	-0.02	-0.17	-0.11	0.09
Total Structures for Sleep	1.43	1.42	1.40	1.48	1.48	1.40	0.75	0.21	0.24	0.54	0.19	0.20	0.95	0.36	0.35	0.03	-0.08	-0.08	-0.10	-0.10	-0.00
Cement or Solid Flooring	0.05	0.06	0.04	0.05	0.06	0.05	0.18	0.41	0.98	0.73	0.63	0.21	0.44	0.46	0.40	0.09	0.06	0.00	-0.03	-0.09	-0.06
Iron Sheet or Solid Roof	0.27	0.24	0.25	0.29	0.29	0.26	0.82	0.18	0.13	0.58	0.27	0.23	0.96	0.32	0.20	-0.02	-0.10	-0.10	-0.09	-0.09	0.00
Burnt Brick or Solid Wall	0.32	0.32	0.30	0.34	0.34	0.32	0.40	0.71	0.63	0.72	0.26	0.22	0.93	0.59	0.99	0.06	-0.04	-0.04	-0.09	-0.10	-0.01
Dwelling Index	0.64	0.63	0.59	0.68	0.69	0.63	0.53	0.52	0.37	0.86	0.22	0.13	0.84	0.42	0.69	0.04	-0.05	-0.07	-0.10	-0.11	-0.02
Borehole Water Source	0.95	0.93	0.93	0.95	0.95	0.96	0.97	0.45	0.38	0.15	0.51	0.44	0.95	0.75	0.51	-0.01	-0.10	-0.11	-0.09	-0.10	-0.01
Infrastructure (Hours Distance to)																					
Tarmac Road	5.47	5.41	5.40	5.65	5.45	5.44	0.94	0.15	0.84	0.78	0.18	0.81	0.30	0.44	0.53	0.01	-0.18	-0.03	-0.18	-0.03	0.14
Public Health Clinic	3.95	3.95	3.90	3.90	4.02	3.98	0.77	0.76	0.71	0.33	1.00	0.46	0.46	0.86	0.93	0.03	0.03	-0.05	-0.00	-0.09	-0.09
Daily Market	3.12	3.23	3.01	3.08	3.19	3.09	0.08	0.25	0.82	0.57	0.54	0.11	0.33	0.23	0.22	0.18	0.12	0.03	-0.06	-0.15	-0.09
Weekly Market	0.13	0.09	0.11	0.17	0.13	0.13	0.81	0.15	0.51	0.89	0.25	0.69	0.45	0.51	0.29	-0.04	-0.23	-0.10	-0.19	-0.07	0.12
Mobile Agent	2.36	2.31	2.39	2.31	2.41	2.36	0.44	0.98	0.36	0.38	0.42	0.87	0.35	0.68	0.48	-0.08	-0.00	-0.09	0.08	-0.01	-0.09
Primary School	2.42	2.48	2.45	2.35	2.44	2.41	0.72	0.14	0.66	0.80	0.30	0.94	0.32	0.50	0.35	0.04	0.15	0.05	0.11	0.01	-0.10
Secondary School	3.98	4.14	3.87	3.94	4.02	3.96	0.04	0.14	0.42	0.28	0.53	0.22	0.53	0.18	0.08	0.23	0.16	0.09	-0.07	-0.13	-0.06
Observations	3,312	625	595	594	599	899															

Control household observations are missing 1 baseline male spouse survey. Total control households at baseline is 626, matching Table A2.

Note: Note:

Table A2: Attrition Table: Household and Partnered Original Female Spouse

	Control	T1 Female Targeted	T2 Male Targeted	T3 Female plus Umodzi	Spillover Households	T1-C p-value	T2-C p-value	T3-C p-value	Sp-C p-value	T3-T1 p-value	T3-T2 p-value	T1-T2 p-value
Follow-up 1												
Mean	0.96	0.96	0.98	0.96	0.95	0.87	0.14	0.98	0.31	0.88	0.11	0.11
Observations	602	571	581	576	854							
Follow-up 1 Original Female, Partnered												
Mean	0.89	0.88	0.90	0.90	0.87	0.52	0.72	0.68	0.11	0.31	0.98	0.35
Observations	559	524	534	539	778							
Baseline Obs	626	595	594	599	899							

Table A3: Individual components of Freedom of Movement and Contact Index at Baseline

	Full Sample Mean
Movement Agency	
Prevented to see family/friends	0.04
Prevented to work outside	0.03
Tries to prevent seeing friends	0.06
Tries to restrict contact with family	0.04
Wants to know where you are always	0.48
Expects you to ask permission to visit health care	0.28
Observations	3,307

Table A4: Multiple Hypothesis Testing for Main Four Headline Indices

	T1vsC		T2vsC		T3vsC		T1vsT2		T1vsT3		T2vsT3	
	p	q	p	q	p	q	p	q	p	q	p	q
Resource Index	0.00	0.00	0.00	0.00	0.00	0.00	0.19	0.28	0.15	0.28	0.00	0.02
Decision Agency	0.47	0.53	0.46	0.53	0.05	0.13	0.19	0.28	0.31	0.45	0.02	0.05
Agency Evidence	0.17	0.28	0.61	0.53	0.24	0.34	0.39	0.52	0.78	0.56	0.54	0.53
Chore Spouse Load	0.38	0.52	0.59	0.53	0.00	0.02	0.75	0.56	0.06	0.13	0.03	0.07

Table A5: Quantitative Variable Definitions

Notes: Index constructions follow Bossuroy et al. (2022) and match the companion paper (Bedi et al., 2026c) exactly; they were fixed before endline data were analysed. The mapping of measures to concepts follows Kabeer (1999b).

Resource	Outcomes
Index	Standardised index of 7 components: 1) Whether the female spouse receives a share of the agriculture income; 2) Whether she has a business; 3) Whether the income from her business goes to herself/jointly shared; 4) Whether she owns livestock; 5) Whether she gets a share of livestock sales; 6) Whether she has her own wage income; 7) Whether she saves via a mechanism that is outside of the household, family and friends.
Agency	Outcomes
	<i>Decision-Making</i>
Decision-Making Index	A standardised weighted average of four decision-making indices based on female spouse response: 1) Financial Decisions Index; 2) Productive Decisions Index; 3) Household Decisions Index; 4) Fertility and Education Decisions Index.
Financial Decisions Index	Standardised index of 3 components: 1) who made the decision on taking a loan out; 2) on who made the decision on how to use this loan; 3) on who made the decision on how to use her savings.
Productive Decisions Index	Standardised index of 4 components. Who makes decisions between her spouse and herself on 1) Agriculture; 2) Spending income she earned personally; 3) Household businesses and income generating activities; 4) Livestock.
Household Decisions Index	Standardised index of 5 components. Who makes decisions between her spouse and herself on 1) Household major purchases; 2) Household current expenses. 3) With income she earned, she can purchase clothing for herself or children without asking the permission of anyone else. 4) When she has small amounts of money, how often can she decide how to spend it on her own. 5) How often her opinion is listened to in decisions related to expensive purchases.
Fertility and Education Decisions Index	Standardised index of 2 questions on who makes decisions on 1) contraception and family planning; 2) children's education.
	<i>Agency Evidence</i>

Political and Community Involvement Agency Index:	A standardised index of 8 standardised questions on the female spouse's involvement in political activities, whether she attended a community meeting, complained/sought help from public authorities, total community groups involved in, whether she mobilises communities for meetings and if she is a community leader in any way. The higher the score, the more involved the individual is politically and/or in the community.
Movement and Contact Agency Index:	Standardised index of 4 questions related to her freedom of movement and contact. The questions look at whether her current or ex-spouse tried to keep her from seeing her friends, restricts contact with her birth family, needs to know where she is at all times and if she needs permission before seeking health care.
	<i>Role Sharing</i>
Chore load for spouse:	A standardised index reported by the female on whether her husband never, rarely or more regularly (weekly or more often) participated in the following 7 activities: 1) preparing food; 2) cleaning the house; 3) caring for your children; 4) purchasing food; 5) collecting water; 6) collecting firewood; 7) washing clothes.
Chore own load:	A standardised index reported by the female on whether she never, rarely or more regularly (weekly or more often) participated in the following 7 activities: 1) preparing food; 2) Cleaning the house; 3) Caring for your children; 4) Purchasing food; 5) Collecting water; 6) Collecting firewood; 7) Washing clothes.