



# Journal of Human Development and Capabilities

A Multi-Disciplinary Journal for People-Centered Development

ISSN: 1945-2829 (Print) 1945-2837 (Online) Journal homepage: <https://www.tandfonline.com/loi/cjhd20>

## Randomized Control Trials and Qualitative Evaluations of a Multifaceted Programme for Women in Extreme Poverty: Empirical Findings and Methodological Reflections

Naila Kabeer

To cite this article: Naila Kabeer (2019) Randomized Control Trials and Qualitative Evaluations of a Multifaceted Programme for Women in Extreme Poverty: Empirical Findings and Methodological Reflections, *Journal of Human Development and Capabilities*, 20:2, 197-217, DOI: [10.1080/19452829.2018.1536696](https://doi.org/10.1080/19452829.2018.1536696)

To link to this article: <https://doi.org/10.1080/19452829.2018.1536696>



© 2019 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group



Published online: 29 Apr 2019.



[Submit your article to this journal](#)



Article views: 5460



[View related articles](#)



[View Crossmark data](#)



Citing articles: 16 [View citing articles](#)

# Randomized Control Trials and Qualitative Evaluations of a Multifaceted Programme for Women in Extreme Poverty: Empirical Findings and Methodological Reflections

NAILA KABEER

Department of International Development, London School of Economics and Political Science, London, UK

**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.*

**KEYWORDS:** Capability approach, Gender, Measurement, Poverty, Well-being

## 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 that the extreme poor were not only poorer than other sections of the poor population, but structurally different because of the chronic, multiple, overlapping and frequently gendered nature of the constraints that kept them trapped in poverty.

---

Correspondence Address: Naila Kabeer, London School of Economics and Political Science, Houghton Street, London WC12 2AE, UK. Email: [N.Kabeer@lse.ac.uk](mailto:N.Kabeer@lse.ac.uk)

© 2019 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group  
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

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 as "end". 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. Certain "conversion factors" can block or differentiate people's ability to translate the resources at their disposal into the capabilities needed to achieve their goals. These conversion factors reflect the characteristics of individuals and their families as well as aspects of the wider physical and socio-economic context in which they are located.

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

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

Local BRAC staff carried out a combination of spatial poverty mapping, participatory wealth ranking and household surveys in order to identify households that met these criteria. As *Matin et al. (2008)* pointed out, while the triangulation process was intended to increase the accuracy of the targeting process, it was also anticipated that the social interactions involved would help to create a shared understanding of the programme's rationale and approach and to build support for it in the wider community.

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

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

Finally, intensive staff engagement was considered central for building women's individual capabilities, their sense of self-confidence and the ability to take risks and plan ahead. These cognitive and subjective qualities were likely to be absent or poorly developed among socially isolated women living close to the margins of survival in a highly

patriarchal society, but were considered essential in enabling these women to translate their assets into viable enterprises (Matin et al., 2008).

### *1.2. Objectives of the Paper*

Positive evaluations of the first phase of the programme suggested that it might offer a generalizable model for tackling extreme poverty within a time-bound period. In 2006, the Consultative Group to Assist the poor (CGAP) and Ford Foundation set up the Graduation Programme which brought together a number of organizations to adapt, pilot and evaluate the TUP approach in eight countries across the world.<sup>1</sup> Randomized control trials (RCTs) were the preferred evaluative methodology and six of these were carried out on a co-ordinated basis between 2007 and 2014 in order to test the validity of the approach across different contexts—including the state of West Bengal in India and the province of Sindh in Pakistan.

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

At the same time, the Graduation programme commissioned a team of researchers, including myself, to carry out qualitative evaluations of two TUP pilots in Sindh and West Bengal, a few hundred miles away from those studied by the RCTs.<sup>2</sup> These fared very differently from each other: at the end of the pilot phase, the project in Sindh was closed down while the one in West Bengal was scaled up (Kabeer et al., 2012).

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

## **2. Counter-factual Approaches to Impact Assessment**

### *2.1. RCTs as Impact Assessment Methodology*

RCTs were prioritized in the Graduation Programme because they are currently held to be the “gold standard” in evaluation methodologies. They use what has been described as the “counterfactual” approach to impact assessment (Shaffer, 2011), seeking to establish how participants of a programme would have fared in the absence of that programme. They do this by randomly assigning a sample of the population deemed eligible for the programme to “treatment” and “control” groups and comparing their progress on selected outcome indicators. RCT advocates claim that, provided what Cartwright calls “the idealisation assumption” is satisfied, they can provide unbiased estimates of the average impacts of the programme by measuring the difference in the mean outcomes reported by “treatment” and “control” groups, its average treatment effects. According to the idealization assumption, the randomization process will ensure that all factors likely to influence the outcome variables, aside from the treatment itself, are distributed identically between

treatment and control groups.<sup>3</sup> The satisfaction of this assumption renders it unnecessary to take account of these other factors—or even to know what they might be (Krauss, 2018). Consequently, RCTs minimize reliance on theoretical predictions or prior knowledge that might be needed otherwise to identify these factors (Barret and Carter, 2010).

However, many critics believe that the claims made for the methodological superiority of RCTs are overstated and that the methodology represents “a baser metal than gold” (Barret and Carter, p. 516). While their criticisms cover a range of issues, two are of particular relevance to this paper. The first is that it is seldom possible to replicate the strict conditions required by a pure RCT once experiments are moved from the laboratory into the messy conditions of the real world where the idealization assumption seldom holds (Shaffer, 2011; Deaton and Cartwright, 2018). As Barret and Carter (p. 521) point out, “the clean identification of randomization gets compromised by human agency”: the wilful and accidental, hidden and overt, expected and unanticipated agency of the multitude of actors that inhabit the project’s domain of operation.

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

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

## *2.2. An RCT Study of TUP Pilots in Sindh*

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

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

This process was not followed in practice. While some organizations did indeed hold public lotteries to randomly select project participants in the identified villages, others “chose to select half of the villages identified” (IDS, p. 10). Unfortunately, the IDS study does not provide any information on the percentages of households that were randomly and non-randomly assigned. What it does point out is that this failure to follow the agreed procedures meant that there were likely to be unknown variations in the background characteristics of the treatment and control groups with no guarantee that they were equally poor (IDS, 2012, p. 10).

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

Putting to one side the serious threats that this posed to both internal and external validity of the Sindh RCT, we find that the findings from the Sindh pilot reported in the synthesis study were extremely uneven, both across indicators and over time. Positive impacts which were sustained across both end lines (at the end of the programme and a year and a half later) were reported for just per capita consumption, assets and political involvement (membership of political party). The impact on the income indicators were insignificant at the first end line but impacts on livestock earnings were positive and significant at the second. Positive impacts were reported for time spent in caring for livestock, food security and women’s empowerment at the first end line but had weakened or become insignificant by the second. Negative impacts on health were reported for the first end line survey but had faded by the second while impacts on financial inclusion were insignificant in the first end line survey and negative in the second. Finally, there did not appear to be any impact on self-reported happiness (collected only in the second end line survey). There is no discussion of possible explanations for this mixed pattern of outcomes but clearly some components of the project worked better than others.

### *2.3. The RCT Study of a TUP Pilot in West Bengal*

The TUP pilot in West Bengal was implemented in rural areas of Murshidabad district by Bandhan, one of India’s largest microfinance organizations. Its goal was to enable women in extreme poverty to qualify within 18 months for regular microfinance loans from Bandhan (Banerjee et al. 2011). Following BRAC’s eligibility criteria and targeting methodology, Bandhan identified 991 eligible households for the trial sample. The base line survey of these households carried out in 2007 confirmed that they conformed to eligibility criteria. 512 (52%) of these households were then randomly selected to participate in the project, while 466 were assigned to the control group.

However, of those selected to participate in the programme, 12.5% had to be dropped on grounds of ineligibility. A further 35% refused to participate, mainly out of suspicion of project motivations. Of these non-compliant households, 58% were Muslims. In the end, only 266 households, or 52% of those originally selected for participation in the project, actually participated.

In order to check whether non-compliance on the part of those selected for the project had biased the sample, the study compared the base line characteristics of households that refused to participate with the rest of the trial sample. It found that they were very similar to the rest of the sample, except that Muslims were significantly less likely to participate than Hindus (by 18 percentage points). The study allowed for the effects of

non-compliance by carrying out the rest of its analysis on the basis of the “intention to treat” (ITT) group, all those who had been selected to participate in the project regardless of whether they participated or not, rather than those actually treated.

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

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

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

The impacts on non-economic indicators were less consistent: positive impacts on physical health, self-reported happiness and stress levels were reported at the first end line, but had disappeared by the second. Positive impacts on political involvement (voting and engagement with village leaders) emerged only in the second end line survey. There was also no evidence of impact on women’s empowerment (as measured by say in decision-making) which was measured in the first end line. The study does not comment on why the economic impacts were stronger and more sustained than the non-economic ones.

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

Finally, while the findings from two RCTs discussed so far relate to average treatment effects for ITT households, the synthesis study used quantile regression analysis to

explore the distribution of the different impacts among the treated populations. It found that while there were positive and significant impacts on income, consumption and assets for different quantiles tested, the largest impacts were found among those who started out better off than the rest. Only in the case of food security did poorer households report larger impacts. It is worth noting that, despite the stated goal of the TUP approach to promote the livelihood capabilities of poor women, the report is studiously gender-neutral in its selection of indicators, aside from the specific indicator on women's empowerment, in its discussion of impacts.

### **3. Mechanism-based Approaches to Impact Assessment**

#### *3.1. Focusing on Causal Processes*

Since RCTs are set up to ensure that the programme is the sole driver of impact, they focus on estimating the size and significance of average impacts. They do not ask further questions about how the programme worked and who it worked for. On the other hand, these questions are pivotal to alternative “mechanism-based” approaches to evaluation which are explicitly concerned with the causal processes which generate observed outcomes. Common elements in these approaches include the need for a programmatic theory of change which spells out the causal pathways through which a programme is expected to work, the relevance of contextual factors in explaining whether, for whom and under what circumstances the programme is likely to work and an analytical narrative which draws on relevant sources of information along with the findings of the evaluation to explain how the programme actually worked (Shaffer, 2011; Pawson and Tilly, 2004; White, 2009).

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

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

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

#### *3.2. The Qualitative Evaluations of TUP Pilots in Sindh and West Bengal: Note on Methodology*

The TUP pilot in Sindh was implemented by Orangi Charitable Trust (OCT) one of the five NGOs selected by the PPAF. It worked with 200 women in 11 villages between February



2008 and May 2010. In West Bengal, the pilot was implemented by Trickle Up, an international NGO, in partnership with a local NGO, Human Development Centre (HDC), which was responsible for field-level operations. It worked with 300 in 10 villages in South 24 Parganas district between August 2007 and August 2009. Our research was carried out between May 2009 and May 2010, covering the final year of the Sindh pilot and the final four months and eight months thereafter of the West Bengal pilot.

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

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

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

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

### *3.3. The Qualitative Evaluation of the TUP Pilot in Sindh*

*3.3.1. The project context in Sindh.* The Sindh study was located in four hamlets near link roads to the highway connecting major urban centres. The nearest market was located five to ten kilometres from the hamlets. The state was largely absent from these villages. None of the hamlets had electricity, water supply or functioning schools nor were there any anti-poverty interventions by either state or NGOs aside from the zakat programme administered by village councils. Costs of transport to the nearest public health facility made it virtually out of reach for most villagers who relied on unqualified local practitioners.

The study hamlets were socially homogeneous, each consisting of around ten to twelve related families of Balochi origin whose ancestors had migrated to coastal Sindh in search of employment. Social relations were organized around the *biraderi*, an endogamous entity based on cross cousin marriage, making for strong cohesive kinship networks (Mohmand and Gazdar, 2007). Each *biraderi* had a head who was generally better off than the rest,

often owning land and livestock and providing employment and patronage to poorer members.

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

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

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

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

The main assets transferred were goats and hens along with support for women's basket weaving efforts. On the advice of a “livestock expert”, OCT decided to transfer three female “Bengal goats”, imported from India, to each household because they were considerably cheaper than the local variety. Because the expert advised against mating between the imported goats and the local variety, each participant was to receive at least one inseminated goat so they could expand their flock. In fact, not only did very few participants receive inseminated goats, but the majority of goats, ill-suited to the aridity of local conditions, died within a few months of the transfer. In addition, the purchase of a virus-infected hen in one of the poorer villages wiped out its entire flock.

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

### 3.3.3. *Project outcomes in Sindh*

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

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

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

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

A number of participants used project support to diversify into new activities. Two were encouraged by project staff to set up grocery shops in their home because they were seen to have male adult family members who could help purchase stock and keep accounts. Two others were advised by their husbands, who were woodcutters, to request assistance to set up a wood shop in one case and purchase a donkey cart in the other.

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

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

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

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

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

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

Others spoke of the restrictions on women's capacity to earn and believed that project should have prioritized assistance to men. As Neelam argued, *You cannot run the household on a woman's income alone. And when a woman has to do all the household work and care for children, this makes it difficult for her to make enough baskets to feed her children ... . Men can engage in various activities: they can cut wood, work in orchards, they can work as drivers or conductors.*

#### *3.4. The Qualitative Evaluation of the TUP Pilot in West Bengal*

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

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

Social identity also differentiated the ability of groups to access economic opportunities and political patronage in the wider community. As we noted, West Bengal has a higher percentage of Muslims than other Indian states. This makes them an important voting

bloc and wooed by politicians, despite their minority status. Adivasi groups, on the other hand, are numerically weak (5%), the poorest of various social groups and exercise little political clout.

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

The state was very active in the study villages. The Left Front Government had been in power between 1977 and 2011. Its land reform programme had benefited some of the TUP participants although the plots they obtained were tiny and often uncultivable because of high levels of salinity. Other examples of state provision for the poor included public works programmes, poverty lending through rural banks to NGO or state organized self-help groups (SHGs)<sup>9</sup>; housing allowances; and subsidized wheat, rice and kerosene. Government-run primary health clinics provided treatment free of cost but were 3–5 km away from TUP households who mainly relied on private practitioners. Although villages had government-run primary and secondary schools, few participants sent their children to school.

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

*3.4.1. Project implementation in West Bengal.* Trickle Up defined its goal as “graduating” TUP participants into self-administered SHGs linked to government lending or into local microfinance organizations. 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. 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 meetings 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 offer 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 the management that many participants had no experience with livestock and that some needed more immediate sources of income. It therefore 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.

3.4.2. *Project outcomes in West Bengal.* Variations in the initial material conditions of households did not emerge in the West Bengal study 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”. Accustomed to supporting their families and managing their livelihood strategies, these women had accumulated a form of human capital over the years. What distinguished Adivasi women further from the rest of these women was their attitude towards the project. Their communities 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 denied credit, “we simply bound our stomachs and slept ... Nobody helped us before. Dada (the project officer) was the first to give us support”.

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

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

Along with illustrating how she had combined new opportunities with her past experience, Sukti’s account draws attention to another factor which proved important in explaining the pace of progress among participants: relations of co-operation and conflict within the household. In Sukti’s household, the co-operation between husband and wife in the pursuit of livelihood strategies reflected marital harmony. In Farida’s household, co-operation was imposed by a domineering husband who took control of the project asset transfers.

But many of the other women reported husbands who were abusive, violent and often irresponsible, not only failing to live up to their breadwinning responsibilities but draining household resources through wasteful habits, including alcohol, drugs and gambling. The women who progressed among this group were those whose husbands’ behaviour improved over the course of the project, a change they attributed directly to the project. Project staff

confirmed that they had taken the decision early on in the project to interact with male household members as it was clear that without their co-operation, women would find it hard to make the most of project opportunities.

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

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

In terms of outcomes, many of the women in the fast-climbing group, the Adivasi women in particular, believed that they started out on the first rung of the graduation ladder, given their earlier condition of abject poverty. One of the most valued capabilities they acquired through the project related to the achievement of food security. Some engaged in paddy husking business to buy paddy when prices were low and either consuming the rice they produced or selling it when the price was high. Others leased in land to grow paddy and ensure at least some of their subsistence for the year. Other valued achievements included reduction in distress migration, not having to beg money from others; receiving credit from those who had previously refused them; reduction in domestic violence; and a greater sense of self-confidence.

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

*3.4.2.1. Slow climbers.* The slow-climbing women in our sample, with the exception of two, all reported restrictions on their ability to engage in work outside the home. The two exceptions were both OBC Muslims, both widows and both engaged in begging, one in her own village, the other commuting to Kolkata. Neither had any prior experience with livestock rearing. Neither was inclined to abandon begging in favour of petty trade, as advised by the project. One said she could not count and would not be able to keep track of those who bought her goods on credit, the other felt that she earned more from begging.

For the rest, restrictions on their movements clearly undermined their capacity for independent economic agency. In the case of two *dalit* women, restrictions were imposed by violent and jealous husbands. One of these women continued to save secretly with her SHG but was afraid to borrow for investment purposes in case her husband sabotaged her efforts.

The five remaining women in this category were OBC Muslims. These women were subject to religious restrictions on their mobility, sometimes internalized by them, other times imposed by family members. In Samira's case, she had been brought up in a very conservative family whose women stayed at home and both she and her husband preferred that she did the same. She saved very irregularly, believed that it was un-Islamic to pay interest

on loans and rejected project advice to set up her own paddy husking business because she was fearful about moving outside the house.

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

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

#### **4. Discussion of Findings**

This paper set out to assess the contributions of alternative approaches to impact assessment. It contrasted the model of causality associated with RCTs with the mechanism-based approach taken by qualitative evaluation, drawing on examples of these approaches from studies of TUP pilots in Sindh and West Bengal. RCTs are regarded by their advocates as the gold standard for evaluation methodologies but we discussed some of the limitations highlighted by their critics.

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

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

##### *4.1. Randomization Failure and Heterogeneity of Impacts: The RCT Studies*

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

In the case of the Sindh RCT, departure from the idealization assumption reflected the agency of the staff of implementing agencies who failed to comply with the agreed randomization procedure. The result here was a treatment group that was heavily biased towards households above the poverty line, with no guarantee that causally relevant characteristics were identical at baseline for both treatment and control households. These failures raise questions about whether the Sindh pilot qualified as an RCT at all.

Our analysis of the RCT studies pointed to another of their limitations. As we noted, provided that the idealization assumption is satisfied, RCTs assume that average treatment



effects establish conclusively whether an intervention has worked or not, absolving them of the need to explore how or for whom these effects materialized. Both the Sindh and West Bengal RCTs were deemed to have worked because they reported positive average treatment effects. But in fact, these effects were stronger for the economic indicators than the non-economic ones in both contexts, suggesting that some components of the TUP approach worked better than others. In addition, the West Bengal RCT reported stronger and more sustained impacts than the Sindh one. Either the intervention was implemented more effectively in West Bengal than Sindh or the West Bengal context was more conducive to the TUP approach—or both. However the ability to distinguish between different explanations for these patterns of outcomes requires background information of a kind that is rarely collected by RCTs.

#### *4.2. Qualitative Evaluations and the Heterogeneity of Impacts*

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

The qualitative evaluation also noted gender-related variations in the pattern of outcomes reported by the two contexts. Its analysis suggests that differences in the patriarchal constraints which characterized the two contexts made it far more difficult for women in Sindh than those in West Bengal to independently convert project resources into livelihood capabilities. As a result, women in our Sindh study relied heavily on the support of an adult male to effect this conversion whereas in West Bengal, we found that women were frequently able to benefit from project support without male assistance.

In addition, the qualitative evaluation found that, within the more socially heterogeneous context of West Bengal, the intersection of gender with other aspects of social identity differentiated the nature of patriarchal constraints for different social groups. This differentiated the pattern of outcomes reported by women from these different groups. It found that Adivasi women in particular were expected to start earning a living from a very early age, working in the fields or rivers “almost from birth”. They displayed the greatest agency with regard to project support. By contrast, Muslim women in the study were subject to greater social restrictions which curtailed their capacity to translate project support into valued capabilities.

One other set of findings from the qualitative evaluations has potential implications for the internal validity of the RCTs. In Sindh, the OCT pilot was carried out with kinship-based communities drawn from the same ethnic group and living together for generations. By contrast, in West Bengal, Trickle Up worked with different socially marginalized communities who lived in a state of considerable tension with each other. In each context, we found examples of actions stemming from community divisions and solidarities which affected the pattern of outcomes.

Resentment towards project participants in West Bengal by non-participating neighbours from other communities often resulted in the destruction and theft of assets, weakening the impact of the project for some of the participants. Equally, there were examples of positive spill-overs from both pilots as women from the participating households shared the information and knowledge they had acquired with non-participants—from their own

community groups in the West Bengal context and from their kinship networks in the Sindh context. While such behaviour would be recorded in qualitative evaluations as aspects of the way that the projects play out in different communities, they violate the non-interference assumption of RCTS, biasing their impacts in unknown ways.<sup>10</sup>

#### *4.3. Explaining Impacts Across Methodologies: The West Bengal Evaluations*

Finally, reading across the two sets of evaluations within each context, we draw on the information they provide to try and explain convergences and divergences in their findings.

The incomplete nature of this information means that the explanations are necessarily speculative but they draw attention to some of the relevant background information that could have strengthened our understanding of how these pilots performed.

First of all, we found that both the RCT and the qualitative evaluations of the pilots in West Bengal reported positive outcomes. We have already attributed this to the dense structure of market and state-supported opportunities evident in this context. This led, as we noted, to considerable livelihood diversification in the West Bengal context. But whereas the RCT found that the better-off participants progressed fastest, the qualitative evaluation reported the counter-intuitive finding that it was the poorest group in the study who made most progress.

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

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

In addition, the divergence in the distribution of outcomes reported by the two studies may also have reflected differences in the implementation of the pilots. The RCT pilot was implemented by Bandhan, a microfinance organization that uses the joint liability group model<sup>11</sup> pioneered by Grameen Bank for its lending activities. For the purposes of the pilot, it excluded participants from any microfinance or self-help groups since these were likely to be associated with government or NGO anti-poverty schemes. While project staff met with participants on a regular basis, they did not make any attempt to promote their collective organization.

Trickle-Up, on the other hand, had long experience with self-help groups which have been found to be generally more effective than joint-liability groups at reaching women from the poorest castes and tribal groups (EDA Rural Systems, 2004; Sarma and Mehta, 2014). It chose to integrate the self-help group approach into its pilot. The qualitative evaluation suggested that this was a particularly valued aspect of the project and that it was of particular value to Adivasi groups.

#### 4.4. Explaining Impacts Across the Two Methodologies: The Sindh Evaluations

Turning to the Sindh studies, both the RCT and the qualitative evaluation agreed that the better-off among project participants reported stronger impacts than poorer ones, but they diverged considerably in their overall assessment: while the RCT pilot in Sindh was deemed to have “worked”, along with other RCT pilots, the majority of the participants in the qualitative evaluation reported little or no progress. The pilot was subsequently closed down.

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

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

By contrast, we have noted that around 90% of the RCT participants were estimated to be above the poverty line. In addition, the IDS study cited earlier contained data on housing conditions that suggests that both treatment *and* control households came from more developed, probably peri-urban contexts. It reported that between 34% and 44% of participating and control households respectively had flush toilets connected to public sewerage; between 45% and 41% had electricity; between 21% and 23% had piped water; and between 7% and 15% cooked with gas.

This is in marked contrast to the hamlets in our qualitative study which had no electricity, no flush toilets, cooked with firewood and, in the absence of piped water, had to either purchase water or carry it from nearby wells. In short, while both the RCT and the qualitative study suggest that better-off participants reported more favourable outcomes than poorer ones, the fact that the RCT participants were generally better off than those in the qualitative study may explain the more favourable overall conclusion of the RCT study.

#### 4.5. Concluding Comments

This paper has used the empirical case studies of RCTs and qualitative evaluations of a similar set of development interventions to explore in greater detail what these alternative approaches to evaluation contribute to our understanding of how development interventions work. In this concluding section, I would like to draw on the broader evaluation literature to make some general points about these approaches as well as draw out some general implications for policy.

First of all, the “real life” threats to the randomization process noted in relation to the RCTs discussed in this paper have been documented in other studies as well. In some cases, RCTs seek to deal with these problems through various *ex post* “econometric or statistical fixes” (Deaton, 2010). As Deaton suggests, these are acceptable solutions, but they carry the danger of “data mining”, viz. trying out different control variables until the

experiment can be shown to have “worked”. In any case, they take RCTs out of the world of the idealization assumptions and back into the world of everyday econometrics and statistics. RCTs then become part of the wider array of quantitative methodologies in which they may sometimes have the edge in terms of rigour over other methods and sometimes not (Deaton and Cartwright 2018).

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

But this takes us to the more general problem posed by the thin empirical base on which the RCT model of causal inference rests. As critics have pointed out, the exaggerated claims made by RCT advocates for the rigour of their methodology has led them to delegitimize “other ways of knowing” as well as the “virtually ritual denigration” of knowledge gained through these other means (Bedecarrats et al., 2015). Yet a great deal of the existing knowledge relevant to improving the design of RCTs and interpreting their findings has not been, and cannot be, collected through RCTs (Deaton and Cartwright, 2018). This includes knowledge of the local contexts in which trials are set and of socio-economic differentials among the trial population that are likely to be causally relevant. Drawing on this wider body of knowledge would improve the design and explanatory power of RCTs.

The explanatory power of RCTs would also be improved through the collection of detailed qualitative information on how the interventions being trialled actually work in particular contexts. Such information would help to triangulate and elaborate on quantitative findings, placing them on firmer grounds. It would also help to tease out causal mechanisms that elude quantitative methods and to interpret unexpected findings that quantitative methods are seldom able to do.

It is worth noting, for instance, that although the synthesis report of the six RCTs found evidence of positive impacts for all six, it also noted that these impacts were “not very large”, leading the authors to engage in theoretical speculation about the nature of the poverty trap that the projects were intended to address. Yet before jumping to these higher level, theoretical explanations, the material discussed in this paper points to more mundane empirical reasons why the impacts found may have been small.

These include the failure to take up the treatment (unique apparently to West Bengal); non-compliance in the randomization process with the possibility that control households started out better off on relevant characteristics than treatment ones; project-induced misbehaviour on the part of non-participants leading to destruction and thefts of project assets; positive spill-over effects from participants to control households, reducing differences in their outcomes; differences in the agro-climatic environment, infrastructure and market conditions between treatment and beneficiary households across different contexts; variations in the effectiveness of project implementation; and variations in impact by different sub-groups which could dilute estimates of average impacts.

Finally, from a policy perspective, a “thicker” understanding of what makes a project work, fail to work or work very partially is essential for a number of reasons. At the practical level, it will help to redesign interventions to adapt them better to different contexts. At a political level, it 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. And it will have important implications for the political economy of scaling up an intervention or extending it to other contexts.

### Disclosure Statement

No potential conflict of interest was reported by the author.

### Notes

1. Ethiopia, Ghana, Haiti, Honduras, India, Pakistan, Peru and Yemen.
2. The project was a collaboration with BRAC Development Institute and the Institute of Development Studies, Sussex and funded by MasterCard. The team consisted of Karishma Huda, Sandeep Kaur, Nicolina Lamhauge, Sanchari Datta, Sajjad Abro and myself. I would also like to express my gratitude to Professor Paul Shaffer for his extremely helpful comments on an earlier version of this paper.
3. This assumes, among other things, that participants comply with the intervention; that its effects do not spill-over into the control group; and that there is no “interference” with the outcomes in the trial sample by influences other than the treatment.
4. It should be pointed out that, while it was not done for the RCTs discussed in this paper, it is possible to gain some insights into project mechanisms by randomizing different aspects of project design. But this would still leave unexplained how the different aspects worked.
5. Used by the MDGs to benchmark extreme poverty.
6. 13 households had to be dropped as they had not been randomized.
7. Over half of these had refused to be re-surveyed, others had migrated.
8. Bonell et al. (2018) make an important distinction between programmatic theories of change which focus on specific interventions and more general “theories of problems”, such as embedded in the capability approach, which seek to explain the phenomena to be addressed.
9. Self-help groups were pioneered by NGOs in India in the 1980s. They bring around 10–20 women within a village together to save regularly and to borrow from pooled savings on terms collectively agreed by the group. This approach has now been adopted by the Indian government to link poorer women to its poverty lending programmes.
10. Spillover results were not examined in either of the two RCTs.
11. Unlike SHGs, joint liability groups are formed for the purpose of guaranteeing loans provided by NGOs to members of the group.

### 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.
- 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* 32 (4): 515–548.
- Bedecarrats, F., I. Guerrin, and F. Roubaud. 2015. *The gold standard for randomised evaluations: from discussion of method to political economy* (Working Paper WP/2015-01). Paris: Development, Institutions and Mondialisation
- Bonnell, C., G. J. Melendez-Torres, and S. Quilley. 2018. “The Potential Role for Sociologists in Designing RCTs and of RCTs in Refining Sociological Theory: A Commentary on.” *Deaton and Cartwright. Social Science and Medicine* 210: 29–31.
- Deaton, A. 2010. “Instruments, Randomization and Learning about Development.” *Journal of Economic Literature* 48: 424–455.
- Deaton, A., and N. Cartwright. 2018. “Understanding and Misunderstanding Randomized Control Trials.” *Social Science and Medicine* 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*. Dhaka: BRAC Development Institute Report.
- 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.
- 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. London: School of Oriental and African Studies.
- Karlan, D. and W. Parienté. 2014. *Impact of 'Targeting the Ultra-poor' Program in Pakistan*. Research Partnerships on Participatory Development Brief 9. Islamabad: PPAF. Downloaded April 2015. <http://www.ppaf.org.pk/PPAF-Conference-2014/Publications/paper9.pdf>.
- Kidd, S., and D. Bailey-Athias. 2017. "The Effectiveness of the Graduation Approach: What Does the Evidence Tell Us?" *Policy in Focus* 14 (2): 22–28.
- Krauss, A. 2018. "Why all Randomised Controlled Trials Produce Biased Results." *Annals of Medicine* 50 (4): 312–322.
- 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 H. Gazdar. 2007. *Social Structures in Pakistan*. Islamabad: Asian Development Bank.
- Pawson, R., and N. Tilly. 2004. *Realistic Evaluation*. London: Sage Publications.
- Robeyns, I. 2005. "The Capability Approach: A Theoretical Survey." *Journal of Human Development* 6 (1): 93–117.
- Sarma, S. K., and M. H. Mehta. 2014. "The Best Model for Micro-lending: Self-help Group or Joint Liability Group?" *Journal of Rural Development* 33 (3): 247–260.
- Sen, A. 1999. *Development as Freedom*. Oxford: Oxford University Press.
- Shaffer, P. 2011. "Against Excessive Rhetoric in Impact Assessment: Overstating the Case for Randomised Controlled Experiments." *Journal of Development Studies* 47 (11): 1619–1635.
- White, H. 2009. "Theory-based Impact Evaluation: Principles and Practice." *Journal of Development Effectiveness* 1 (3): 271–284.
- Zaidi, S. A., H. Jamal, S. Javeed, and S. Zaka. 2007. *Social Impact Assessment of Microfinance Programmes*. Islamabad: Study submitted to European Union-Pakistan Financial Services Sector Reform Programme.

### **About the Author**

**Naila Kabeer** is joint Professor of the Departments of International Development and Gender Studies, London School of Economics and Political Science. She has extensive experience in research, teaching and advisory work in the related fields of gender, poverty, labour markets and livelihoods, social protection and collective action. Her recent publications include "Organizing women in the informal economy: beyond the weapons of the weak" (Zed Press, 2013) and "Gender, livelihood capabilities and women's economic empowerment: reviewing evidence over the life course" (Overseas Development Institute, 2018).