A Multi-Stage Approach to Qualitative Sampling within a Mixed Methods Evaluation: Some Reflections on Purpose and Process

Purnima Ramanujan and Suman Bhattacharjea
ASER Centre, New Delhi, India

Benjamin Alcott
UCL Institute of Education, England

Abstract: *We share experiences from a mixed methods evaluation in rural India that combines a randomized controlled trial (RCT) of 400 villages with embedded case studies in four villages. Specifically, we present two lessons from the multi-stage sampling approach adopted to select the four case-study villages, which first prioritized key-informant observations regarding intervention status in order to shortlist locations and subsequently used data from the RCT's baseline survey to select the final sample. In doing so, we highlight how large-scale mixed methods program evaluations in education can go beyond questions of "what works" to answering those of "how," "why," and "why not."*

Keywords: *India, international development, mixed methods, program evaluation*

Résumé : *Nous faisons part d'expériences découlant d'une évaluation par méthodes mixtes en Inde rurale qui combinait un essai clinique randomisé (ECR) de 400 villages avec des études de cas intégrées dans quatre villages. Plus précisément, nous présentons deux leçons tirées de l'approche d'échantillonnage à plusieurs degrés adoptée pour sélectionner les quatre villages de l'étude de cas, qui a d'abord priorisé les observations d'informateurs clés concernant l'état de l'intervention afin de présélectionner les emplacements, puis utilisé les données de l'enquête de base de l'ECR pour sélectionner l'échantillon final. Ce faisant, nous soulignons la manière dont les évaluations de programmes à grande échelle par méthodes mixtes dans le domaine de l'éducation peuvent aller au-delà de questions de « ce qui fonctionne » pour répondre à des questions de « comment », « pourquoi » et « pourquoi pas ».*

Mots clé : *Inde, développement international, méthodes mixtes, évaluation de programme*

Corresponding Author: Purnima Ramanujan, ASER Centre-Pratham Education Foundation, B- 4/58, B4 Block, Safdarjung Enclave, New Delhi, 110029, India; purnima.ramanujan@ asercentre.org

Concurrent with the rising prominence of randomized control trials (RCTs) as the "gold standard" for evidence in policymaking, there have also been calls to maintain a broader vantage point beyond the simplistic question of "if" a program works (Cartwright & Hardie, 2012). In other words, it is important also to examine the mechanisms underpinning why and how programs are effective. This often entails comparing, synthesizing, or integrating findings from quantitative and qualitative methods to answer questions around process, scalability, and sustainability (Kabeer, 2019; Outhwaite et al., 2020; Peterson, 2016).

Sampling plays an essential role in the insights that can be generated by mixed methods evaluation designs. Within the mixed methods research literature, sampling decisions have received increasing attention owing to their complex ramifications for achieving coherence between research objectives, inferences, generalizations, and interpretive consistency (Collins et al., 2006, 2007; Corrigan & Onwuegbuzie, 2020; Onwuegbuzie & Leech, 2007; Poth, 2018). In particular, the literature highlights that sampling approaches need to match the given study's research purpose (Corrigan & Onwuegbuzie, 2020). The research purpose shapes how the qualitative and quantitative components relate to one another (Newman et al., 2003), and this in turn has implications for the sampling approaches used. For example, greater independence between the qualitative and quantitative sampling approaches may suit simultaneous designs aiming at convergence or complementarity, in contrast to more contingent sampling strategies in sequential structures aiming at expansion or explanation (Palinkas et al., 2015).

However, our search of prior literature on sampling strategies found little discussion of the specific case of mixed methods evaluations, by which we mean those forms of mixed methods research that include an experimental or quasi-experimental evaluative component. Mixed methods evaluations deserve particular attention given the tendency for the logic and needs of the quantitative component to dominate decision making in the research design process (Onwuegbuzie & Hitchcock, 2017). Hence, in this article we share the qualitative sampling approach that we adopted in a large-scale mixed methods evaluation in rural India, which paired an RCT with embedded case studies in selected sites. Like most mixed methods evaluations, the quantitative component's sampling followed the Neyman–Rubin model's logic of causal inference (Holland, 1986; Rubin, 1974), which focuses on randomization and sufficient sample size to minimize standard errors and thus enable causal inference. The qualitative sampling component, though, was less straightforward: Finding systematic ways of choosing from among 200 potential case-study sites posed practical challenges to the use of traditional qualitative sampling strategies. As we will show, the research team's solution was to use a multi-stage purposive sampling approach, drawing upon both key-informant perspectives and the RCT's baseline survey data.

## DESCRIPTION OF THE RESEARCH DESIGN

Underway in 400 villages of one district in rural Uttar Pradesh, India, this mixed methods evaluation tests an intervention that aims to increase the involvement

of community and school actors in support of children's foundational learning, focusing on children in government schools who are often from relatively less affluent and less educated families, with consequently limited access to home-based support for learning.

Two forms of the intervention are being tested: a community-based initiative (in 100 villages) and a community and school collaboration initiative (in 200 villages). A third set of villages form the comparison group (of 100 villages). In the first intervention format, volunteers from the community are trained to conduct classes with children and set up study groups to inculcate the habit of studying daily and foster peer learning. Additionally, village meetings and events are organized to address children's learning and attempts made to engage influential village residents and parents to support and guide these efforts. The second format includes all elements of the first and, in addition, headmasters and teachers are encouraged to assess and understand the learning levels of children in their classroom, adapt their teaching practice in line with children's learning needs, and engage with parents and community members in a variety of activities. The full study aimed to assess the intervention's impact upon children's learning levels and explore the key dynamics, at the village level, that shape the efficacy of the intervention.

Figure 1 depicts the intended sampling strategies, stages of data collection for the respective research questions, and the intention of the data collection to be longitudinal to document changes over time. However, this timeline was subsequently disrupted by the COVID-19 pandemic.

This practice note discusses the sampling strategies employed to select sites for the qualitative strand. The qualitative strand aimed to understand how the intervention was perceived and received at the village level by focussing on key actors in the intervention's theory of change and exploring how they engaged with each other and with activities to support children's learning. Since the project's central focus was on school-community relationships, these sites were selected only from the second intervention format, that is, from 200 villages. While the decision to select four sites was dictated by what was feasible given the available resources in terms of time, money, and people, selecting four out of 200 sites
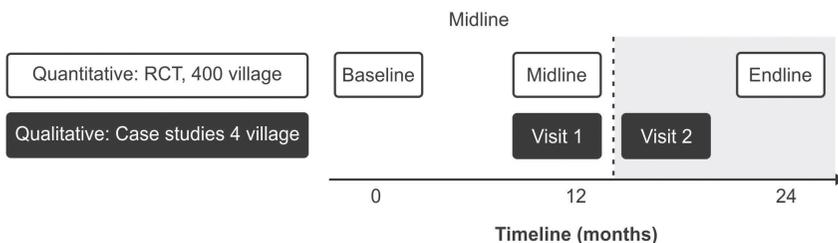
Figure 1. Initial fieldwork timeline

Note: Fieldwork disruptions caused by the COVID-19 pandemic are shown by the period shaded in grey, separated by a vertical dotted line. At the time of writing this article, the endline for the quantitative fieldwork and Visit 2 for the qualitative fieldwork had not been undertaken.

entailed parsing through a variety of options connected to the qualitative study's aims, objectives, and sequencing within the larger project.

OVERVIEW OF QUALITATIVE SAMPLING STRATEGY

Several factors influenced our qualitative sampling approach. One was the research objective, which aimed to identify potential barriers and enablers in the intervention's theory of change and uncover explanatory narratives underlying the intervention (Yin, 2003). Second, the qualitative strand was designed to begin after the quantitative baseline survey had been completed; hence, baseline data were available to inform qualitative sampling. Third, this work was to begin once the intervention had begun, thus generating an additional source of information in the form of program teams' assessments of intervention sites as well as their characterization of the responses they were receiving to their work.

Given these multiple timepoints and sources of information, a decision had to be made on how best to utilize them. Should the process of sampling prioritize the baseline data from randomly selected villages, or should it privilege the lived experience of the teams working in them? In other words, should the baseline data be interrogated to generate a shortlist of villages, which could then be further shortened using field teams' experiences, or should the process operate in the reverse direction, starting with field teams' assessments and moving to baseline data? There were proponents of each approach: Those seeking to privilege the former emphasized the representative, objective nature of the quantitative data, while those in favour of privileging the latter viewed the process elements identified by field teams as being more relevant to the research question than the baseline data.

Early on in this process, discussions with program staff revealed that villages differed greatly in their responsiveness to the intervention. For example, volunteer-led community classes were more difficult to organize in villages with low levels of youth and adult educational attainment, while in poorer villages, children were often involved in household or income-generating work and thus unavailable to attend such classes. Issues stemming from caste, class, gender, and power dynamics were often starkly visible in the relationships between community and school actors—all of which impeded progress toward the intended shared focus on children's learning. These discussions revealed important insights about adaptations made to suit the needs of these sites, such as flexibility in scheduling volunteer classes and organizing night visits by staff to meet caregivers and other community stakeholders who were unavailable during the day.

Overall, intervention teams reported pockets of action that were visible as a result of their work in villages where, for example, mothers began taking an active role in ensuring that their children attended school regularly, residents and teachers attended community events and facilitated finding volunteers, or parents and teachers reported improved attendance levels in schools. But teams also described villages where, despite equivalent efforts, no progress was apparent. Crucially, these discussions provided the impetus to focus the qualitative research

not only on sites demonstrating engagement with the intervention, as originally planned, but also on sites where teams struggled to implement the program, especially since the forms of deprivation that made some sites challenging also made the intervention arguably more relevant for children and communities in these locations.

In the end, we decided to focus our research on "information-rich" locations (Patton, 1990) and adopted a multi-stage purposeful approach to sampling sites (Collins et al., 2007). The RCT's baseline data, though vast, did not reflect the dynamics of community responses to intervention teams and activities. In this context, we defined "information-rich" villages as those that the program staff identified at extreme ends of the response spectrum: that is, villages where school and community actors had shown signs of early, positive engagement with the intervention and those where no evidence of engagement was visible. Following this decision, in the first stage, we used recommendations from the program team and in the second stage further narrowed down the selection based on quantitative baseline survey indicators.

Stage 1: Shortlisting using key informants

Based on early field visits and conversations with program staff, the research team developed a set of parameters to describe and classify the response and engagement of community and school actors to the intervention, accompanied by questions to facilitate discussions (see Table 1).

Table 1. Parameters for key-informant led selection of sites

Stakeholder	Guiding questions
Community	• Are volunteers in the village engaged and regular in running learning camps? • Do volunteers take initiative in organising community school events? • Do parents send their children to participate in intervention activities, and/or organize peer-groups themselves? • Have parents' views about the government teachers changed? Are they ensuring that their child attends school? • Have influential members in the community engaged and helped advocate for the project, either directly or indirectly?
School	• Do teachers know about the intervention? Do they take an active interest? • Are teachers aware of the volunteers who teach children in the villages? Do they engage with program teams when they visit the school? • Do teachers involve themselves in community activities? Has there been any change in what teachers do? • Have the teachers' thinking about their own role and/or that of the community/parents changed?

Sixteen senior program staff were included as key informants based on their leadership and supervision of field personnel in the program. Following criterion sampling (Collins et al., 2007; Palinkas et al., 2015), participants were asked to nominate two "responsive" and two "difficult" sites in each of the 15 blocks (rural administrative units) where the program was operating. "Responsive" sites were those where teams noted positive engagement from both teachers and community members with the likelihood of further improvement in relationships. Sites categorized as "difficult" were those where, despite similar programmatic effort, the situation after eight months of the intervention remained more or less the same as at the start of the project.

Following this process, key informants identified 25 "responsive" and 26 "difficult" sites, accompanied by a written rationale explaining their choices.

Stage 2: Narrowing down the selection using quantitative criteria

Members of the research team then followed a series of steps to narrow down the candidate list of villages from 51 to 4, using a set of quantitative indicators from the baseline data. Table 2 lists these indicators and the accompanying rationale.

Researchers compared these indicators for the shortlisted villages and narrowed down the selection in three steps. The first step was to reduce the number of shortlisted villages while maintaining geographic spread, by selecting one village per category (i.e., responsive and difficult) per administrative block. In the next step, access and convenience of fieldwork were prioritized, with villages shortlisted based on their distance from district headquarters. The final step prioritized two indicators considered crucial for the overall study: teacher and parental awareness of children's learning levels, because improving stakeholder awareness on learning was a major program goal; and student enrolment and attendance, because these were considered to be key proxies for school size and functioning.

Throughout this process, the explicit objective was to maximize variation (Collins et al., 2007; Palinkas et al., 2015) and select villages that differed from one another on a variety of dimensions, so that emerging conclusions could be triangulated across different contexts. Baseline data revealed that although sample children across all four sites had similar learning levels, there was substantial variation in school and household characteristics as well as in the awareness of teachers and parents with respect to learning outcomes. For example, one shortlisted site could have low school enrolment, low attendance, and high levels of misalignment in teacher and parental awareness of children's learning levels, while another could have the opposite characteristics. This comes back to the value of integrating the different elements of a mixed methods evaluation: Our sampling process was enhanced by having data from diverse data sources and methodological perspectives.

The final sample included three "responsive" sites and one "difficult" site. The imbalance in this selection was intentional, aiming to prioritize learnings from sites characterized as showing high engagement with the intervention in order to

Table 2. Criteria used for shortlisting sites

	Indicator	Rationale
School	• Student enrolment and attendance in sample grades • Number of teachers employed and number present • Number of multi-grade classes • Number of sample teachers with travel time less than 30 minutes • Number of sample teachers living in the same village	• Understand school context (size) and functioning (attendance). • Understand teacher allocation across grades as well as challenges of organising teaching and learning. • Understand teacher proximity and connection to the community and children.
Household	• Caste distribution • Construction of walls (Permanent, Semi-permanent, Temporary) • Asset ownership • Education of household members	• Understand distribution of socio-economic advantage and disadvantage among sample households. • Understand household education and learning support context.
Teacher and parental awareness of children's learning	• % of children at different levels of reading and numeracy • % teachers and % parents by whether they correctly or incorrectly identified children's learning levels	• Understand alignment or misalignment of teacher and parental perception with children's learning levels.
Distance from district headquarters	• Distance in kilometres	• Logistical consideration for ease of travel and access by the research team.

feed these findings back to the intervention team to help them incorporate more effective strategies. Notably, the "difficult" site selected for the final sample bordered a "responsive" site, thus enabling the study of seemingly different processes and outcomes within the same geographic and administrative unit.

WHAT LESSONS WERE LEARNED?

We identified two key learnings with regard to integrating mixed methods in sampling decision making. The first is that utilizing multiple data sources to take sampling decisions, though tricky, is crucial—especially in the context of complex mixed methods evaluations. No one data source can do alone what data from different sources can offer together. For complex mixed methods evaluations aiming to answer questions of "why" and "why not," utilizing multiple data sources for sampling allows researchers to balance and meet competing needs of contextuality and variation.

Within the field of qualitative research, the use of key informants for sampling decisions is well established (see, for example, Hammersley & Atkinson, 2007; Mason, 2018; Onwuegbuzie & Leech, 2007). Because our objective was not to generalize to other contexts but rather to understand processes, key informants' perspectives on what was happening on the ground provided an understanding of human dynamics in these locations that was not reflected in the baseline data and enabled the shortlisting of information-rich locations. But these observations did not allow us to systematically discriminate between shortlisted sites on the basis of other key parameters of interest, such as socioeconomic characteristics of households or school characteristics, that were reflected in the quantitative baseline data.

Using quantitative methods alone, it would be nearly impossible to select just four unique cases from the extensive baseline data collected from the 200 candidate sites. While every site was unique in some respects, many were similar across key dimensions such as school size, or even caste or gender distribution among sampled children. In any case, no process of selection of just four sites based on this data could lead to the possibility of generalizing results to a larger set of locations. Most importantly, exclusive use of baseline data for sampling could have resulted in selecting only sites where the intervention had failed to take root, leading to the impossibility of answering the specific research question for which the qualitative component was designed.

The second is that timing plays a critical role in shaping the opportunities to collect and utilize data for decision making, especially in mixed methods research. Sequencing study components is essential to laying out a systematic process and enabling project agility. However, in this study it was not simply a question of sequencing the quantitative and qualitative strands of the work (i.e., quals led, quants led, or both in parallel), but equally a question of the specific stage at which both the quantitative strand as well as the intervention itself were at when these decisions were being made. In our project, qualitative sampling decisions benefited only from baseline data, but if these decisions had been taken after both baseline *and* midline data collection had been completed, we might have been able to identify quantitative proxies for the process indicators that we were most interested in capturing, and to potentially compare these with key informants' assessments of the situation. Similarly, if qualitative sampling decisions had been taken in the very early days of the intervention's rollout, key informants' assessments of what was happening on the ground might have looked very different than they did when this process was undertaken a few months later.

CONCLUSIONS

The complexities and dynamics unique to mixed methods evaluations make it worth focusing specifically on how to sample for qualitative purposes within such designs. Further developments on this topic would be helpful, both at the conceptual level and through more first-hand reflections from fieldwork. Ultimately, it may be impossible to identify a single "best" way to select four sites from

200 options; perhaps searching for such a formula is asking the wrong question. Context matters in evaluation (White, 2009; Rog, 2012); hence, it should form a key element, alongside the study's goals and research questions, in shaping the sampling strategy. This is just as important when integrating mixed methods in a single evaluation study—the focus of this special issue. In sharing one such sampling strategy that fitted our study goals and research context, we hope to provide a useful example to future research in the field of mixed-methods evaluation.

ACKNOWLEDGMENTS

REFERENCES

Cartwright, N., & Hardie, J. (2012). *Evidence-based policy: A practical guide to doing it better*. Oxford University Press. https://doi.org/10.1093/acprof:osobl/9780199841608.001.0001

Collins, K. M. T., Onwuegbuzie, A. J., & Jiao, Q. G. (2006). Prevalence of mixed-methods sampling designs in social science research. *Evaluation & Research in Education*, *19*(2), 83–101. https://doi.org/10.2167/eri421.0

Collins, K. M. T., Onwuegbuzie, A. J., & Jiao, Q. G. (2007). A mixed methods investigation of mixed methods sampling designs in social and health science research. *Journal of Mixed Methods Research*, *1*(3), 267–294. https://doi.org/10.1177/1558689807299526

Corrigan, A. J., & Onwuegbuzie, A. J. (2020). Toward a meta-framework for conducting mixed methods representation analyses to optimize meta-inferences. *The Qualitative Report*, *25*(3), 785–812. https://doi.org/10.46743/2160-3715/2020.3579

Hammersley, M., & Atkinson, P. (2007). The process of analysis. In *Ethnography: Principles in practice* (pp. 158–190). Routledge. https://doi.org/10.4324/9780203944769

Holland, P. W. (1986). Statistics and causal inference. *Journal of the American Statistical Association*, *81*(396), 945–960. https://doi.org/10.1080/01621459.1986.10478354

Kabeer, N. (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. https://doi.org/10.1080/19452829.2018.1536696

Mason, J. (2018). *Qualitative researching*. Sage.

Newman, I., Ridenour, C. S., Newman, C., & DeMarco, G. M. P. (2003). A typology of research purposes and its relationship to mixed methods. In A. Tashakkori & C. Teddlie (Eds.), *Handbook of mixed methods in social and behavioral research* (pp. 167–188). Sage. https://doi.org/10.4135/9781506335193

Onwuegbuzie, A. J., & Hitchcock, J. H. (2017). A meta-framework for conducting mixed methods impact evaluations: Implications for altering practice and the teaching of

evaluation. *Studies in Educational Evaluation*, *53*, 55–68. https://doi.org/10.1016/j.stueduc.2017.02.001

Onwuegbuzie, A. J., & Leech, N. L. (2007). Sampling designs in qualitative research: Making the sampling process more public. *The Qualitative Report*, *12*(2), 238–254. https://doi.org/10.46743/2160-3715/2007.1636

Outhwaite, L. A., Gulliford, A., & Pitchford, N. J. (2020). A new methodological approach for evaluating the impact of educational intervention implementation on learning outcomes. *International Journal of Research & Method in Education*, *43*(3), 225–242. https://doi.org/10.1080/1743727x.2019.1657081

Palinkas, L. A., Horwitz, S. M., Green, C. A., Wisdom, J. P., Duan, N., & Hoagwood, K. (2015). Purposeful sampling for qualitative data collection and analysis in mixed method implementation research. *Administration and Policy in Mental Health and Mental Health Services Research*, *42*(5), 533–544. https://doi.org/10.1007/s10488-013-0528-y. Medline:24193818

Patton, M. Q. (1990). *Qualitative research and evaluation methods* (2nd ed.). Sage.

Peterson, A. (2016). Getting "what works" working: Building blocks for the integration of experimental and improvement science. *International Journal of Research & Method in Education*, *39*(3), 299–313. https://doi.org/10.1080/1743727x.2016.1170114

Poth, C. N. (2018). *Innovation in mixed methods research: A practical guide to integrative thinking with complexity*. Sage.

Rog, D. J. (2012). When background becomes foreground: Toward context-sensitive evaluation practice. *New Directions for Evaluation*, *2012*(135), 25–40. https://doi.org/10.1002/ev.20025

Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and non-randomized studies. *Journal of Educational Psychology*, *66*(5), 688–701. https://doi.org/10.1037/h0037350

White, H. (2009). Theory-based impact evaluation: Principles and practice. *Journal of Development Effectiveness*, *1*(3), 271–284. https://doi.org/10.1080/19439340903114628

Yin, R. (2003). *Case study research: Design and methods*. Sage.

AUTHOR INFORMATION

Purnima Ramanujan is Head of Research at ASER Centre, Pratham Education Foundation, and leads the design, planning, and implementation of large-scale research studies on education and policy in India. Her research interests include early childhood education, school-family engagement, and qualitative and mixed methods research.

Suman Bhattacharjea is Director (Research) at ASER Centre, the research and assessment unit of Pratham Education Foundation in India, where she leads the design and implementation of small- and large-scale research studies focusing on what and how children and youth learn in rural India.

Benjamin Alcott is an associate professor at the UCL Institute of Education, in London, England. His research focuses on education policies in low- and middle-income countries.