

[Naila Kabeer](#)

## Randomized control trials and qualitative evaluations of a multifaceted program for women in extreme poverty: empirical findings and methodological reflections

**Article (Accepted version)  
(Refereed)**

**Original citation:**

Kabeer, Naila (2018) *Randomized control trials and qualitative evaluations of a multifaceted program for women in extreme poverty: empirical findings and methodological reflections*. *Journal of Human Development and Capabilities*. ISSN 1945-2829

© 2018 [Human Development and Capability Association](#)

This version available at: <http://eprints.lse.ac.uk/90486/>

Available in LSE Research Online: October 2018

LSE has developed LSE Research Online so that users may access research output of the School. Copyright © and Moral Rights for the papers on this site are retained by the individual authors and/or other copyright owners. Users may download and/or print one copy of any article(s) in LSE Research Online to facilitate their private study or for non-commercial research. You may not engage in further distribution of the material or use it for any profit-making activities or any commercial gain. You may freely distribute the URL (<http://eprints.lse.ac.uk>) of the LSE Research Online website.

This document is the author's final accepted version of the journal article. There may be differences between this version and the published version. You are advised to consult the publisher's version if you wish to cite from it.



**Randomized control trials and qualitative evaluations of a multifaceted program for women in extreme poverty: empirical findings and methodological reflections**

Journal:	<i>Journal of Human Development and Capabilities</i>
Manuscript ID	CJHD-2018-0061.R1
Manuscript Type:	Special Issue Article
Keywords:	Capability Approach, Gender, Measurement, Poverty, Well-being
Abstract:	<p>This paper sets out to synthesize key lessons from studies using alternative methodologies to impact assessment. Drawing on Sen's capability approach as a conceptual framework, it analyses two pairs of impact assessments which were carried out in West Bengal and Sindh around the same time and within close proximity to each other. Each pair consisted of a randomized control trial and a qualitative assessment of attempts to pilot BRAC's approach to transferring assets to women in extreme poverty. The paper reports on the findings of these studies, their strategies for establishing their claims about causality and the information base they drew on to establish these claims. It finds that not only did the RCTs fail to meet their own criteria for establishing causality, but they also provided very limited explanation for the patterns of outcomes observed. Such information formed the substance of the qualitative studies. The paper concludes that greater use of mixed methods could help to offset some of limitations of RCTs and to place their findings on much firmer ground.</p>

SCHOLARONE™  
Manuscripts

# Randomized control trials and qualitative evaluations of a multifaceted program for women in extreme poverty: empirical findings and methodological reflections

Naila Kabeer

## Abstract

*This paper sets out to synthesize key lessons from studies using alternative methodologies to impact assessment. Drawing on Sen's capability approach as a conceptual framework, it analyses two pairs of impact assessments which were carried out in West Bengal and Sindh around the same time and within close proximity to each other. Each pair consisted of a randomized control trial and a qualitative assessment of attempts to pilot BRAC's approach to transferring assets to women in extreme poverty. The paper reports on the findings of these studies, their strategies for establishing their claims about causality and the information base they drew on to establish these claims. It finds that not only did the RCTs fail to meet their own criteria for establishing causality, but they also provided very limited explanation for the patterns of outcomes observed. Such information formed the substance of the qualitative studies. The paper concludes that greater use of mixed methods could help to offset some of limitations of RCTs and to place their findings on much firmer ground.*

## 1. Introduction

### 1.1 Background to the research

In 2002, BRAC launched the first phase of its new Targeting the Ultra Poor (TUP) programme in Bangladesh. The programme was intended to promote entrepreneurship among women in extreme poverty in order to 'graduate' them into its mainstream microfinance programme within a period of two years. Several decades of experience of working with poor people in Bangladesh had established for BRAC that the extreme poor were not only poorer than other sections of the poor, but structurally different because of the chronic, multiple, overlapping and frequently gendered nature of the constraints that kept them trapped in poverty.

The comprehensive approach to addressing extreme poverty which evolved from this theoretical understanding can be seen to exemplify important aspects of Sen's capability approach (Sen, 1999; Robeyns, 2005). First of all, while the transfer of livelihood assets to women in extreme poverty was a central component of the programme, it was clearly conceptualized as 'means' rather than 'ends'. The ultimate objective of the programme was to promote the capabilities of these women to convert their assets into viable enterprises and thereby achieve goals that they had reason to value.

And secondly, the programme incorporated one of the key insights of the capability approach: that access to resources (such as those distributed by a programme) do not automatically or uniformly translate into valued goals, because certain 'conversion factors' can block or differentiate people's ability to translate the resources at their disposal into the capabilities needed to achieve these goals. These conversion factors reflect the characteristics

1  
2  
3 of individuals and their families as well as aspects of the wider physical and socio-economic  
4 context in which they are located.  
5

6 These insights are reflected in the approach taken by the TUP programme. Along with  
7 transfer of livelihood assets, most often livestock and poultry, the TUP programme included  
8 a number of other components intended to work in tandem on various constraints in order to  
9 promote women's ability to convert their assets into, in the first instance, entrepreneurial  
10 capabilities and subsequently into improvements in different aspects of their lives.  
11

12 One of these components was its approach to selecting eligible households. Eligibility was  
13 defined in terms of two or more of the following criteria: lack of land; lack of productive  
14 assets; absence of an adult male breadwinner; reliance on daily wage labour or begging for a  
15 living; and school-going children were at work. Excluded from the programme were  
16 households participating in any microfinance or other anti-poverty programmes and those  
17 that did not have an adult female member.  
18

19 Local BRAC staff carried out a combination of spatial poverty mapping, participatory wealth  
20 ranking and household surveys in order to identify households that met these criteria. As  
21 Matin et al (2008) pointed out, not only was this triangulation process intended to increase the  
22 accuracy of the targeting process, but it was anticipated that the social interactions involved  
23 would help to create a common understanding of the programme's approach and rationale  
24 among those responsible for implementing the programme and in the wider community.  
25  
26

27 A second component was a monthly cash transfer for the period of the programme to allow  
28 participants to focus on building their enterprises rather than having to engage in other  
29 livelihood activities to meet their daily needs. Once their enterprises started to yield returns,  
30 participants were required to save part of their stipend so that they had a lump sum to fall  
31 back on when the programme ended. Health support, in the form of enabling access to free  
32 government services, helped to protect households from the debilitating effects of health  
33 shocks.  
34

35 The TUP programme also set up Village Assistance Committees made up of local elites to  
36 provide advice and support to participants. Considerable thought given to this component,  
37 given the unequal nature of the relationships between TUP participants and village elites. But,  
38 in the light of the reliance of the ultra-poor on powerful households within their communities  
39 for patronage and protection, it was decided that on balance, efforts to enlist elite  
40 involvement in the project was preferable to leaving them out and risking their antagonism.  
41  
42

43 Finally, intensive staff engagement was considered central for building women's individual  
44 capabilities, their sense of self-confidence and the ability to take risks and plan ahead. These  
45 cognitive and subjective qualities were likely to be absent or poorly developed among socially  
46 isolated women living close to the margins of survival in a highly patriarchal society, but were  
47 considered essential if they were to translate their resources into viable enterprises (Matin et  
48 al., 2009).  
49

## 50 **1.2 Objectives of the paper**

51  
52 Positive evaluations of the first phase of the program suggested that it might offer a  
53 generalizable model for tackling extreme poverty within a time-bound period. In 2006, CGAP  
54 and Ford Foundation set up the Graduation Programme which brought together a number of  
55 organisations to adapt, pilot and evaluate the TUP approach in eight countries across the  
56  
57

1  
2  
3 world<sup>1</sup>. Randomized control trials (RCTs) were the preferred evaluative methodology and six  
4 of these were carried out on a co-ordinated basis between 2007 and 2014 in order to test the  
5 validity of the approach across different contexts – including the state of West Bengal in India  
6 and the province of Sindh in Pakistan.  
7

8 A synthesis report of their findings was very positive about TUP impacts, concluding that the  
9 programme had achieved its primary goals of bringing about sustained increases in  
10 consumption and income (Banerjee et al., 2015). The findings also appeared to endorse the  
11 external validity of the approach. According to the authors, the diversity of circumstances in  
12 which these pilots were carried ‘should give us a high level of confidence in the robustness of  
13 the impact to variations in both the context and the implementation agency’ (2015, p.3).  
14

15 At the same time, the Graduation programme commissioned a team of researchers, including  
16 myself, to carry out qualitative evaluations of two TUP pilots in Sindh and West Bengal, a few  
17 hundred miles away from those studied by the RCTs<sup>2</sup>. These fared very  
18 differently from each other: at the end of the pilot phase, the project in Sindh was closed  
19 down while the one in West Bengal was scaled up.  
20

21 The objective of this paper is to use the two pairs of studies in West Bengal and Sindh to  
22 reflect on what we can learn in both empirical and methodological terms from alternative  
23 approaches to impact assessment. What is of particular value in this comparison is that the  
24 interventions under study were very similar, indeed adaptations of a common programme  
25 approach, and that within each geographical location, they were carried out in close proximity  
26 to each other. The paper will discuss the findings of the different studies and assess their  
27 strategies for establishing claims about causality.  
28  
29

## 30 **2. Counter-factual approaches to impact assessment**

### 31 **2.1 RCTs as impact assessment methodology**

32  
33 RCTs were prioritized in the Graduation Programme because they are currently held to be the  
34 ‘gold standard’ in evaluation methodologies. They use what has been described as the  
35 ‘counterfactual’ approach to impact assessment (Shaffer, 2011), seeking to establish how  
36 participants of a program would have fared in the absence of that programme. They do this  
37 by randomly assigning a sample of the population deemed eligible for the programme to  
38 ‘treatment’ and ‘control’ groups and comparing how they fare on selected outcome indicators.  
39 RCT advocates claim that, provided what Cartwright calls ‘the idealisation assumption’ is  
40 satisfied, they can estimate the average impacts of the project by measuring the difference in  
41 the mean outcomes reported by ‘treatment’ and ‘control’ groups (the ATE or average  
42 treatment effect). According to the idealisation assumption, the randomization process will  
43 ensure that all factors likely to influence the outcome variables, aside from the treatment  
44 itself, are distributed identically between treatment and control groups<sup>3</sup>. The satisfaction of  
45 this assumption renders it unnecessary to take account of these other factors - or even to  
46 know what they might be (Krauss, 2018). Consequently, RCTs minimize reliance on  
47 theoretical predictions or prior knowledge that might be otherwise needed to identify these  
48 factors (Barret and Carter, 2010).  
49  
50  
51

52 However, many critics believe that the claims made for the methodological superiority of RCTs  
53 are overstated and that the methodology represents ‘a baser metal than gold’ (Barret and Carter,  
54 p. 516). While their criticisms cover a range of issues, two are of particular relevance to this  
55 paper. The first is that it is seldom possible to replicate the strict conditions required by a pure  
56  
57

1  
2  
3 RCT once experiments are moved from the laboratory into the messy conditions of the real  
4 world where the idealisation assumption seldom holds (Shaffer, 2011; Deaton and Cartwright,  
5 2018). As Barret and Carter (p. 521) point out, ‘the clean identification of randomization gets  
6 compromised by human agency’: the wilful or accidental, hidden or overt, expected or  
7 unanticipated agency of the multitude of actors that inhabit the project’s domain of operation.  
8

9  
10 The second criticism relates to its approach to establishing causality. According to its  
11 proponents, given a sufficiently large sample size, any differences in outcomes that are  
12 observed between treatment and control groups can be confidently attributed to the  
13 programme, since all other observed and unobserved factors that might otherwise have  
14 influenced these differences have been eliminated (Karlan and Parienté, 2014). The focus of  
15 RCTs is therefore on the ‘effects’ of causes rather than the causal processes themselves  
16 (Shaffer 2011). As a result, most RCTs are not in a position, nor do they seek, to specify  
17 what it is about a programme or the context in which it is implemented that explains  
18 observed outcomes<sup>4</sup>.

19  
20 This is less of a problem when observed outcomes accord with expected outcomes since  
21 studies can then, with some justification, attribute outcomes to the programme. But when  
22 outcomes are heterogeneous, fail to emerge or are of an unexpected nature, RCTs are unable  
23 to distinguish whether this reflects factors internal to the programme or some aspect of the  
24 wider context (White, 2009). The next sections discuss the RCTs in West Bengal and Sindh  
25 to illustrate both sets of problems.  
26

## 27 **2.2 An RCT study of TUP pilots in Sindh**

28  
29 In Pakistan, the Pakistan Poverty Alleviation Fund (PPAF), a government-donor partnership  
30 which co-ordinates microfinance efforts across the country, was charged with managing the  
31 TUP pilots. PPAF selected five NGOs (Badin Rural Development Society, Aga Khan  
32 Planning and Building Service, Sindh Agricultural and Forestry Workers Coordinating  
33 Organization, Indus Earth Trust and the Orangi Charitable Trust) to carry out the pilots in  
34 villages within the Sindh Coastal Area Development, a particularly deprived part of Sindh  
35 province. The first four of these were included in the RCT study while our study focused on  
36 the fifth.  
37

38  
39 The findings from the Sindh RCT are summarized in a CGAP briefing paper (Karlan and  
40 Parienté 2014) and elaborated in the synthesis paper by Banerjee et al., but the problems  
41 encountered in implementing the randomization process are not discussed in either of these  
42 publications. Information on these can be found instead in an independent assessment study  
43 carried out for the PPAF by Innovative Development Strategies, a local research organization,  
44 at the end of pilot period (IDS 2012). According to this study, each of the five NGOs were  
45 required to select the least developed villages within their designated project area, use PRA  
46 techniques to identify 400 eligible households within these selected villages and then  
47 randomly assign 200 of these households to the project while assigning the remaining 200 to  
48 the control group. This would give a total of 1600 households for the trial sample.  
49

50  
51 This process was not followed in practice. While some organizations did indeed hold public  
52 lotteries to randomly select project participants in the identified villages, others ‘chose to  
53 select half of the villages identified’ (IDS: p. 10). Unfortunately, the IDS study does not  
54 provide any information on the percentages of households that were randomly and non-  
55 randomly assigned. What the study does point out is that this failure to follow the agreed  
56 procedures meant that there were likely to be unknown variations in the background  
57

1  
2  
3 characteristics of the treatment and control groups with no guarantee that they were equally  
4 poor (IDS, 2012, p. 10).  
5

6 In fact, according to Kidd and Bailey-Athias (2017), most of treatment households were not  
7 in extreme poverty. Estimating the proportion of beneficiary households in the six RCT  
8 studies that were below the poverty line of \$1.25 (PPP)<sup>5</sup>, they found that only 10% of those in  
9 the Pakistan study were below this line. This was far lower than estimates for the most of the  
10 other RCTS including West Bengal where it was 70%. Attrition rates of 21% were also higher  
11 in Pakistan than the other RCTs.  
12

13 Putting to one side the serious threats that these posed to both internal and external validity  
14 of the Sindh RCT, we find that the findings from the Sindh pilot reported in the synthesis  
15 study were both fewer and less consistent than those from West Bengal. Positive impacts  
16 sustained across both end lines were reported for just per capita consumption, assets and  
17 political involvement (membership of political party). There were inter-temporal variations  
18 for all other impacts. The impact on the income indicators were insignificant at the first end  
19 line but impacts on livestock earnings were positive and significant at the second. Positive  
20 impacts were reported for time spent in caring for livestock, food security and women's  
21 empowerment at the first end line but had weakened or become insignificant by the second.  
22 Negative impacts on health were reported for the first end line survey but had faded by the  
23 second while impacts on financial inclusion were insignificant in the first end line survey and  
24 negative in the second. Finally, there did not appear to be any impact on self-reported  
25 happiness (collected only in the second end line survey). There is no discussion of possible  
26 explanations for this very mixed pattern of outcomes.  
27  
28

### 29 **2.3 The RCT study of a TUP pilot in West Bengal**

  
30

31 The TUP pilot in West Bengal was implemented in rural areas of Murshidabad district by  
32 Bandhan, one of India's largest microfinance organizations. Its goal was to enable women in  
33 extreme poverty (defined by criteria that closely mirrored those of BRAC) to qualify within 18  
34 months for regular microfinance loans from Bandhan. Following BRAC's eligibility criteria  
35 and targeting methodology, Bandhan identified 991 eligible households for the trial sample.  
36 The base line survey of these households carried out in 2007 confirmed that these  
37 households conformed to these criteria. 512 (52%) of these households were then randomly  
38 selected to participate in the project, while 466 were assigned to the control group.  
39

40 However, of those selected to participate in the programme, 12.5% (64 households) had to be  
41 dropped on grounds of ineligibility. A further 35% (182 households) refused to participate,  
42 mainly out of suspicion of project motivations. Of these non-compliant households, 58%  
43 were Muslims. In the end, only 52% (266 households) of those selected for participation in  
44 the project actually participated.  
45

46 In order to check whether non-compliance on the part of those selected for the project had  
47 biased the sample, the study compared the base line characteristics of non-compliant  
48 households with the rest of the trial sample. It found that they were very similar to the rest of  
49 the sample, except that Muslims were significantly less likely to participate than Hindus (by 18  
50 percentage points). The study allowed for the effects of non-compliance by carrying out the  
51 rest of its analysis on the basis of the 'intention to treat' (ITT) group, all those who had been  
52 selected to participate in the project regardless of whether they participated or not, rather  
53 than those actually treated.  
54  
55

1  
2  
3 Of the 978 households that were included in the base line study<sup>6</sup>, 166 households, 85 from  
4 the 'treatment' group and 81 from the control group, were not available for the first end line  
5 survey<sup>7</sup>. As a result, only 812 (83% of the original sample) were included in the first end line  
6 survey and hence in subsequent estimates of impacts. 429 of these came from ITT group  
7 while 388 came from the control group.  
8

9 To check whether the attrition of households had biased the randomization design, the study  
10 compared households that had dropped out of the end line survey with those who had been  
11 included. Here it found significant differences: attrition households were significantly more  
12 land-poor and had higher dependency ratios than those included in the survey. In addition,  
13 70% were Muslim compared to just 28% of the included population. The authors point out  
14 that this evidence of systematic differences between surveyed and attrition households raised  
15 the need for caution in extrapolating the results of the study to other contexts. At the same  
16 time, the fact that they found very little difference between attrition households in the ITT  
17 and control groups and little evidence that treatment assignment was correlated with attrition  
18 suggested there was less need for concern about its internal validity.  
19

20 The findings of the study for the two end line surveys carried out by the study, one at the end  
21 of the project and one a year later, are discussed in the synthesis study (Banerjee et al., 2015).  
22 It reports a number of positive economic impacts that were sustained over both end line  
23 surveys. These included impacts on per capita consumption, asset holdings, income, food  
24 security, financial inclusion, time spent in productive activities, particularly livestock, and self-  
25 reported economic status. Impacts were also evident on livelihood diversification by the  
26 second end line.  
27

28 The impacts on non-economic indicators were less consistent: positive impacts on physical  
29 health, self-reported happiness and stress levels were reported for at the first end line, but had  
30 disappeared by the second. Positive impacts on political involvement (voting and engagement  
31 with village leaders) emerged only in the second end line survey. There was also no evidence  
32 of any impact on women's empowerment (as measured by their say in decision-making)  
33 which was measured in the first end line. The study does not comment on possible reasons  
34 for this pattern of inconsistent and absent, but it seems clear that certain components of the  
35 project did not work as anticipated.  
36  
37

38 Before concluding the discussion of the West Bengal RCT, it is important to note that while it  
39 took steps to address the possible *statistical* biases introduced by non-compliance and attrition,  
40 it ignored the *social* biases that such behaviour introduced. The impacts reported by the ITT  
41 households were driven by the (at most) 52% of the treatment group who agreed to  
42 participate in the project although we do not know what percentage of this compliant group  
43 were also included in the follow up surveys. Not only was this 52% better off than those who  
44 refused to participate but it was disproportionately drawn from the Hindu population while  
45 those who excluded themselves from the project and later from the follow up surveys were  
46 disproportionately drawn from the Muslim population. The social significance of this pattern  
47 of exclusion is that while Muslims are a marginalized minority in India generally (frequently  
48 classified as Other Backward Castes), West Bengal has a higher percentage of Muslims in its  
49 population than the national average (27% compared to 14%) while Murshidabad has a  
50 higher percentage of Muslims than the average for West Bengal (66% compared to 27%). The  
51 West Bengal RCT thus systematically failed to reach the state's most significant religious  
52 minority group who also tend to be poorer than those in the religious majority.  
53  
54  
55  
56  
57  
58  
59  
60



1  
2  
3 Finally, while the findings from two RCTs discussed so far relate to average treatment effects  
4 for ITT households, the synthesis study used quantile regression analysis to explore the  
5 distribution of the different impacts among the treated populations. It found that while there  
6 were positive and significant impacts on income, consumption and assets for different  
7 quantiles tested, the largest impacts were found among those who started out better off than  
8 the rest. Only in the case of food security did poorer households report larger impacts. It is  
9 interesting to note that, despite the stated goal of the TUP approach to promote the  
10 livelihood capabilities of poor women, the report is studiously gender-neutral in its discussion  
11 of impacts, part from an indicator on women's empowerment.  
12

### 13 **3. Mechanism-based approaches to impact assessment**

#### 14 **3.1 Focusing on causal processes**

15  
16 Since RCTs are set up to ensure that the programme is the sole driver of impact, they focus  
17 on establishing and estimating average impacts. They do not ask further questions about how  
18 the programme worked and who it worked for. These questions are pivotal to 'mechanism-  
19 based' approaches to evaluation which are explicitly concerned with understanding the causal  
20 processes which generate observed outcomes. Common elements in these approaches include  
21 the need for a programmatic theory of change which spells out the causal pathways through  
22 which a programme is expected to work, the relevance of contextual factors in explaining  
23 whether, for whom and under what circumstances the programme is likely to work and an  
24 analytical narrative which draws on relevant sources of information along with the findings of  
25 the evaluation in order to explain how the programme actually worked (or failed to work) and  
26 for whom (Shaffer, 2011; Pawson and Tilly, 2004; White, 2009).  
27  
28  
29  
30

31 Such approaches clearly resonate with the capabilities-based conceptualization of the TUP  
32 approach we set out earlier but as evaluation frameworks, they are explicitly concerned with  
33 causal processes through which programme interventions convert into impacts<sup>8</sup>. They pay  
34 particular attention to human agency as a key mechanism mediating this conversion process.  
35 As Pawson and Tilly (2004) put it: 'the triggers of change in most interventions are ultimately  
36 located in the reasoning and resources of those touched by the programme' (p. 5).  
37

38 Mechanism-based approaches are not necessarily tied to any particular methodology. They  
39 may prioritize quantitative methods in order to measure the association between program  
40 activities and outcomes and disentangle the impact of other possible causal factors. They may  
41 rely qualitative methods in order to identify potentially generalizable mechanisms that explain  
42 how and in what contexts projects can be expected to work. In addition, there is growing  
43 literature which calls for combining different methods in order to benefit from the strengths  
44 of each (White, 2009; Shaffer, 2011; Deaton and Cartwright, 2018).  
45  
46

47 Unfortunately, while a number of RCT studies have begun to incorporate mixed methods,  
48 many leading advocates of RCTs remain resistant. This was the main reason why our  
49 qualitative evaluations of TUP pilots in Sindh and West Bengal, although part of the larger  
50 Graduation program, were carried out as stand-alone studies rather than as part of a mixed  
51 methods approach.  
52  
53

#### 54 **3.2 The qualitative evaluations of TUP pilots in Sindh and West Bengal: note on 55 methodology**

1  
2  
3  
4 The TUP pilot in Sindh was implemented by Orangi Charitable Trust (OCT) one of the five  
5 NGOs selected by the PPAF. It worked with 200 women in 11 villages between February,  
6 2008 and May, 2010. In West Bengal, the pilot was implemented by Trickle Up, an  
7 international NGO, in partnership with a local NGO, Human Development Centre (HDC),  
8 which was responsible for field-level operations. It worked with 300 in 10 villages in South 24  
9 Parganas district between August 2007 and August 2009. Our research was carried out  
10 between May 2009 and May 2010, covering the final year of the Sindh pilot and the final four  
11 months and eight months thereafter of the West Bengal pilot.  
12

13 The research team worked closely with the implementing organizations in the two contexts to  
14 select the study locations and participants for the study. It used village mapping to select five  
15 villages in the West Bengal project area, some relatively well-endowed and others more  
16 isolated and deprived. In Pakistan, where OCT's project locations were dispersed across  
17 urban, peri-urban and rural locations, we selected four hamlets in the more disadvantaged  
18 rural locations.  
19

20 A local member of the research team carried out semi-structured interviews with project staff  
21 and more detailed, in-depth interviews with participants over the course of the year. Project  
22 staff in each location were asked at the start of our research to identify twenty participants,  
23 broadly divided between 'fast' and 'slow' climbers in relation to the progress they had made  
24 so far. These women were interviewed every two months over the period of our research.  
25 The interviews were loosely structured around their early life histories, the processes through  
26 which they came into the project, their experiences of the project and the changes it had  
27 brought about.  
28  
29

30 In the final interview, participants were asked to place themselves on a 'graduation ladder' to  
31 illustrate how they thought they had fared over the life of the project. The bottom rung  
32 represented a state of extreme hardship, while the seventh represented security and wellbeing.  
33 Each participant was asked which rung of the ladder her family had occupied at the start of  
34 the project and where they were currently positioned.  
35

36 The results of the exercise showed that only nine out of 20 of the participants in Sindh  
37 reported improvements in their situation, including the seven identified as 'fast climbers' by  
38 project staff, while the rest reported either unchanged or worsened conditions. By contrast,  
39 16 of the 20 participants in the West Bengal pilot reported improvements in their situation,  
40 including some who had been classified by project staff as 'slow', while four reported  
41 unchanged positions or did not respond. These are, of course, highly subjective assessments  
42 on the part of participants. They cannot provide an objective measure of the extent to which  
43 change happened or failed to happen. Their usefulness lies instead in the discussions they  
44 generated about participants' assessments about their own trajectories and what they revealed  
45 about their views of progress.  
46  
47

### 48 **3.3 The qualitative evaluation of the TUP pilot in Sindh**

#### 49 **The project context in Sindh**

50  
51 The Sindh study was located in four hamlets near link roads to the highway connecting major  
52 urban centres. The nearest market was located five to ten kilometres from the hamlets. The  
53 state was largely absent from these villages. None of the hamlets had electricity, water supply  
54 or functioning schools nor were there any anti-poverty interventions by either state or NGOs  
55  
56  
57

1  
2  
3 aside from the zakat programme administered by village councils. Costs of transport to the  
4 nearest public health facility made it virtually out of reach for most villagers who relied on  
5 unqualified local practitioners.  
6

7 The study hamlets were socially homogeneous, each consisting of around ten to twelve  
8 related families of Balochi origin whose ancestors had migrated to coastal Sindh in search of  
9 employment. Social relations were organised around the *biraderi*, an endogamous entity based  
10 on cross cousin marriage, making for strong cohesive kinship networks. Each *biraderi* had a  
11 head who was generally better off than the rest, often owning land and livestock and  
12 providing employment and patronage to poorer members.  
13

14 The aridity of the soil meant that it was only cultivable for a very short period after the rains.  
15 Those who owned or leased in land grew vegetables which they either consumed themselves,  
16 saving on expenditure, or sold to local traders. The rains also led to the growth of natural  
17 fodder for livestock for a limited period of time but fodder had to be purchased the rest of  
18 the year. Other occupations for men included cutting wood, breaking stones and fishing along  
19 the coast. Some migrated illegally to work on fishing boats in Iran.  
20

21 High levels of fertility together with strong norms of female seclusion restricted women's  
22 mobility in the public domain, confining them mainly to their hamlets. There was no market  
23 for female wage labour in the study villages. Women engaged in productive activities within  
24 or near the home, mostly basket weaving but also livestock rearing. Some also helped male  
25 family members on the land for the brief period when cultivation was possible. Female  
26 earnings were generally very low and irregular, placing considerable pressure on men's earning  
27 capacity to support their households.  
28  
29

### 30 **Project implementation in Sindh**

  
31

32 OCT defined the goal of the TUP pilot as graduating participants into its microfinance  
33 programme or encouraging them to save formally through bank accounts set up in their name.  
34 Possibly due to the fact that OCT had very little experience in rural areas – it normally engaged  
35 in providing microfinance to established male entrepreneurs in urban contexts (Zaidi et al., 2007)  
36 - its management took a number of decisions which seriously undermined its ability to achieve  
37 these goals. The first related to its selection process. OCT was clearly one of the organizations  
38 which had misunderstood the requirements of randomization.  
39  
40

41 The hamlets were selected by senior management while the field staff were asked to identify  
42 eligible households on the basis of consultations with *biraderi* heads. They later found out that the  
43 heads frequently selected their closest kin or client families, advising them to conceal their true  
44 economic condition from project staff. Three of the 10 women originally classified as 'fast'  
45 climbers by the project were subsequently reclassified as 'slow' when it became clear that  
46 assumed improvements in their conditions reflected pre-existing assets.  
47

48 The main assets transferred were goats and hens along with support for women's basket  
49 weaving efforts. On the advice of a 'livestock expert', OCT decided to transfer three female  
50 'Bengal goats', imported from India, to each household because they were considerably  
51 cheaper than the local variety. Because the expert advised against mating between the  
52 imported goats and the local variety, each participant was to receive at least one inseminated  
53 goat so they could expand their flock. In fact, not only did very few participants receive  
54 inseminated goats, but the majority of goats, ill-suited to the semi-aridity of local conditions,  
55  
56  
57

1  
2  
3 died within a few months of the transfer. In addition, the purchase of a virus-infected hen in  
4 one of the poorer villages wiped out its entire flock.  
5

6 There was uneven implementation and experience with other components of the TUP  
7 approach. OCT decided to dispense with Village Assistance Committees for fear of undue  
8 influence by village elites, ironic in the light of its selection procedure. The idea of setting up  
9 formal bank accounts in participants' names foundered because of the distance and  
10 procedures involved. Most preferred to save at home or in informal village savings groups.  
11 Other components of the project were appreciated by participants, such as its mobile health  
12 clinic, the opening of a primary school in collaboration with the Rotary Club and their regular  
13 interactions with project staff.  
14

## 15 **Project outcomes in Sindh**

### 16 *The fast climbers*

17  
18  
19 One of the key findings of our research in Sindh was that project gains accrued  
20 disproportionately to participants who were already better off. All seven women who were  
21 classified as fast climbers came from the two better off villages where land was more fertile,  
22 and all reported household ownership of land, varying in size from one to 11 acres. They  
23 were thus engaged in own cultivation, considered to be the most lucrative activity in the area.  
24  
25

26 The other important factor that explained the pace of progress was household labour  
27 endowments. The ratio of able-bodied earners to dependents mattered, as did the gender  
28 composition of earners, given the very low returns to female livelihoods. With the exception  
29 of Miriam who was a widow, all of the fast climbers were married. All of them, including  
30 Miriam, had one or more able-bodied male members in their family.  
31

32 The stronger 'initial conditions' of the faster-climbing households worked in their favour in a  
33 number of ways. First of all, the greater fertility of the land around their villages meant that  
34 rain-fed fodder grew in greater abundance and for a longer period. Not only did most of the  
35 households in these villages rear their own goats but many also share-reared goats belonging  
36 to others in less favoured villages. Their long-standing experience with goat rearing meant  
37 that while some were fortunate enough to receive one of the few inseminated goats  
38 distributed by the project, most of them managed to expand their flocks because they ignored  
39 project advice against mating between the imported goats and local variety; indeed they  
40 ensured that it happened.  
41  
42

43 Secondly, the fast climbers were better able to use project support for activities other than  
44 those directly supported by the project. Having adult males in the household was the main  
45 factor that enabled this. The availability of investible resources was another. An important  
46 indicator of their favourable starting positions was the fact that many of these women were  
47 able to save their consumption stipends because they could cover their daily needs from  
48 existing income sources. Most chose to use their savings and enterprise profits to purchase  
49 agricultural inputs such as traction, fertilizer and pesticide outright rather than on credit from  
50 buying agents from the urban centres. This had not only reduced their share of the proceeds  
51 from selling vegetables but also tied them to the buying agent rather than allowing them to  
52 search for the best prices.  
53

54 A number of participants used project support to diversify into new activities. Two were  
55 encouraged by project staff to set up grocery shops in their home because they were seen to  
56  
57

1  
2  
3 have male adult family members who could help purchase stock and keep accounts. Two  
4 others were advised by their husbands, who were woodcutters, to request assistance to set up  
5 a wood shop in one case and purchase a donkey cart in the other.  
6

7 The fast climbers described themselves as 'poor but not destitute': their children did not go to  
8 bed hungry, they did not have to beg others for money, they had land but had to buy inputs  
9 on credit. The project clearly strengthened entrepreneurial capabilities within their  
10 households, in some cases, those of the participant herself, in other cases, those of a male  
11 member. This in turn led to other valued achievements ranging from improvements in the  
12 quality of their everyday lives, such as being able to afford milk with their tea, to longer-term  
13 goals such as saving for emergencies and reduced dependence on others in times of crisis.  
14

### 15 *The slow climbers*

16  
17 The slow climbers in our study, with the exception of two, came from the poorer villages.  
18 Only three of them reported land ownership although some leased land in for cultivation.  
19 They had high dependency ratios which reflected not only the presence of children too young  
20 to work but also ill or elderly family members and the lack of able-bodied adult male labour.  
21 These factors, which defined them as poorer than others, also prevented them from taking  
22 full advantage of project support.  
23

24  
25 The stories of the slow climbers were variations on themes of dependency ratios, deaths of  
26 livestock and poultry and ill health in the family. The importance of human resources in  
27 differentiating progress was exemplified by the case of two slow climbers from the better-off  
28 villages. Both were widows and both stated that they started out on the second rung of the  
29 ladder. But while one reported that she had descended to the bottom rung, the other  
30 believed she had remained on the second step. The dependency ratio in their households at  
31 the time of their husbands' deaths played a role in explaining this difference.  
32

33 Bujra's husband had died recently while her three children were very young. She now lived  
34 with her twelve-year-old son and ten-year-old daughter. Her older married son had his own  
35 family to support. Her request for a grocery shop was turned down as she did not have an  
36 adult male in the household to help her. Her health had been deteriorating steadily, she could  
37 barely weave any baskets and her stipend had been spent on medicine. Her surviving goats  
38 had not reproduced and the costs of purchasing fodder merely added to her financial burden.  
39

40  
41 Farhat, on the other hand, had eleven children, but they had been much older when her  
42 husband died. She now lived with two sons, one at school and one working in stone breaking,  
43 and two daughters who wove baskets with her. She used her stipend to purchase a sewing  
44 machine which allowed her two daughters an additional source of income. One of her goats  
45 had produced a kid and while nine of her twelve hens died, the rest produced chicks. She did  
46 not think she had progressed through the project, but did believe that it had acted as a safety  
47 net, preventing her descent to a lower rung.  
48

49 Many of slow climbers were of the view that the project had failed to tackle what they  
50 considered to be the main barriers to progress. One was the dearth of health facilities. The  
51 mobile health clinic could help with minor ailments but could do little about serious health  
52 problems, leaving women like Bujra in a state of continued dependence.  
53

54 Others spoke of the restrictions on women's capacity to earn and believed that project should  
55 have prioritized assistance to men. As Neelam argued *A woman's income just supports the household*  
56  
57

1  
2  
3 *-you cannot run the household on a woman's income alone. And when a woman has to do all the household*  
4 *work and care for children, this makes it difficult for her to make enough baskets to feed her children....Men*  
5 *can engage in various activities: they can cut wood, work in orchards, they can work as drivers or conductors.*  
6  
7

### 8 **3.4 The qualitative evaluation of the TUP pilot in West Bengal**

9

10 The West Bengal evaluation was spread over 5 villages of varying sizes located in the low-lying  
11 saline marshland areas of the South 24 Parganas. The village populations were a mixture of social  
12 groups recognized by the state as historically disadvantaged: previously 'untouchable' *dalit* castes,  
13 officially classified as Scheduled Castes (SC), *adivasis* (tribal groups) officially classified as  
14 Scheduled Tribes (ST) and deprived sections of the Muslim population officially classified as  
15 Other Backward Classes (OBC). These are not only among the poorest groups in West Bengal  
16 (and India more generally) but also face severe discrimination.  
17

18 Social identity differentiated the experiences of the extreme poor in these villages in ways that  
19 proved relevant to project outcomes. It differentiated gender relations within households,  
20 particularly women's mobility in the public domain and hence their capacity to engage in work  
21 outside the home. Restrictions on mobility were particularly severe for women from Muslim  
22 households.  
23

24 Social identity also differentiated the ability of groups to access economic opportunities and  
25 political patronage in the wider community. As we noted, West Bengal has a higher percentage  
26 of Muslims than other Indian states. This makes them an important voting bloc and wooed by  
27 politicians, despite their minority status. Adivasi groups, on the other hand, are numerically weak  
28 (5%), the poorest of various social groups and exercise little political clout.  
29  
30

31 Despite their shared marginalized status, there was considerable tension between these group,  
32 which was exacerbated in some cases by jealousy and resentment on the part of neighbours who  
33 had not been selected for the project. This led to frequent quarrels over alleged encroachments  
34 by project livestock as well as many incidents of theft and poisoning of fish and livestock.  
35

36 The state was very active in the study villages. The Left Front Government had been in power  
37 between 1977 and 2011. Its land reform programme had benefited some of the TUP participants  
38 although the plots they obtained were tiny and often uncultivable because of high levels of  
39 salinity. Other examples of state provision for the poor included public works programmes,  
40 poverty lending through rural banks to NGO or state organized SHGs; subsidized wheat, rice  
41 and kerosene and housing allowances. Government-run primary health clinics provided  
42 treatment free of cost but were 3-5 km away from TUP households who mainly relied on private  
43 practitioners. Although villages had government-run primary and secondary schools, few  
44 participants sent their children to school.  
45  
46

47 A diverse range of livelihood options was evident for men but also for women. It included  
48 farming on own or rented land, fish cultivation in ponds, daily wage labour of various kinds,  
49 including government public works, and a range of trading activities. In addition, well-developed  
50 transport systems meant that both men and women could commute daily or migrate periodically  
51 to Kolkata and its environs in search of work.  
52

### 53 **Project implementation in West Bengal**

54  
55  
56  
57

Trickle Up defined its goal as 'graduating' TUP participants into self-administered SHGs linked to government lending or into local microfinance organizations. It adopted the main components of the TUP approach with some adaptation to the local context. The organization had previously focused on cash transfers to poor women using the self-help group (SHG) approach. The TUP pilot in West Bengal allowed it to test the SHG approach with asset transfers, mainly sheep and goats. The organization also experimented with, but abandoned, the idea of forming committees of village elites when it found that they became prone to partisan politics.

The TUP pilot in West Bengal closely followed BRAC's targeting methodology to identify the extreme poor. Its effectiveness was borne out by a mid-term process evaluation which estimated that 73% of TUP members lived below \$1 a day with 45% earning half of this. (Huda, 2009). Its SHGs met on a weekly basis under a rotating leadership, saved regularly and used their pooled savings as a source of loans. The groups were used to teach financial skills and build social networks of their members. The project also provided various forms of training relating to livestock raising, fish farming and cultivation. Project staff interacted regularly with participants on an individual basis and at SHG meetings to discuss their problems, encourage them to save and to diversify their livelihoods and offering advice on family matters.

The main assets transferred were goats, sheep and poultry, but lack of familiarity with the area meant that staff were not aware that it was prone to water logging and fluke worm during the monsoons. This led to considerable loss of livestock and poultry early in the project. It became clear to project staff that many participants had no experience with livestock and that some needed more immediate sources of income. Trickle Up management re-allocated livestock from the poorer to better performers while supporting the poorer performers to take up some form of small trade, such as to vegetable vending, fish cultivation and paddy husking, which would allow them to start earning straightaway.

### **Project outcomes in West Bengal**

Variations in the initial material conditions of households did not emerge as the most important explanation for variations in the pace of progress. More significant were variations in their human resource endowments. Given declining fertility rates in West Bengal and the considerable female mobility in the public domain, particularly among poor households, dependency burdens here most often reflected illness or unemployment in the family rather than, as in Sindh, the presence of large numbers of young children or the absence of a male breadwinner.

Social identities were also relevant in differentiating project outcomes. It was one of the unexpected findings of our study that all five Adivasi women in the study were classified as fast climbers - as were Adivasi women in the rest of the project. Only two of the four *dalit* women and just four of the 11 Muslims were classified as fast climbers.

One factor common to all the women classified as fast climbers was that they had been engaged in paid work outside the home before the project started, with Adivasi women engaged in such work '*almost from birth*'. These women were accustomed to supporting their families and managing livelihood options. What distinguished Adivasi women further from the rest of these women was their attitude towards the project. Their community had been systematically by-passed by all previous development efforts. They subsisted on what they could earn on a daily basis. If they didn't earn, they bought food on credit and if they were

1  
2  
3 denied credit, *'we simply bound our stomachs and slept...Nobody helped us before. Dada (the project*  
4 *officer) was the first to give us support'*.

5  
6 As a result, Adivasi women embraced TUP as a once-in-a-lifetime opportunity. More than  
7 other participants, these women valued not only the assets they were given but project  
8 training and advice which they regarded as a substantive resource in its own right, offering  
9 new knowledge and skills and new ways to think about their livelihoods. Supti Sardar's  
10 account of her experience provides a text book case study of the faithful adherence to project  
11 advice demonstrated by these women.  
12

13 Before the project, her husband used to migrate periodically to Kolkata in search of wage  
14 work while she combined crab catching with running a liquor shop from their home. She was  
15 given two goats and 10 ducks and later a doe and three goats. She saved her consumption  
16 stipend in order to lease a small pond at low rent from their landlord/patron. Her husband  
17 took a loan from his uncle to stock the pond. They pooled the proceeds from the sale of their  
18 fish and two goats to buy a pig and to lease 60 decimals of cultivable land from the same  
19 landlord. They then combined savings from her husband's income with a loan from her SHG  
20 to lease in land to cultivate paddy using the System of Rice Intensification method they had  
21 been taught. Later, as her herd of pigs grew, she gave two to her sister to share-rear. Her  
22 husband began to do wage work within their village so that he could help her to look after the  
23 livestock.  
24  
25

26 Sukti's account draws attention to another factor which proved important in explaining the  
27 pace of progress among participants which was relations of co-operation and conflict within  
28 the household. In Sukti's household, the co-operation between husband and wife in the  
29 pursuit of livelihood strategies reflected marital harmony. In Fareeda's case, co-operation was  
30 imposed by a domineering husband who took control of the livelihood asset transfers.  
31

32 Other women reported husbands who were abusive, violent and often irresponsible, not only  
33 failing to live up to their breadwinning responsibilities but draining household resources  
34 through wasteful habits, including alcohol, drugs and gambling. The women who progressed  
35 among this group were those whose husbands' behaviour improved over the course of the  
36 project, a change they attributed directly to the project. Project staff confirmed that they had  
37 taken the decision early on in the project to interact with male household members as it was  
38 clear that without their co-operation, women would find it hard to make the most of project  
39 opportunities.  
40

41 While irresponsible or antagonistic husbands could prove a major drawback to women's  
42 ability to progress, the absence of husbands did not. The fact that three of the four divorced  
43 women in our study, all from different communities, were classified as fast climbers, though  
44 none of them had adult male support, was indicative of the capacity for independent  
45 economic agency by poorer women in this context.  
46  
47

48 A final factor that both contributed to, and was symptomatic of, the progress of the fast  
49 climbers was their active participation in their SHGs. Some had been saving prior to the  
50 project, but they valued the discipline and regularity that came with SHG saving, *'saving by the*  
51 *book'*, as well as the ability to access loans at lower interest rates than charged by informal  
52 sources. The SHGs also provided a space where they discussed livelihood matters such as  
53 caring for livestock and new forms of rice cultivation as well as their personal problems and  
54 shared concerns.  
55  
56  
57



1  
2  
3 Many of the women in the fast-climbing group, the Adivasi women in particular, believed that  
4 they started out on the first rung of the graduation ladder, given their conditions of abject  
5 poverty. One of the most valued capabilities they acquired through the project related to the  
6 achievement of food security. Other valued achievements included reduction in distress  
7 migration, not having to beg money from others; receiving credit from those who had  
8 previously refused them; reduction in domestic violence; and a greater sense of confidence  
9 and 'mental strength'.

10  
11 SHG loans were clearly an important mechanism through which some of these material  
12 impacts were realized. In addition, SHGs had come to represent a valued new set of  
13 relationships for some of the women with impacts on their consciousness, agency and  
14 engagement in collective action. Majida described how some of these changes came about: *We*  
15 *have met together, talked to each other and slowly we have developed...this is not knowledge you gain by*  
16 *reading books, but by talking to people, meeting with people.*

### 19 ***Slow climbers***

20  
21 While engagement in paid work outside the home prior to the project was common to the  
22 fast-climbing women in our study, the slow-climbing women, with the exception of two, all  
23 reported restrictions on their ability to engage in such work. The two exceptions were both  
24 OBC Muslims, both widows and both engaged in begging, one in her own village, the other  
25 commuting to Kolkata. Neither had any prior experience with livestock rearing. Neither was  
26 inclined to abandon begging in favour of petty trade, as advised by the project. One said she  
27 could not count and would not be able to keep track of those who bought her goods on  
28 credit, the other felt that she earned more from begging.

29  
30 For the rest, restrictions on their movements clearly undermined their capacity for  
31 independent economic agency. In the case of two *dalit* women, restrictions were imposed by  
32 violent and jealous husbands. One objected to his wife stepping out of the home, even to  
33 purchase food. The other had turned hostile to the project because project staff had refused  
34 his wife's request for assistance for his paddy husking business because of his continued  
35 alcoholism. She saved secretly with the SHG but was afraid to borrow for investment  
36 purposes in case he sabotaged her efforts.

37  
38 The five remaining women in this category were OBC Muslims. These women were subject  
39 to religious restrictions on their mobility, sometimes internalized by them, sometimes  
40 imposed by family members. In Samira's case, she had been brought up in a very conservative  
41 family whose women stayed at home and both she and her husband preferred that she did the  
42 same. She saved very irregularly, believed that it was un-Islamic to pay interest on loans and  
43 rejected project advice to set up her own paddy husking business because she was fearful  
44 about moving outside the house.

45  
46 Raima Bibi's husband had reduced his days of work since the project began. Given  
47 restrictions on her mobility, the only way that she could make up the shortfall was through  
48 poorly paid activities that could be done at home. Her comments support the observation we  
49 made earlier about the importance of intra-household co-operation: *I could not move up faster as*  
50 *we had differences amongst us. If in any family there are differences between husband and wife then can they*  
51 *improve?*

1  
2  
3 Notwithstanding the project's lukewarm assessment of their progress, around half of those  
4 classified as slow climbers reported some improvement in their situation, particularly those  
5 who had started out barely able to meet their survival needs. The ability of save emerged as an  
6 important source of security and they valued the discipline of SHG 'savings by the book'  
7 which protected their saving from their own need or temptation to spend as well as from but  
8 also from appropriation by husbands. As Kamala put it: *Savings give us hope and courage ....The*  
9 *book has been my major benefit. Whether we have food or not, at least we have savings.*  
10

#### 11 **4. Discussion of findings**

#### 12

13  
14 This paper set out to assess the contributions of alternative approaches to impact assessment. It  
15 contrasted the model of causality associated with RCTs with that associated with mechanism-  
16 based qualitative evaluations, drawing on examples of these approaches from studies of TUP  
17 pilots in Sindh and West Bengal. It explored two of the critiques commonly levelled at RCTs: the  
18 real-life challenges of satisfying the idealization assumption which underpins their model of  
19 causal inference and their empirically-thin account of causality.  
20

21 The main limitation of qualitative evaluations is that their study samples are too small to be  
22 regarded as statistically representative of the relevant population. Their main strength is that  
23 they provide detailed insights into the causal processes that explain observed patterns of  
24 outcomes, insights that we would argue are analytically generalizable and can shed light on  
25 findings reported by quantitative studies.  
26

27 In this section, we bring the findings of the two sets of evaluations together to ask what they  
28 contribute to our understanding of TUP impacts in the two contexts studied, to explain  
29 convergences and divergences in the findings reported by the two sets of studies and to draw  
30 out the implications of our analysis for evaluation research more generally.  
31

#### 32 **4.1 Randomization failure and heterogeneity of impacts: the RCT studies**

#### 33

34  
35 Our analysis of the RCTs found that the randomization process departed from ideal RCT  
36 conditions in both contexts. In both contexts, the failure could be attributed to human  
37 agency. In the West Bengal study, randomization was undermined by non-compliance on the  
38 part of a sizeable percentage of 'treatment' households. Not only did this pose a threat to the  
39 external validity of the project, but it also gave rise to a 'treated' group that was skewed  
40 towards the better-off religious majority in the area.  
41

42 Departure from the idealisation assumption in the Sindh RCT reflected the agency of the staff  
43 of implementing agencies who failed to comply with the agreed randomization procedure.  
44 The result here was a treatment group that was heavily biased towards households above the  
45 poverty line, with no guarantee that relevant characteristics were identical at baseline for both  
46 treatment and control households. These failures raise questions about whether the Sindh  
47 pilot qualified as an RCT at all.  
48

49  
50 Our analysis of the RCT studies pointed to another of their limitations. As we noted,  
51 provided that the idealization assumption is sufficiently satisfied, RCTs assume that average  
52 treatment effects establish conclusively whether an intervention has worked or not, absolving  
53 them of the need to explore how or for whom these effects materialized. Both the Sindh and  
54 West Bengal RCTs were deemed to have worked because they reported positive average  
55 treatment effects. But in fact, these effects were stronger for the economic indicators than the  
56 non-economic ones, suggesting that some components of the TUP approach worked better  
57

1  
2  
3 than others. In addition, West Bengal reported more, and more sustained, impacts than  
4 Sindh. Either the intervention was implemented more effectively in West Bengal than Sindh or  
5 the West Bengal context was more conducive to the TUP approach than Sindh – or both.  
6 However the ability to distinguish between these different explanations for these patterns of  
7 outcomes requires background information of a kind that is rarely collected by RCTs.  
8

#### 9 **4.2 Qualitative evaluations and the heterogeneity of impacts**

10  
11 Our qualitative evaluation collected precisely this kind of information and hence was better  
12 placed to explain variations in the impacts that it observed. Like the RCT, it also found  
13 stronger evidence of impacts for the West Bengal pilot than the Sindh one. It reported that,  
14 compared to Sindh, the West Bengal context was characterised by a more dense structure of  
15 livelihood opportunities, generated by both market and government support. It also had  
16 better transport and infrastructure. That the scope for livelihood diversification was greater in  
17 West Bengal than Sindh is borne out by both the RCTs and the qualitative evaluation.  
18

19  
20 The qualitative evaluation also noted gender-related variations in the pattern of outcomes  
21 reported by the two contexts. Its analysis suggests that differences in the patriarchal  
22 constraints which characterised the two contexts made it far more difficult for women in  
23 Sindh to translate project support into livelihood capabilities than those in West Bengal. As a  
24 result, women in our Sindh study relied to a greater extent on the presence of an adult male to  
25 convert project assets into entrepreneurial capabilities whereas in West Bengal, we found a  
26 number of examples of women who were able to benefit from project support without male  
27 assistance. As we noted, the RCTs do not provide any information on the gender distribution  
28 of impacts.  
29

30  
31 Within the socially heterogeneous context of West Bengal, the qualitative evaluation found  
32 that the intersection of gender with other aspects of social identity differentiated the nature of  
33 patriarchal constraints for different social groups and hence the outcomes reported by  
34 women from these different groups. It found that Adivasi women in particular were expected  
35 to start earning a living from a very early age, working in the fields or rivers 'almost from  
36 birth'. By contrast, Muslim women in the study were subject to greater social restrictions so  
37 that even those who welcomed the opportunities offered by the project were curtailed in their  
38 capacity to translate them into valued capabilities.  
39

40  
41 Our qualitative evaluations also suggested that social composition of beneficiary households  
42 in the two contexts gave rise to 'real-world' violations of the idealisation assumption with  
43 implications, often hidden, for the outcomes reported. In Sindh, the OCT pilot was carried  
44 out with kinship-based communities drawn from the same ethnic group and living together  
45 for generations. By contrast, in West Bengal, Trickle Up worked with different socially  
46 marginalized communities who lived in a state of considerable tension with each other. We  
47 found examples of actions stemming from community divisions and solidarities in each  
48 context that affected the size and distribution of impacts. Resentment towards project  
49 participants in West Bengal by neighbours from other communities who had not been  
50 selected often resulted in the destruction of their assets, weakening the impact of the project  
51 for some of the participants. Equally, however, there were examples of positive spill-overs  
52 from both pilots as women from the participating groups shared the information and  
53 knowledge they had acquired with others - from their own community groups in the West  
54 Bengal and from their kinship networks in the Sindh context. While these would be recorded  
55 in qualitative evaluations as aspects of the way that the projects play out in different  
56  
57  
58  
59  
60

1  
2  
3 communities, they would have constituted violence of the non-interference principle in RCTs,  
4 biasing their results in unknown ways.

#### 5 6 **4.2 Explaining impacts across the two methodologies: the West Bengal** 7 **evaluations**

8  
9 Reading across the two sets of evaluations within each context, we find examples of  
10 convergences and divergences in their overall conclusions as well the distribution of impacts  
11 they reported in each context. We can use the information discussed in the paper to try and  
12 explain these. The incomplete nature of this information means that the explanations are  
13 necessarily speculative but they draw attention to some of the relevant background  
14 information that could have strengthened our understanding of how these pilots performed.

15  
16 First of all, we found that both the RCT and the qualitative evaluations of the pilots in West  
17 Bengal reported positive outcomes. We have already attributed this to the dense structure of  
18 market opportunities backed up by government support evident in this context. As we noted,  
19 both the RCT and the qualitative evaluation found strong evidence of livelihood  
20 diversification in the West Bengal context. But whereas the RCT found that it was the better-  
21 off participants who progressed fastest, the qualitative study reported the counter-intuitive  
22 finding that it was the poorest group who made the greatest progress.

23  
24 The discussion in this paper suggests a number of explanations for this. One explanation for  
25 the RCT finding relates to the randomization failure noted earlier. As we noted, patterns of  
26 non-compliance and attrition in the RCT study meant that it was largely better-off, mainly  
27 Hindu, households among the treatment group who received project support while a  
28 substantial percentage of poorer Muslims not only refused or returned project assets, they  
29 were also missing from the end line surveys. They would have registered zero impact in ITT  
30 estimates of average treatment effect. It is consequently not surprising that it was better-off  
31 households that reported greater impacts; indeed, it is likely that they were driving much of  
32 the impact.

33  
34 The qualitative evaluation did not find any examples of outright non-compliance but it did  
35 find considerable variation in the zeal with which selected beneficiaries participated. We  
36 explained the greater progress reported by the Adivasi women in terms of the greater zeal  
37 with which members of the poorest and historically most excluded group in West Bengal (and  
38 much of India) embraced what they saw as a once-in-a-lifetime opportunity as well as the  
39 experience they had accumulating in managing household livelihoods.

40  
41 In addition I would like to suggest that divergence in the distribution of outcomes reported  
42 by the two studies may also have reflected differences in the implementation of the pilots.  
43 The RCT pilot was implemented by Bandhan, a microfinance organization that uses the joint  
44 liability group model pioneered by Grameen Bank for its lending activities. For the purposes  
45 of the pilot, it excluded participants from any microfinance or self-help groups since these  
46 were likely to be associated with government or NGO anti-poverty schemes. While project  
47 staff met with participants on a regular basis, they did not make any attempt to promote their  
48 collective organization.

49  
50 Trickle-Up, on the other hand, had long experience with self-help groups which have been  
51 found to be generally more effective than joint-liability groups at reaching women from the  
52 poorest castes and tribal groups (EDA Rural Systems, 2004; Sarma and Mehta, 2014). It chose  
53 to integrate self-help groups into the design of its pilot. The qualitative evaluation suggested

1  
2  
3 that this was a particular valued aspect of the project and that it was of particular value to  
4 Adivasi groups.  
5  
6

#### 7 **4.4 Explaining impacts across the two methodologies: the Sindh evaluations**

8

9 Turning to the Sindh studies, both the RCT and the qualitative evaluation agreed that the  
10 better-off among project participants reported greater impacts than poorer ones, but they  
11 diverged considerably in their overall assessment: while the RCT pilot in Sindh was deemed to  
12 have 'worked', along with other RCT pilots, we found that the majority of the participants in  
13 the evaluation pilot reported little or no progress and the pilot was subsequently closed down.  
14

15 Once again, there are a number of possible strands to the explanation. One obvious strand  
16 related to project implementation. OCT made some glaring mistakes in the selection of  
17 livelihood assets and advice provided to project participants which may well have reflected its  
18 lack of experience working with poor rural women. No information is provided on the  
19 implementation process of the four NGOs that featured in the RCT study but we must  
20 assume that they performed sufficiently well for participating households to report positive  
21 average treatment effects.  
22

23  
24 However, differences in the beneficiary groups and contexts covered by the two studies also  
25 needs to be taken into account. Our qualitative evaluation chose to focus on the poorest rural  
26 locations covered by the OCT pilot. These were characterised by a semi-arid environment,  
27 limited infrastructure and health facilities, the near-absence of government interventions of  
28 any kind and a dearth of local livelihood opportunities for men and even fewer for women.  
29 Given this context, it is not surprising that those who started out in stronger positions were  
30 better able to avail themselves of project opportunities. Poorer participants were hampered  
31 by their lack of assets and human resources and the debilitating effects of ill-health.  
32

33 By contrast, we have already noted that around 90% of the RCT participants were estimated  
34 to be above the poverty line. In addition, the IDS study cited earlier contained data on  
35 housing conditions that suggests that both treatment *and* control households came from more  
36 developed, probably peri-urban contexts. It reported that between 34% and 44% of  
37 participating and control households in its study had flush toilets connected to public  
38 sewerage; between 45% and 41% had electricity; between 21% and 23% had piped water; and  
39 between 7 and 15% cooked with gas. This is in marked contrast to the hamlets in our  
40 qualitative study which had no electricity, no flush toilets, no piped water so that households  
41 either purchased water to store in tanks or fetched it from neighbouring wells and finally,  
42 mainly cooked with firewood. In short, while both the RCT and the qualitative study suggest  
43 that better-off participants reported more favourable outcomes than poorer ones, the fact  
44 that the RCT participants were generally better off than those in the qualitative study may  
45 explain the more favourable overall conclusion of the RCT study.  
46  
47

#### 48 **4.5 Concluding comments**

49

50 This paper has used the empirical case studies of RCTs and qualitative evaluations of pilots to  
51 transfer assets to women in extreme poverty in to explore in greater detail what they contribute  
52 to our understanding of how development interventions work. In this concluding section, I  
53 would like to draw on the broader evaluation literature to make some general points about these  
54 alternative approaches and draw out some general implications for policy.  
55  
56  
57

1  
2  
3 First of all, the threats to the randomization process noted in relation to the RCTs discussed in  
4 this paper have been noted in other studies as well. In some cases, RCTs seek to deal with these  
5 problems through various *ex post* 'econometric or statistical fixes' (Deaton, 2009). As Deaton  
6 suggests, these are acceptable solutions, but they carry the danger of 'data mining', trying out  
7 different control variables until the experiment can be shown to have 'worked'. In any case, they  
8 take RCTs out of the world of the idealization assumptions and back into the world of everyday  
9 econometrics and statistics. RCTs then become part of the wider array of quantitative  
10 methodologies in which they may sometimes have the edge in terms of rigor over other methods  
11 and sometimes not (Deaton and Cartwright).  
12

13 In other cases, real life threats can be avoided by better design of the trial. For instance, causally  
14 relevant variables could be built into the design of the trial through the *ex ante* stratification of  
15 the randomization process. However, RCTs frequently do not collect information on these such  
16 variables because they do not consider these relevant to their experiment or even know what  
17 they might be. Yet theoretical analysis of problems and prior knowledge of contexts could play a  
18 valuable role in helping with the pre-identification of relevant variables as well as offering  
19 plausible explanations for observed patterns of outcomes.  
20

21 This takes us to the more general problem posed by the thin empirical base on which the RCT  
22 model of causal inference rests. As critics have pointed out, the exaggerated claims made by RCT  
23 advocates for the rigour of their methodology has led them to delegitimize 'other ways of  
24 knowing' and the 'virtually ritual denigration' of knowledge gained through these other means  
25 (Bedcarrat). Yet a great deal of the existing knowledge relevant to improving the design of RCTs  
26 and interpreting their findings has not been, and cannot be, collected through RCTs (Deaton and  
27 Cartwright, 2018). This includes knowledge of the local contexts in which trials are set and of  
28 socio-economic differentials among the trial population that are likely to be causally relevant.  
29  
30

31 The explanatory power of RCTs could also be improved through the collection of detailed  
32 qualitative information on how the intervention being trialled actually works in particular  
33 contexts. Such information would help to triangulate and elaborate on quantitative findings,  
34 placing them on firmer grounds. It would also help to tease out causal mechanisms that elude  
35 quantitative methods and to interpret unexpected findings that quantitative methods can seldom  
36 do.  
37

38 It is worth noting, for instance, that although the synthesis report of the six RCTs found  
39 evidence of positive impacts for all six, it also noted that these impacts were 'not very large',  
40 leading to it to raise theoretical questions about the nature of the poverty trap that the projects  
41 were intended to address. Yet before jumping to these higher level reflections, the material  
42 discussed in this paper points to more mundane, lower-level findings from Sindh and West  
43 Bengal that could contribute to the explanation: these failure to take up the treatment (unique  
44 apparently to West Bengal), failure to randomize with the possibility that control households  
45 started out better off on certain characteristics than treatment ones, project-induced  
46 misbehaviour on the part of non-participants leading to destruction and thefts of project assets;  
47 positive spill-over effects from participants to control households; differences in project  
48 implementation strategies; differences in the agro-climatic environment, infrastructure and  
49 market conditions between treatment and beneficiary households; and variations in impact by  
50 different sub-groups which could dilute or cancel out average impacts.  
51  
52

53 Finally, from a policy perspective, a 'thicker' understanding of what makes a project work or fail  
54 to work or work very partially is essential for a number of reasons. At the practical level, it will  
55 help to redesign interventions to adapt them better to different contexts. At a political level, it  
56  
57

can provide insights into the distribution of the effects of an intervention, why certain sub-groups failed to benefit or proved non-compliant while others made unexpected gains or monopolized the benefits. Such an understanding is critical if interventions are to address, rather than reproduce, inequalities that are present, even among those classified as very poor. It will have important implications for the political economy of scaling up an intervention or extending it to other contexts.

## References

- Banerjee, A., E. Duflo, R. Chattopadhyay and J. Shapiro (2011) *Targeting the hard-core poor: an impact assessment* (<http://www.povertyactionlab.org/publication/targeting-hard-core-poor-impact-assessment>) Downloaded November 2017
- Banerjee, A., E. Duflo, N. Goldberg, D. Karlan, R. Osei, W. Pariente, J. Shapiro, B. Thuysbaert, and C. Udry. (2015) 'A Multifaceted Program Causes Lasting Progress for the Very Poor: Evidence from Six Countries', *Science* 348 (6236): 1260799–1260799
- Barrett, C.B. and M.R. Carter (2010) 'The power and pitfalls of experiments in development economics: some non-random reflections'. *Applied Economic Perspectives and Policy* Vol (32); 4: 515–548.
- Deaton, A. (2010) 'Instruments, randomization and learning about development' *Journal of Economic Literature* Vol. 48 : 424-455
- Deaton, A. and N. Cartwright (2018) 'Understanding and misunderstanding randomized control trials' *Social Science and Medicine* Vol. 210 (2-21).
- EDA Rural Systems (2004) *The Maturing of Indian Microfinance: A Longitudinal Study*. Report prepared for SIDBI Foundation for Micro-credit, New Delhi
- Huda, K. (2009) *Process Evaluation of Trickle Up's Targeting Ultra Poor Programme*. BRAC Development Institute Report. Dhaka
- IDS (2012) *Assessment Survey: PPAF's Social Safety Net Targeting Ultra Poor (TUP) Program* (<http://www.ppaf.org.pk/Research/TUP.pdf>. Downloaded November 2017)
- IPA (2014) *Graduating the ultra-poor in Pakistan* Innovations for Poverty Action. (<http://www.poverty-action.org/study/graduating-ultra-poor-pakistan>). Downloaded November 2015
- Kabeer, N., K. Huda, S. Kaur and N. Lamhauge (2012) *Productive safety nets for women in extreme poverty: lessons from pilot projects in India and Pakistan* CDPR Working Paper 28/12. School of Oriental and African Studies, London
- Karlan, D. and Parienté, W. (2014) *Impact of 'Targeting the Ultra-Poor' Program in Pakistan* Research Partnerships on Participatory Development Brief 9. PPAF: Islamabad. (<http://www.ppaf.org.pk/PPAF-Conference-2014/Publications/paper9.pdf>). Downloaded April 2015.
- Kidd, S. and D. Bailey-Athias (2017) 'The effectiveness of the Graduation Approach: what does the evidence tell us?' *Policy in Focus* Vol 14(2): 22-28.
- Krauss, A. (2018) 'Why all randomized controlled trials produce biased results' *Annals of Medicine* Vol 50(4): 312-332
- Matin, I., M.Sulaiman and M.Rabbani (2008) *Crafting a graduation pathway for the ultra poor: lessons and evidence from a BRAC programme* CPRC Working Paper No. 109. Manchester: Manchester University
- Mohmand, S.R. and Gazdar, H. (2007) *Social structures in Pakistan* Islamabad: Asian Development Bank
- Pawson. R. and N. Tilly (2004) *Realistic evaluation* London: Sage Publications

- 1  
2  
3 Robeyns, I. (2005) 'The capability approach. A theoretical survey' *Journal of Human*  
4 *Development* Vol. 6(1): 93-114  
5 Sarma, S.K. and M.H.Mehta (2014) 'The best model for micro-lending: self-help group or  
6 joint liability group?' *Journal of Rural Development* Vol. 33 (3): pp. 247-260  
7 Sen, A. (1999) *Development as Freedom*, Oxford University Press, Oxford.  
8 Shaffer, P. (2011) 'Against excessive rhetoric in impact assessment: overstating the case for  
9 randomized controlled experiments' *Journal of Development Studies* Vol. 47 (11): 1619-1635  
10 White, H. (2009) 'Theory-based impact evaluation: principles and practice' *Journal of Development*  
11 *Effectiveness* Vol. 1 (3): 271-284  
12 Zaidi, S.A., Jamal, H., Javeed, S. and Zaka, S. (2007) *Social impact assessment of microfinance*  
13 *programmes* Study submitted to European Union-Pakistan Financial Services Sector Reform  
14 Programme, Islamabad  
15

---

16 <sup>1</sup> Ethiopia, Ghana, Haiti, Honduras, India, Pakistan, Peru and Yemen.

17 <sup>2</sup> The project was a collaboration with BRAC Development Institute and the Institute of Development Studies,  
18 Sussex and funded by MasterCard. The team consisted of Karishma Huda, Sandeep Kaur, Nicolina Lamhauge,  
19 Sanchari Datta, Sajjad Abro and myself. I would also like to thank, with the usual disclaimers, Professor Paul Shaffer  
20 for his extremely helpful comments on an earlier version of this paper.

21 <sup>3</sup> This assumes, among other things, that participants comply with the intervention; that its effects do not spill-  
22 over into the control group; and that there is no 'interference' with the outcomes in the trial sample by  
23 influences other than the treatment

24 <sup>4</sup> It should be pointed out that, while it was not done for the RCTs discussed in this paper, it is possible to gain  
25 some insights into project mechanisms by randomizing different aspects of project design. But this would still leave  
26 unexplained how the different aspects worked.

27 <sup>5</sup> Used by the MDGs to benchmark extreme poverty

28 <sup>6</sup> 13 households had to be dropped as they had not been randomized

29 <sup>7</sup> Over half of these had refused to be re-surveyed, others had migrated.

30 <sup>8</sup> Bonell et al (2018) make an important distinction between programmatic theories of change which focus on  
31 specific interventions and more general 'theories of problems', such as embedded in the capability approach, which  
32 seek to explain the phenomena to be addressed.  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60