Theme 2: Causal Inference and Politics of Evidence

NEP 2020: From Policy to Praxis

Theme 2: Causal Inference, Scale and the Politics of Evidence in Education Reform

Consolidated Extended Abstracts

1. **Designing, Implementing, and Evaluating School Leadership Training: Reflections from an RCT within a Dynamic Government Education System in Telangana** By Apurva Sankar and Ismeet Gulati (Alokit)

Extended Abstract:

1. Context & Policy Linkage (NEP 2020)

The National Education Policy (NEP) 2020 positions teachers at the centre of educational reform, recognising them not merely as implementers of curriculum but as key agents of systemic change. The policy emphasises continuous professional development, recommending a minimum of 50 hours of annual professional learning, mentoring-based support structures, and a shift from Passive learning–oriented training to reflective, practice-based teacher training. NCERT's teacher professional development (TPD) guidelines further conceptualise professional development as a continuous, career-long process that may include formal courses, school-based learning, mentoring, peer collaboration, and reflective practice.

Despite this policy vision, teacher professional development in many Indian states continues to be implemented through large-scale cascade models that emphasise information dissemination over instructional improvement. An examination of TPD structures and practices in the district revealed several limitations. These limitations are outlined below.

First, RPs are seen as 'need-based' trainers. Whenever a complex meeting or any other state-wide training has to be conducted, the RPs are given the duty. However, this approach has its demerits. RPs have limited opportunities to develop depth in the training content, as pre-designed decks typically combine multiple topics into a single-day training. They are given information only on 'what' the content is and not on 'how to deliver' the content to teachers. Their inability to effectively facilitate learning for teachers becomes a major roadblock. Moreover, RPs often experience hesitation in leading training spaces, as teachers are their peers, and underlying power dynamics make it difficult to manage large groups of experienced teachers.

Second, Complex Meetings are state-mandated monthly spaces that must be conducted for all teachers by the Resource Persons. However, these trainings are largely done in a lecture method, meetings become a space to overload teachers with more information or operational details about new interventions, lacking brainstorming and peer learning opportunities.

Third, post-training classroom implementation is seldom monitored, making TPD compliance-based.

To address these, a district-embedded, practice-based TPD model was designed. The approach focused on:

A. Strengthening instructional leadership by creating a full-time DRP cadre,

B. Redesigning TPD spaces to be immersive and practice-oriented, and

C. Establishing structured post-training school visits and class observations to ensure implementation quality.

A key structural change was the introduction of preparatory meetings before each complex meeting. These preparatory spaces are conducted in two stages:

A. Our team members work with DRPs on both content and facilitation, and

B. DRPs subsequently conduct preparatory sessions with MRPs.

These sessions focus on deep engagement with content, anticipation of challenges in classroom implementation, contextual adaptation, and facilitation skills such as questioning, managing resistance, and supporting peer learning. The preparatory spaces also gave opportunities to the RPs to practice the given Literacy and Numeracy strategies themselves, to build their confidence and conviction before delivering it to the teachers.

2. Research Questions

The study attempts to examine the model of TPD that improves teacher practices at scale through the following specific research questions:

What features of district-level TPD spaces (specifically, monthly complex meetings) enable observable improvements in classroom teaching-learning practices?

What forms of professional and structural support enable District and Mandal Resource Persons (DRPs and MRPs) to function as effective instructional leaders rather than ad-hoc knowledge-delivering trainers?

How does strengthening feedback loops between TPD spaces and classroom practice influence the quality and consistency of implementation across schools?

3. Methodology & Data Sources

The study adopts a mixed-method approach, examining the District-Embedded Practice-Based TPD model over the last 18 months across primary schools in Peddapalli district. The intervention involved working closely with 6 District Resource Persons (DRPs) and 88 Mandal Resource Persons (MRPs), with 7 complex meetings conducted annually across 36 school complexes.

Multiple quantitative and qualitative data sources were used to study the impact of the program such as observations of preparatory meetings and complex meetings, focusing on facilitation quality, teacher engagement, and nature of learning activities, classroom observations during monthly school visits, teacher feedback forms and reflective discussions, documenting teacher perceptions of usefulness, confidence, and applicability and RPs feedback form after each capacity-building session and preparatory space.

4. Key Findings / Results

The findings indicate that the components of the program, such as preparatory spaces before the teacher training, development of content mastery through practice in classrooms, structured facilitation tools, and feedback mechanisms, have strengthened the leaders' ability to design interactive, practice-oriented teacher professional development. Moreover, it has enhanced their facilitation skills, making the teacher learning spaces participatory and relevant. The RPs mentioned significant improvement in their confidence while facilitating teacher training spaces. They reported growth in their facilitation skills, including maintaining eye contact with participants, clarity in communication and instructional delivery, managing disruptions, using checks for understanding (CFUs), and engaging participants through diverse learning activities.

Second, experiential and practice-based learning within complex meetings enhanced teacher engagement and belief in new pedagogical strategies. Teachers participated in demo lessons, peer practice, and real-time classroom teaching with students. When teachers practiced strategies with students in real-time during the complex meeting, their questions became more nuanced and grounded in classroom realities, moving away from hypothetical or assumption-based concerns. On average, for 4 complex meetings conducted so far in 2025-26, teachers have rated the usefulness at 3.37 out of 4. They reported that these complex meetings provided them with useful strategies to support students who are struggling the most.

Third, post-training classroom observation has helped DRPs to check the implementation quality, identify misconceptions in teacher practices, and their challenges in strategy implementation. They are also able to incorporate these observations in the subsequent complex meetings. For instance, during multiple school visits, the teachers asked for strategies to teach word problems to students. The RPs addressed this in the next complex meeting. Subsequently, in the next complex meeting, misconceptions in strategy implementation were brought to notice and opened up space for teachers to brainstorm solutions. This has also improved the accountability for teachers to implement the learnings from the complex meetings. In the interviews conducted with DRPs, they stated that through their school visits, they can identify good practices, understand implementation of strategies, give grounded suggestions based on their learnings from other schools, and also identify challenges faced by students and teachers, which can further be addressed in the upcoming complex meeting spaces. They are also able to build interpersonal skills and cordial relationships with teachers through these visits.

Additionally, a full-time DRP cadre can bring cohesiveness and continuity to practices across the district, which has also resulted in substantial FLN growth in students from baseline to midline outcomes.

5. Policy Implications & Relevance to NEP 2020

The findings of this study offer several implications for the design and implementation of teacher professional development aligned with NEP 2020. The study demonstrates that strengthening instructional leadership at the district level is critical for translating policy intent into classroom practice. This can be operationalised by redefining the role of Resource Persons and reimaging the design of monthly Complex Meetings.

The study indicates that such a district-embedded, practice-based TPD model is feasible at scale within existing administrative structures. The cascade model can be strengthened through targeted role clarity, preparatory support, and systematic follow-up, offering a practical pathway for improving TPD.

2. **Evaluation of Educate Girls' Primary Learning Program** By Ashwini Maslekar and Shovik Chatterjee (Educate Girls)

Extended Abstract:

1. Context & Policy Linkage (NEP, 2020)

Although primary enrolment rates have improved, India continues to encounter persistent challenges in girls' education, especially in rural and socio-economically marginalised areas. These challenges extend beyond initial access to include retention through secondary education, where dropout rates increase significantly during adolescence (Muralidharan & Prakash, 2017). Evidence from Rajasthan, which has one of the lowest female literacy rates nationally, demonstrates that entrenched gender norms related to early marriage, domestic labour, and mobility restrictions create multiple barriers that undermine girls' educational trajectories.

The National Education Policy 2020 explicitly prioritises addressing these structural barriers. The policy emphasises equitable access for underrepresented groups, particularly girls from Socially and Economically Disadvantaged Groups (SEDGs). It recognises that achieving gender parity requires interventions extending beyond enrolment to address retention, transition to secondary Education, and the development of girls' agency (Ministry of Education, 2020). Critically, NEP 2020 underscores the role of community participation in educational outcomes and calls for holistic approaches that address socio-cultural determinants of educational exclusion.

The Vidya Program by Educate Girls addresses both policy priorities and practical implementation. Operating in multiple states with high gender gaps, the program utilises community-based volunteers (Team Balika) to identify out-of-school girls, mobilise households for enrolment, and promote normative change regarding girls' education. Previous experimental evaluations reported significant learning gains during implementation (IDinsight, 2018). However, there is limited evidence regarding the durability of outcomes, post-program sustainability, and normative change. This gap is particularly relevant for the NEP 2020 reform agenda, which relies on understanding whether community-level interventions can produce lasting change.

2. Research Questions

This study assessed the sustainability of outcomes from the Educate Girls' Vidya Program in regions where the organisation had exited at least three years earlier. The primary research questions were as follows:

1) To what extent have key changes in community behaviour been maintained post-program, particularly regarding: (a) sustained enrolment and retention of girls in school, and (b) girls' agency in financial and social decision-making?

2) What pathways of change—including parental norms, community ownership, and volunteer networks—appear to have persisted after program exit, and which have attenuated?

3) How do sustainability outcomes vary across districts and socio-demographic contexts, and what does this heterogeneity imply for scalability?

These questions address a critical gap. While impact evaluations increasingly demonstrate effectiveness under controlled conditions, policymakers implementing NEP 2020 reforms require evidence on the sustainability of interventions, particularly those targeting deeply embedded social norms.

3. Methodology & Data Sources

The study utilised Contribution Analysis (CA), a theory-based evaluation method designed to assess whether an intervention made a credible and significant contribution to observed outcomes in contexts where experimental designs are not feasible (Mayne, 2012). This methodological approach addresses several common constraints in real-world program evaluation: the program concluded three or more years earlier, making prospective experimental design impossible; no comparable control group was available due to geographic targeting and phased implementation; outcomes such as normative change, sustained agency, and community ownership are not easily measured by standardised metrics; and ethical considerations prevented withholding intervention from identified out-of-school girls.

Contribution Analysis provides a rigorous alternative by systematically examining the program's Theory of Change against empirical evidence, testing underlying assumptions, identifying rival explanations, and assessing whether observed outcomes align with predicted causal pathways (Mayne, 2001). Instead of estimating counterfactual impact, Contribution Analysis evaluates whether it is plausible that the intervention made a significant contribution based on observed outcomes.

Data Sources: Mixed-methods data were collected across four districts in Rajasthan (Bhilwara, Jhalawar, Pali, and Sirohi), selected through purposive sampling based on enrolment figures and retention rates (Robinson, 2014). Quantitative data were obtained from structured surveys administered to 422 girls aged 7 to 23 years. Qualitative data included 54 semi-structured interviews with parents, Team Balika volunteers, ASHA workers, principals, government officials, and program staff.

Analytical Approach: For each sustainability outcome, the analysis systematically examined the causal chain from intervention to outcome, verified underlying assumptions, assessed risks to sustainability, identified unintended consequences, and considered rival explanations such as government schemes, economic factors, and school-level initiatives.

Limitations: Potential recall bias in retrospective self-reports, inherent limitations of causal attribution in non-experimental designs, and the use of purposive sampling constrain the statistical generalisability of the findings.

4. Key Findings/Results

Sustained Enrolment with Shifting Parental Attitudes: Across all districts, 91.7% of surveyed girls were currently enrolled and attending school—a notably high rate given the program had exited these geographies over three years prior. Qualitative evidence revealed substantial shifts in parental attitudes: Education is increasingly framed as a pathway to self-reliance and delayed marriage rather than mere marriage preparation. Fathers emerged as key supporters of enrollment across districts, with 62.3% of girls citing fathers as primary enrollers. Persistent Adolescent Dropout: Despite high overall enrolment, dropout rates increase significantly from age 15, with 38% of 15-year-olds and 54% of 18-year-olds out of school. Child marriage (accounting for 25% of dropouts), household labour demands, and schooling costs remain primary barriers. These findings suggest that although enrolment norms have shifted, structural constraints during adolescence continue to persist.

Uneven Sustainability Across Districts: Outcomes varied substantially by context. Sirohi demonstrated near-universal enrolment (98.8%) and strong community mobilisation, whereas Jhalawar exhibited higher dropout rates (83.5% enrolled) and weaker continuity of Team Balika volunteers. Where volunteers remained active, normative shifts and school engagement persisted. In contrast, disengagement of volunteers corresponded with diminished community-level change.

Financial Access Without Autonomy: Although 74.5% of girls aged 13 and above possessed bank accounts, only 39% used them independently, with fathers typically operating the accounts on their behalf. This finding highlights a gap between access to financial infrastructure and the development of meaningful financial agency.

Emerging Social Agency with Limited Public Participation: Girls are increasingly described as confident and vocal within their households, and more than half of enrolled girls encourage their peers to attend school. However, participation in community-level activities remains minimal (12.8%), indicating that gains in agency have not yet extended to public spheres.

5. Policy Implications & Relevance to NEP 2020

Rethinking Evidence for System Reform: The study demonstrates that credible causal narratives can be constructed without experimental designs when rigorous methods are applied. For outcomes central to NEP 2020, such as norm change, sustained agency, and post-program sustainability, experimental evidence alone is insufficient. Contribution Analysis provides a complementary approach for assessing the long-term, system-level changes required by policy reform.

Implications for NEP 2020 Implementation: The findings have specific implications for the equity agenda of NEP 2020. First, post-primary retention strategies should address the increase in adolescent dropout by targeting marriage norms, household labour expectations, and access to secondary education, rather than focusing solely on primary enrolment. Second, community-anchored interventions show potential for sustainability but require mechanisms to maintain volunteer engagement after program completion. Third, financial inclusion initiatives for girls should extend beyond account opening to include financial literacy and address household norms that limit girls' economic agency.

Implications for Evaluation Practice: As NEP 2020 implementation advances, evaluation frameworks should expand beyond RCT-derived impact estimates to address the complexity of educational reform. Mixed-methods approaches that integrate quantitative outcome measurement with theory-based causal assessment can generate actionable evidence for policymakers engaged in adaptive, long-term reform."

3. **Peer Teaching for Foundational Literacy and Numeracy: A Scalable and Equitable Pedagogical Model in Government Schools** By Anna Daniel and Atmaja Acharya (Involve Learning Solutions Foundation)

Extended Abstract:

Introduction

India’s education system today is at a critical juncture. Despite reaching over a million schools and serving the world’s largest youth population, foundational learning remains alarmingly low. ASER 2024 reports that 70% of Grade 5 children cannot solve a simple division problem, and two-thirds of Grade 3 students struggle with basic subtraction. The crisis is particularly visible in states such as Bihar, where multi-grade classrooms, teacher shortages, and top-down pedagogy limit opportunities for personalised learning.

Involve Learning Solutions Foundation challenges this paradigm by placing agency at the heart of pedagogy not as an abstract ideal, but as a practical response to India’s learning crisis. As a nonprofit committed to ensuring that all children develop age-appropriate agency to thrive, Involve has worked across six states over the last seven years, reaching more than one million students through government & ecosystem partnerships. The organization’s core intervention, Peer Teaching (PT), reframes learners not simply as beneficiaries of reform but as co-creators of it.

Reimagining Pedagogy Through Peer Teaching

Peer Teaching transforms the classroom from a site of instruction to that of co-creation. Student Champions selected for their conceptual clarity support 4 -5 peers through structured learning circles embedded within regular school hours. Within these circles, children explain basic mathematical concepts, engage in multimodal learning through locally sourced materials and self-created teaching–learning materials.

Research Questions:

1. What is the impact of Peer Teaching on foundational learning outcomes in Indian government primary classrooms?
2. How does Peer Teaching reshape classroom dynamics, particularly peer interactions, emotional safety, and inclusion?

Methodology:

This study uses a cluster randomised controlled trial (RCT) to examine the effects of same-grade Peer Teaching within government primary school classrooms. The research was conducted in 177 public schools in Bhagalpur district, Bihar, focusing on students in Grades 3–5. Schools were randomly assigned to a treatment group (81 schools) or a control group (96 schools). Data were collected through baseline and endline assessments covering approximately 14,000 students. In addition to student assessment, semi-structured interviews (25-30 minutes each) with 35 teachers were conducted to understand teachers' experience and perspectives on Peer Teaching. These interviews were conducted in Hindi and audio-recorded. Recordings were transcribed and translated into English using manual and AI-assisted tools for accuracy.

In addition to the RCT, a six-month qualitative study was conducted across five government schools to closely examine classroom processes and inclusion under Peer Teaching. Data were collected through classroom observations across 15 peer groups, focusing on participation patterns, peer interaction, and indicators of emotional safety.

Theoretical Framework

The theoretical grounding of this work draws from Paulo Freire’s concept of conscientization and Lev Vygotsky’s theory of social learning. Freire critiques what he calls the “banking model” of education, where students are treated as empty vessels to be filled with information (Freire, 1970). He argues that learners must be recognised as active participants who construct knowledge through reflection and dialogue. Peer Teaching aligns closely with this vision by positioning students as co-creators of knowledge.

Vygotsky similarly situates learning within social interaction, proposing that understanding develops through collaboration with others, particularly within the Zone of Proximal Development; the space where learners can achieve more with support than they can independently (Vygotsky, 1978). Peer Teaching leverages this social dimension by enabling students to explain concepts to one another using shared experiences, and relevant examples. This results in a pedagogy that is grounded in everyday classroom interactions, shaped by students’ voices, experiences, and collaborative problem-solving.

Key Findings & Results

The students who attended Peer Teaching sessions showed a mathematical score improvement of 0.15 - 0.16 standard deviations relative to control schools, with statistically significant gains. Learning improvements are observed across multiple assessed domains including arithmetic operations, division, and word problems, indicating effects beyond basic skills. Significantly, the gains extended to both high-performing tutors and lower-performing learners, indicating that Peer Teaching strengthens conceptual clarity. Students in treatment schools also report improved classroom experiences, including greater ease of studying and lower anxiety. The RCT study documented a 0.10 SD reduction in learning-related anxiety among students attending Peer Teaching sessions.

Using directed classroom social network measures based on in-degree and out-degree of study-group, help-seeking, and friendship ties, we find that peer tutoring produces smaller, more effective academic networks and connects low-performing learners to high-performing peers, with the average academic quality of their learning networks increasing by about 0.16 - 0.20 standard deviations.

The six-month qualitative study captured rich shifts as well: migrant students who initially hesitated to speak evolved into group leaders; children switched languages, learned more expressive interactions to include all peers; conflicts decreased as shared responsibility grew; and students began designing curriculum elements themselves. In addition to this, indicators of emotional safety, reflected through joy markers such as smiling, requesting turns, and initiating dialogue, were observed in over 80% of sessions.

As student agency and peer leadership increased, teachers reported a parallel shift in their role, with reduced day-to-day instructional load and improved classroom management. When students began to take greater ownership of learning and step into leadership roles, teachers experienced a noticeable reduction in day-to-day instructional burden. At the same time, learning within peer groups accelerated: students progressed more quickly as they learned together and moved through the curriculum at a faster pace through shared understanding. Teachers noted reduced behavioural issues, improved student engagement, and greater collaboration among the students. These transformations highlight that agency-based pedagogy benefits not only students but also educators who have long been constrained by systemic pressures and limited autonomy.

Policy Implication & Relevance to NEP

The National Education Policy (NEP) 2020 identifies foundational learning as an “urgent and necessary prerequisite” for systemic improvement in school education. Evidence from a large-scale cluster randomized controlled trial indicates that structured peer tutoring (PT), when embedded within regular classroom practice, improves mathematics learning outcomes by approximately 0.15–0.16 standard deviations, demonstrating its potential as a scalable approach to addressing foundational learning gaps. The PT model supports NEP's emphasis on competency-based education by enabling continuous, low-stakes formative assessment through student leaders, facilitating early identification and remediation of learning deficits among lower-performing students. In addition to academic gains, the intervention advances NEP's focus on 21st-century skills by strengthening communication, collaboration, problem-solving, and leadership through structured peer interactions. The small-group design operationalizes NEP-recommended pedagogical shifts toward collaborative and discussion-based learning without requiring additional classroom resources. While PT is not a standalone teacher training or capacity-building intervention, it contributes to NEP's system-strengthening agenda by reducing learning heterogeneity within classrooms, creating additional in-class learning facilitators, and enabling teachers to focus on effective grade-level instruction. Collectively, these findings suggest that achieving quality at scale may not require uniform delivery alone; rather, enabling structured classroom-level agency can be a critical lever for sustainable, cost-effective, and system-wide implementation of NEP reforms.

Conclusion

Involve's work offers a reimagined vision of pedagogy grounded in agency, reciprocity, and indigenous ways of knowing. Peer Teaching demonstrates that learning can flourish when classrooms shift from hierarchical spaces of instruction to collaborative spaces of shared authorship.

By weaving together evidence, insights, and frameworks, this work argues that Peer Teaching is not merely a strategy but a transformative pedagogy. It restores the student as a central actor in the learning process and positions teachers as facilitators of agency and offers a pathway for governments and educators to build classrooms where every child learns in relations rooted in community, strengthened by collaboration, and empowered by agency.

References:

Freire, P. (1970). Pedagogy of the oppressed. Continuum.

Ministry of Education, Government of India. (2020). National Education Policy 2020. Government of India.

Vygotsky, L. S. (1978). Mind in society: The development of higher psychological processes. Harvard University Press."

4. **Do Digital Add-ons improve Phone-based Education? Evidence from four field Experiments in India** By Neaketa Chawla and Dr. Ambrish Dongre (IIM-A)

Extended Abstract:

Context and Policy Linkage

A significant share of young children in developing countries fail to attain basic literacy and numeracy, even after completing the early grades of primary schooling (Grades I–III). This challenge, often referred to as the learning crisis, has drawn considerable attention. Recognising the urgency of this issue, the National Education Policy (2020) identifies the achievement of universal FLN in primary schooling as the highest priority of the education system. Given the scale of the problem, it also underscores exploring all viable methods in order to achieve the same. Most interventions to address it have focused on schools and teachers, while the role of parents has received relatively little attention (Cardim et al., 2023; Duflo et al., 2024). This paper examines a series of interventions conducted in partnership with a non-profit organization in India that sought to increase parental involvement through mobile devices in children’s education and thereby improve their learning outcomes.

Background

The partner organisation works with children enrolled in government primary schools in Delhi. Most of these children come from households where parents have low levels of literacy. In a typical household, the mother’s education extends up to Grade V (primary level), while the father has completed Grade VIII. Most participating households earn between INR 10,000 and 20,000 per month.

The partner organization has a distinctive intervention to strengthen foundational literacy and numeracy among children in primary grades. Field workers approach parents of children enrolled in government schools, either through partnerships with schools or door-to-door outreach in specific neighbourhoods. With parental consent, families are enrolled on a platform through which they receive monthly phone calls for assessing the child’s progress and level appropriate learning material on a WhatsApp chatbot managed by the organisation.

Once the parent is enlisted, they receive a call to complete the telephonic assessment of their child in Mathematics and Hindi. The assessment is administered by a trained caller using a context-specific and widely recognized assessment tool. The caller communicates the results to the parent and shares them over WhatsApp. Subsequently instructional material tailored to the child's learning level is also shared on WhatsApp. Parents are expected to show this material to their child and facilitate their learning at home. After a month, the same caller reassesses the child, and the process continues until the child achieves foundational literacy and numeracy.

Research Questions

One of the central challenges identified by the organisation was that households dropped out before children attained Foundational Literacy and Numeracy (FLN). The primary difficulty lay in scheduling: both parent and child needed to be available at the same time for the child to access the phone. In practice, this was often not possible. Calls had to be rescheduled repeatedly, and completing even one assessment required multiple attempts by the organization's callers. In cases of repeated non-availability, households were eventually dropped from the programme.

As a result, fewer children reached the expected levels of literacy and numeracy. To address this, we designed a series of experiments aimed at increasing the proportion of children who completed the required assessments. The experiments evaluated four interventions implemented as slight modification or addition to the base programme - (i) a WhatsApp-based commitment contract to schedule assessments; (ii) digital rewards to motivate parents to complete assessments; (iii) information on class-level assessment completion rates; and (iv) an option to join a WhatsApp group to connect with other participating parents. These experiments in all help us to answer the following research questions.

1. To what extent can additional digital features in a phone-based educational programme improve programme effectiveness?
2. What are the limits of delivering digital programmes in resource-constrained settings?

Methodology and Data Sources.

The four experiments were conducted between 2023 and 2025 as part of an ongoing collaboration with the organisation to evaluate interventions aimed at reducing costs and improving programme effectiveness. Most of the interventions had a duration of approximately one month, corresponding to the typical interval between two assessment calls. The digital rewards intervention was the sole exception which lasted for approximately three months- typical time period required to deliver three assessments.

All participating households were already enrolled in the programme. Most had completed multiple rounds of assessment calls at the start of the experiment, however, their children had not yet achieved FLN. Households were randomly assigned to either a control group receiving the base intervention or a treatment group receiving a modified version of the base intervention. Balance checks were conducted on selected baseline characteristics, including the child's grade level and their Hindi and Math learning levels. The samples were balanced across all three variables.

We estimate the intent-to-treat (ITT) effects of these interventions using the following empirical specification: $Yi=\alpha+\beta 1Ti+ \gamma\ Xi +\varepsilon i$

where Y_i denotes the outcome variables: (i) whether the subsequent assessment was completed; (ii) whether the subsequent assessment was completed in one call; and (iii) whether the subsequent assessment was completed in one or two calls. Ti is an indicator equal to 1 if household i is assigned to the treatment group and 0 otherwise. Xi is a vector of baseline covariates.

Key Findings and Results

Overall, only one intervention, the WhatsApp-based commitment contract, generated significant efficiency gains. Parents were more likely to complete the assessment on the first call attempt. This reduced the time and effort required by agents to schedule assessments.

The remaining interventions did not produce statistically significant effects on parental engagement. They neither increased completion of the subsequent assessment call nor reduced the number of calls required per completed assessment. These findings highlight the limitations of purely digital interventions, particularly in resource-constrained settings. Organisations often assume that incremental additions, such as WhatsApp groups or digital nudges, will enhance engagement. However, systematic testing yielded limited benefits.

Policy Implications and Relevance to NEP (2020)

These results suggest that underlying barriers are more fundamental. Many parents face low digital literacy and limited time and cognitive capacity to process information delivered through messaging platforms. Providing multiple simultaneous prompts may increase the likelihood that messages are ignored or overlooked. Anecdotal evidence from the experiments indicates that parents value one-to-one phone interactions more than digital communication. Programmes may therefore benefit from finding ways to deliver personalised human engagement at lower cost over additional digital features.

The study also highlights the operational challenges of conducting A/B experiments within non-governmental organisations that face resource constraints and lack advanced technological infrastructure. Technical limitations can compromise the fidelity of digital programme delivery. For example, WhatsApp messages frequently failed to reach a substantial proportion of parents due to changes in Meta's platform policies. In addition, organisations often lack the technical capacity to systematically track engagement-mediating metrics, such as click-through rates on content shared via WhatsApp. Together, these results highlight the limitations of digital interventions when programmes are delivered remotely or through phones.

References:

Cardim, J., Molina-Millán, T., & Vicente, P. C. (2023). Can technology improve the classroom experience in primary education? An African experiment on a worldwide program. Journal of Development Economics, 164, 103145.

Duflo, A., Kiessel, J., & Lucas, A. M. (2024). Experimental evidence on four policies to increase learning at scale. The Economic Journal, 134(661), 1985-2008.

Ministry of Education, Government of India. (2020). National Education Policy 2020. https://www.education.gov.in/sites/upload_files/mhrd/files/NEP_Final_English_0.pdf

5. **Strengthening Validity and Reliability of Outcomes Measures in Large Scale Assessments** By Anuradha Ganesan (Independent Consultant)

Extended Abstract:

1. Context and Policy-relevance

The National Education Policy 2020 marks a significant shift in India's education reform agenda, foregrounding efforts towards evidence-based decision-making especially in evaluating system accountability (NEP 2020, Para 4.41, Para 8.7 – 8.10). Central to this vision is the use of large-scale assessments (LSAs), administrative datasets, and psychometric tools to evaluate progress across Foundational Literacy and Numeracy (FLN), school quality, and teacher effectiveness. As per NEP 2020, instruments such as the National Achievement Survey (NAS), state-level FLN assessments, and other program-linked evaluations are to increasingly inform resource allocation and instructional reform, through evidence-backed policies.

NEPs reliance on data-driven governance implicitly assumes that outcome measures used in causal evaluations are reliable, valid, and contextually appropriate. However, evidence suggests that this assumption is frequently violated. Weakly validated assessment tools, inadequate pre-testing and piloting undermine the fidelity of causal inference, leading to distorted estimates of program impact and, in turn, flawed policy learning in large scale evaluations. This risk is evident in the Foundational Learning Study (FLS) 2022 conducted under the NIPUN Bharat Mission, where cross-language comparability in early literacy outcomes is implicitly assumed without explicitly evidencing how test items function across languages. Languages included in the study differ substantially in script, orthographic depth, and morphological structure. These factors are known to influence early reading acquisition and perceived item difficulty. In this case, observed differences in minimum proficiency may reflect linguistic or script-based effects rather than variation in learners' reading ability. With such comparisons there is a risk of scores being misinterpreted as indicators of system performance or instructional effectiveness, potentially leading to misplaced policy attention and inequitable resource allocation. This study aims to highlight the importance of rigorous pretesting and piloting in surfacing potential risks to quality of the assessment and of the importance of triangulation in ensuring construct validity and defensible causal inference.

2. Research Questions:

1. What constitutes a comprehensive and methodologically sound piloting process for

language and literacy assessments such that item functioning is reliable and construct-validity is ensured?

2. How can findings from pilot studies be triangulated across quantitative item analyses, qualitative response evidence, and instructional context data to refine assessment items and strengthen construct validity?

3. Methodology & Data Sources

This qualitative research study launches a design-based inquiry into how piloting classroom-level summative assessments can generate evidence related to item validity, cognitive demand, and testing conditions prior to operational use. Two summative assessment instruments were developed and deployed in a middle-school classroom of 22 students at a private school in Bangalore to be validated. Design rationales were tied to an assessment framework developed at the beginning of the process. One assessed prose comprehension and the other poetry interpretation. Each instrument included a combination of multiple-choice and constructed-response items, designed to elicit varying levels of cognitive processing aligned with curricular expectations. The pilot was conducted under typical classroom testing conditions to reflect authentic administration contexts.

Multiple qualitative data sources were used to examine assessment quality through triangulation.

A. Design and deployment of one prose-based test instrument and one poetry-based test-instrument.

B. In- classroom observations were conducted during test administration to assess the adequacy of time allocation, particularly for higher-order interpretive tasks.

C. Student responses were used as artefacts and analysed to identify patterns in interpretation, reasoning strategies, and sources of confusion.

D. Structured post-assessment FGDs were held to elicit learner perspectives on item clarity, language accessibility, response formats, and perceived difficulty. Emergent insights were captured on teacher-researchers' reflective notes.

A qualitative research method was deliberately employed as a quality assurance tool, enabling fine-grained examination of how items functioned in practice and informing iterative refinement. This approach demonstrates how classroom-level piloting can produce robust validity evidence relevant to both summative assessment design and large-scale evaluation contexts.

4. Key Findings:

With respect to cognitive process validity, findings from the poetry question paper were especially instructive. The paper comprised two multiple-choice questions and one constructed-response item, with an allotted time of 15 minutes. However, the cognitive demands of the poetry text requiring interpretation of implicit meaning, and the unfamiliarity of the constructed-response task [Figure 1] for the test-takers resulted in longer response times than anticipated. This misalignment highlighted a gap between the intended cognitive processes and the practical time constraints imposed, thereby necessitating reconsideration of both task design and timing.

Piloting also revealed issues related to item clarity and formatting. In one multiple-choice item in the prose question paper, students reported significant confusion. Upon review, this was traced to the use of a sequencing-based numbering format that was uncommon in language assessments. This insight underscored how seemingly minor design choices, such as item formatting conventions, can substantially shape test-taking experiences, influence comprehension, and potentially increase learner anxiety. Student feedback further contributed to improving item quality and distractor effectiveness. During post-pilot discussions, students suggested the inclusion of an additional distractor in an assertion–reasoning item and pointed out instances where distractor wording closely mirrored the contextual language of the passage, leading to unintended selection by a notable proportion of test-takers. The precision with which students articulated these concerns prompted reflection on the potential role of learners as active contributors in various stages of criterion-referenced assessment design and review. Student responses further helped us triangulate our understanding of instructional complexity of test items and allowed us to design scoring rubrics accounting for a larger range of responses.

5. Policy Implications & Relevance to NEP 2020:

The Standards for Educational and Psychological (2014) treat validity and reliability as key pillars of test quality assurance, especially when scores are used for accountability. also suggest that validity as an argument, must be planned in advance and cannot be retrofitted after scores come into existence. Ensuring validity and reliability in LSAs is a complex proposition and conducting effective pretrials and pilots play a role beyond being good practices in test development. They provide opportunity for analyses of test items for evaluating test quality and appropriateness prior to large-scale use.

These findings underscore the importance of the cumulative outcome of deliberate design, empirical checking, and iterative refinement to ensure assessment quality. In the context of NEP 2020's emphasis on evidence based governance, strengthening the validity and reliability of outcome measures must be treated as a foundational design requirement rather than a technical afterthought. Without such deliberate attention, large scale evaluations risk producing precise but misleading estimates, undermining both policy learning and public trust in educational reform.