WHAT WE LEARNED: APPLYING BEHAVIORAL SCIENCE TO IMPROVE FOUNDATIONAL LITERACY AND NUMERACY

ideas⁴²

Authors: Kate Rinehart-Smit (ideas42), Sharon Schroen (VVOB), Stefaan Vande Walle (VVOB), Dr. Sneha Shashidhara (Centre for Social and Behaviour Change), Akashi Kaul (Central Square Foundation), Morgan Kabeer (Busara), Varsha Ashok (Busara), Mamadou Amadou Ly (ARED), Sihle Bikitsha (Funda Wande)

From 2022–2025, the Gates Foundation funded a cohort of five behavioral science projects focused on improving foundational literacy and numeracy (FLN) outcomes. Each project took a unique approach, but all shared a common goal: to understand how behavioral science could help teachers adopt or more consistently implement effective teaching practices, ensuring these practices were embedded into daily routines with a focus on fidelity and feasibility. The cohort designed and tested low-cost solutions aimed at driving change not through overhauls, but through small, sustainable shifts in teacher behavior.

In July 2025, members of the behavioral science for FLN cohort came together for a closed reflection session to candidly discuss what worked, what didn't, and what they'd recommend for future efforts. The session included representatives from four of the five projects and seven of the nine organizations involved. What followed was an honest and thought-provoking conversation about what worked, what didn't, and what we'd recommend for the future.

Lessons and learnings

1. Teachers are motivated and are central to the work: Across all projects, one message rang clear: teachers care. They want to help students learn, and they're open to trying new instructional practices. But systemic and behavioral barriers often get in the way. Some teams found that tools like learner progress trackers and feedback from coaches helped reinforce motivation. Equally powerful was engaging teachers in the research process itself. From in-depth interviews to large-scale quantitative surveys to co-design workshops, teachers were eager to share their experiences. This participatory approach not only led to richer insights but made the work more rewarding for researchers and implementers.

2. Mid-level system actors matter: Teachers did not operate in isolation. Their practices are influenced by coaches, district officials, subject advisors, inspectors, and others in the "middle" of the education system. Teams found that if these mid-level actors weren't aligned with or supportive of an intervention,

it was unlikely to stick. Treating the middle-tier as change agents themselves is essential in reinforcing and scaling teacher behaviour change.

3. Deep diagnosis pays off: Rich and multi-method diagnosis including classroom observations, in-depth interviews, focus group discussions and large-scale surveys uncovered deep and rich insights, around gender disparities, correlations between specific behavioral characteristics and student outcomes and uncovered other contextual and behavioral barriers. These insights shaped not only teams' behavioral interventions but broader aspects of their program design.

4. But rushed designs can undermine rich diagnosis: Despite the value of deep diagnosis, some teams felt pressured to move too quickly into the design phase due to tight timelines or budgets. This limited their ability to iterate, user-test, and co-design with teachers, and saw these as missed opportunities that could have strengthened implementation

5. Embed, don't add on: Teachers are more likely to adopt and sustain new practices when they are integrated into existing systems whether within a government system or NGO-led program. Embedding behavioral tools into materials, routines, or training that teachers were already using is key. Teams found that interventions seen as "add-ons" or those that were misaligned with system, official curriculum or program priorities were less likely to stick.

6. Evaluation expectations should fit the work: While rigorous evaluation is important, many teams found the pressure to conduct randomized controlled trials (RCTs) constraining. It limited the behavioral challenges they could tackle, introduced data collection hurdles, and left little room for iteration. Spillover effects between treatment and control groups also made it difficult to isolate impact in some contexts. Going forward, teams recommend more flexible, fit-for-purpose evaluation approaches that are tailored to the realities of behavioral design and of the sample size and budget available.

7. Measuring teacher behavior change is tricky: Not all good teaching shows up in a checklist. Many current data tools and the training behind them are built to measure whether teachers comply with teacher guides. However, good teaching is not merely about following the teacher guide word-for-word or exercise-by-exercise but also about understanding and adapting it to meet students' needs. Teams emphasized the need to pair quantitative tools with qualitative methods to capture the complexity of

how and why teachers implement new practices. Otherwise, we risk missing meaningful behavior changes.

8. A behavioral lens adds real value: Across the board, teams agreed that behavioral science helped them define problems more clearly, design practical interventions, and understand the deeper drivers of behavior. It provided a structured way to move beyond surface-level explanations and get at the systems, habits, and constraints that shape teacher practice. Importantly, the behavioral insights generated often extended beyond the interventions themselves, informing broader strategies for program implementation and adaptation. Teams without behavioral science expertise or partners mentioned the challenge of conducting behavioral research without this expertise.

9. But not all problems are behavioral: While powerful and important, behavioral science alone cannot resolve the full set of challenges in foundational learning. Structural barriers such as overcrowded classrooms, special classes (two-track or multi-grade), teachers with too much on their plate, absence of adequate teacher training, student and teacher absenteeism or curriculum misalignment can limit the effectiveness of even well-designed behavioral interventions. Behavioral science may be most effective when integrated into broader program and system design and reforms. Integrating behavioral insights into broader system and program design is key. So is aligning expectations about what behavioral science can (and can't) do.

10. Peer learning is energizing: Teams found it deeply valuable to be part of a collective learning effort. The structure created a sense of camaraderie, validation, and shared purpose. But many wished it had started sooner. Earlier opportunities to exchange ideas and troubleshoot challenges could have strengthened the work even further.

Recommendations for funders supporting behavioral science work

1. Co-define challenges and learning goals upfront: Foster early alignment between funders, implementing partners, and researchers to co-define the behavioral problem and clarify strategic learning goals. This ensures that projects are both contextually grounded and positioned to generate insights relevant to the broader sector, beyond a single program or geography.

2. Treat frontline actors as co-designers, not just end-users: Whether in education, health, or other domains, projects benefit from engaging frontline workers as active contributors in research and design. Their first-hand experiences surface valuable insights, build buy-in, and lead to more feasible and context-sensitive solutions. While this kind of participatory approach is encouraged, it's also important to be mindful of the additional demands it may place on already stretched frontline workers. Their time and contributions should be respected and whenever possible, adequately compensated.

3. Expand focus to mid-level system actors: To support and sustain teacher behavior change, behavioral interventions should also target coaches, subject advisors, inspectors, district officials and other mid-system actors. These individuals shape norms, provide reinforcement, and act as critical levers for broader system change. Consider supporting projects that target behavior change across multiple levels of a system, not just teachers, recognizing the collective influence of actors who sit between policy and practice.

4. Budget for deep diagnosis and iterative design: Applying a behavioral lens requires time for robust diagnosis, multiple rounds of user testing, and refinement. Funders should structure grants to allow for flexible, multi-phase implementation with adaptable timelines and evaluation strategies, rather than assuming a fixed solution or evaluation method from the start.

5. Pair quantitative metrics with qualitative understanding: Teacher behavior change is complex and not always observable through checklists or structured data collection tools. Future projects should include qualitative methods as complementary to quantitative methods to better understand adaptation, fidelity, and underlying motivations.

6. Ensure access to behavioral science expertise: Education implementers saw the value in behavioral science and would like to find ways to continue to integrate it and leverage it in their program design and implementation. However, not all implementers have in-house behavioral science capacity and projects without behavioral science expertise struggled more in early phases. Funders can increase project success by providing access to technical support, through embedded advisors, a shared expert pool, or flexible advisory mechanisms that grantees can access as needed.

7. Build in cohort models and structured peer learning: Cohort-based approaches create opportunities for shared reflection, cross-organizational exchange, and collective problem solving. Funders should intentionally design for peer learning from the outset, through facilitated touchpoints, shared milestones, and learning spaces to help projects move further, faster, and with broader relevance.