Shaping Future Success: Evidence from an Early Childhood Human Capital Formation Intervention

Deepak Saraswat ⓡ [*] Shwetlena Sabarwal ⓡ [†] Lindsey Lacey ⓡ [‡]
Natasha Jha ⓡ [§] Nishith Prakash ⓡ [¶] Rachel Chazan-Cohen ⓡ [‖]

September 20, 2026

Abstract

Nearly 250 million children under five in low- and middle-income countries face developmental deficits despite expanding access to early childhood services. We present evidence from a randomized controlled trial (3,131 children in 201 schools) in Nepal's government school system testing three implementation models that combine classroom quality improvements with parental engagement. Following a 15-day teacher training, schools were randomly assigned to deliver caregiver sessions through teachers alone, teachers with in-class helpers, or an external facilitator for parental engagement. Models preserving the teacher's involvement in both domains raised children's developmental outcomes by 0.16–0.21 standard deviations and improved parental engagement by similar magnitudes, while delegating parental engagement to an external facilitator produced no detectable child development gains and worsened caregiver outcomes. We further characterize this heterogeneity across children's and families' baseline profiles, finding that developmental gains concentrate among the least developed children while the rise in caregiver stress falls on families already most strained. These findings demonstrate that how responsibilities are organized among frontline workers shapes program effectiveness as much as program content.

JEL Codes: J13, J24, I21, I24, O15

Keywords: Early childhood education, Cognitive skills, Non-cognitive skills, Ages and Stages Questionnaire, Child development, Nepal

---

The early stages of this project benefited greatly from the guidance of Nathan Fiala. We thank Mo Alloush, Ashutosh Bhuradia, Alex Eble, Peter Eggen, Willa Friedman, Alejandro J. Ganimian, Pamela Jakiela, Gaurav Khanna, Matthew A. Kraft, Owen Ozier, Aaryav Sharma, and Sandip Sukhtankar and participants at NEUDC 2025 for valuable comments that strengthened the paper. We are grateful to the Early Learning Partnership and the World Bank for generous financial support, without which this work would not have been possible. Excellent research assistance was provided by Vharun Boopathy, Shivanshi Garodia, Neil Oommen, Jae Wook Jung, Mridul Mishra, Nikhilesh Prakash, and Deepika Shrestha. We are grateful to the Center for Education and Human Resource Development (formerly Department of Education, Nepal), Officials in the local governments (municipalities) of Jhapa District, Provincial Education Training Center, Koshi Province, Seto Gurans National Child Development Services, Solutions Consultant, and the Nepal Development Research Institute for their invaluable collaboration in implementing the program and facilitating data collection. Author order was randomized using the American Economic Association's author-randomization tool. We used ChatGPT to assist with grammar checks and minor sentence-level improvements; all analysis, arguments, and interpretations are our own. Ethical approval for the study was obtained from the Nepal Health Research Council (NHRC Approval No. 862/2020) and from the Institutional Review Board of the University of Connecticut. The study was pre-registered with the AEA Trial Registry (RCT ID: AEARCTR-00113181).

[*]Independent Researcher, Inc. Email: saraswat.d@gmail.com

[†]The World Bank. Email: ssabarwal@worldbank.org

[‡]Allegheny County Analytics. Email: Lindsey.Lacey@alleghenycounty.us

[§]The Centre for Advanced Financial Research and Learning. Email: natasha.jha@cafral.org.in

[¶]Northeastern University, BREAD, CESifo, IZA, GLO, HiCN, and CReAM. Email: n.prakash@northeastern.edu

[‖]University of Connecticut. Email: rachel.c.cohen@uconn.edu

1 Introduction

Children's earliest years are foundational for later life outcomes. Early interactions shape cognitive, socio-emotional, and physical development (Phillips and Shonkoff, 2000). Yet millions of children in low- and middle-income countries (LMICs) begin school without these foundations, limiting their ability to benefit from education and economic opportunities. Nearly 250 million children under age five in LMICs are at risk of not reaching their developmental potential due to poverty, malnutrition, or inadequate learning environments. This number remains high despite increased service provision (Black et al., 2017). These early deficits compound across the life course, reducing educational attainment, earnings, and health while increasing criminality (Brutti and Montolio, 2021; Campbell et al., 2014; Heckman and Karapakula, 2019). Addressing them requires strengthening the two primary contexts for early learning: homes and classrooms. A growing body of evidence shows that responsive parenting combined with enriched early learning environments produces stronger developmental gains than either alone (Jeong et al., 2021; Shonkoff and Fisher, 2013). But programs that target both contexts simultaneously place substantial demands on frontline teachers, and how to organize such delivery effectively remains poorly understood. Using a randomized controlled trial in Nepal, we study whether program effectiveness depends on how responsibilities for classroom instruction and parental engagement are divided between teachers and support staff — holding program content constant while varying only the organizational structure of delivery.

As preschool enrollment expands in LMICs, concerns about quality are becoming increasingly urgent (Masino and Niño-Zarazúa, 2016). Governments recognize the long-term value of early childhood education for human capital development and are increasing investments in pre-primary programs (Jakiela et al., 2024; McCoy et al., 2017). Yet fewer than one in three children aged 3–4 attend formal programs, and pre-primary education receives less than 2 percent of education budgets (World Bank and UNICEF, 2020a). Program designs often import standards from high-income settings that fit poorly in resource-constrained contexts where classes are large and cultural practices differ (Alan, 2025; Chen and Wolf, 2021). Without addressing these design and resource constraints, expanding access risks repeating the trajectory of universal primary education, where enrollment

gains and effort-focused interventions that leave delivery structures unchanged both failed to translate into learning improvements (Islam et al., 2024; Pritchett, 2013).

Delivering quality at scale hinges on two challenges: how frontline workers are trained, and how their responsibilities are organized. In LMIC education systems, teachers are often underprepared and must balance instruction with administrative duties, home visits, and community engagement, typically with minimal external support (World Bank and UNICEF, 2020a). These competing demands can compromise both classroom quality and family engagement. Such constraints are especially acute in Nepal, where early childhood classrooms are typically staffed by a single, under-trained facilitator responsible for instruction, family engagement, and routine administration alike. As access has expanded, teacher preparation and staffing have become the binding constraints on turning enrollment into learning.[1] Yet, despite growing evidence that specific organizational inputs shape returns in early childhood settings (Walters, 2015), how responsibilities should be allocated among staff, and how that allocation affects coordination, has received limited empirical attention.

We evaluate an intervention in Nepal designed to address both constraints: it combines teacher professional development with parental engagement, and compares three implementation models that vary whether the teacher's combined role across classroom and home domains is preserved, augmented through workload relief, or partitioned between specialized staff. Each arm targets one of the two constraints described above: all arms address the preparation gap through a common training program, while the two support arms address the resource and workload gap in different ways — one by adding capacity within the teacher's existing role, the other by moving a function out of it. The design holds program content constant while creating exogenous variation in organizational structure.

We implement this comparison through a randomized trial in Jhapa district, Nepal, assigning 201 schools to one of three treatment arms or a control group. All teachers in treatment schools

[1]We document these constraints for Nepal's Early Childhood Education (ECE) setting in Section 2 in detail. Briefly, most ECE centres in Nepal are run by one facilitator who combines teaching with management, parental engagement, and administration; nearly one-third of these centres exceed the government's recommended 20:1 student-teacher ratio, and only about 37% had received more than five days of ECD training (Buck et al., 2020).

completed a 15-day professional development program: ten days focused on pedagogy, classroom practices, and national early childhood education standards, followed by five days training them to lead parental education sessions on health, nutrition, hygiene, disability awareness, and cognitive and socio-emotional stimulation. Teachers were then expected to deliver 20 two-hour sessions over six months. Treatment Arm 1 (T1, 50 schools) implemented this model with teachers working alone. Treatment Arm 2 (T2, 50 schools) added an in-class helper on a seven-month contract to assist with classroom management. Treatment Arm 3 (T3, 50 schools) assigned caregiver sessions to an external facilitator trained on the same curriculum, allowing teachers to focus on classroom instruction. The remaining 51 schools served as controls.

Whether to relieve the teacher by adding capacity to their role or by reassigning one of their functions is a question about the division of labor, and the theory offers competing predictions. Specialization may increase efficiency through focus, but it may also weaken value if the tasks being separated share an information base (Becker and Murphy, 1992; Holmström and Milgrom, 1991; Milgrom and Roberts, 1990). In education, the teacher who observes children daily is often the same person advising parents. These predictions pull in opposite directions, making the tradeoff between workload relief and coordination loss fundamentally empirical.

We measure two types of outcomes. Our primary outcomes are three, namely child development, measured with the caregiver-completed Ages and Stages Questionnaire (ASQ); parental engagement, measured by an index of stimulation and learning activities in the home (PESL); and caregiver stress, measured by a standard parental stress scale. We also measure a set of intermediate outcomes that help trace the channels behind these effects, namely classroom quality, captured through structured observations of the learning environment and of teacher–student interactions, and caregiver attitudes and beliefs toward disability and physical discipline. We describe each measure in Section 4.

Our main empirical results suggest that organizational structure shapes program effectiveness as much as program content. Models preserving the teacher's involvement in both classroom instruction and parental engagement (Training Only and Training with Helper) produce statistically equivalent child development gains of 0.21 and 0.16 standard deviations respectively. The model that delegates

parental engagement to an external facilitator (Training with external Facilitator) shows no detectable child development effects (0.08 SD, $p = 0.20$) and raises caregiver stress. The estimated marginal effect of each staffing model over training alone confirms this pattern: adding a helper is neutral (-0.053 SD, $p = 0.38$), while adding a facilitator significantly reduces child development gains (-0.139 SD, $p = 0.02$) and raises stress (0.117 SD, $p = 0.03$). All three arms improve classroom environments and teacher-student interactions to a similar degree. But only models maintaining teacher-family continuity translate these improvements into developmental gains. This indicates that improved pedagogy may be necessary but not sufficient without coordination between what the teacher observes at school and what families do at home.

These results are consistent with several mechanisms that we document. First, directly-observed teacher-student interactions improve uniformly across all three arms, yet child development gains emerge only where teachers also maintain family engagement. This suggests that improved classroom pedagogy translates into developmental gains only when teachers coordinate across both domains. Second, the helper model generates higher parental attendance at education sessions, yet this does not produce proportionally larger child development effects relative to training alone. The teacher's coordinating role contributes independently of engagement dosage. Third, caregiver stress rises only in the facilitator arm — the sole model where families interact with someone other than their child's teacher. This is consistent with the teacher-family relationship providing coherence that an external worker does not. The facilitator holds formal responsibility for parental education sessions but lacks the informational basis to personalize advice (Aghion and Tirole, 1997). Fourth, although individual comparisons between Training with Helper and Training with external Facilitator generally lack statistical power, the helper model produces higher point estimates than the facilitator on nearly every outcome domain. This suggests that capacity augmentation within the existing role structure differs from reassignment of a relational function to a separate worker.

Heterogeneity analysis reveals that the three models reach different children and families. We estimate child-specific effects using a causal forest, which identifies the baseline characteristics that predict larger or smaller effects without a pre-specified subgroup split. Developmental gains are

largest for children with lower baseline development, and the increase in caregiver stress under the facilitator model is concentrated among families already reporting high stress at baseline. This heterogeneity is informative for targeting such programs and for choosing delivery models that do not impose costs on the most vulnerable families. Characterizing this heterogeneity is important for understanding the welfare implications of this program. Returns to early investment depend on the child's starting point (Aizer and Cunha, 2012; Johnson and Jackson, 2019); therefore, identifying which children and families gain most, or gain least, allows a deeper insight into the return on investment for these programs, especially in resource-constrained settings like Nepal.

The intervention improved developmental outcomes by 0.13–0.21 SD and parental engagement by 0.11–0.16 SD at a cost of USD 21 per child. These magnitudes align with a cross-country synthesis finding average cognitive gains of 0.15 SD across more than 50 studies (Holla et al., 2021). They are comparable to other early childhood interventions in low- and middle-income countries, including teacher professional development in Colombia (0.16–0.30 SD; Andrew et al., 2024), a group-based parenting program in Rwanda (0.30–0.38 SD; Justino et al., 2023), and home visiting in urban India (0.30 SD; Andrew et al., 2020). Our estimates exceed access-only programs (Blimpo et al., 2022) and transfer-based models (López Bóo and Creamer, 2019) but fall below the most intensive interventions (Dean and Jayachandran, 2020; Ganimian, Muralidharan, and Walters, 2024). Magnitudes depend on the quality of care available to control children, which varies substantially across settings (Kline and Walters, 2016). The training-only model achieves the largest child development gains (0.21 SD) at the lowest per-child cost, suggesting feasibility for scaling within government systems.

We make three main contributions. First, we provide experimental evidence on the consequences of task specialization in frontline educational service delivery. Our experimental design is novel in that it creates exogenous variation in whether a complementary task — parental engagement — is performed by the same teacher who observes children daily or by a separate specialist, and whether routine classroom burden is reduced through an in-class helper. All program content is held constant. No prior study has randomized organizational structure in this way within an education

program. Organizational theory predicts that tasks sharing an information base should be bundled (Becker and Murphy, 1992; Holmström and Milgrom, 1991), but experimental tests are rare because organizational structures are difficult to randomize, especially in LMIC settings. Our results are consistent with this prediction: separating the parental engagement function from the classroom role eliminates child development gains entirely. Preserving the teacher's bundled role, with or without routine support, produces significant effects.

Second, we extend the education literature on input complementarities and staffing design. Prior work has shown that inputs are complementary between categories: textbooks and teachers together (Mbiti et al., 2019), pedagogical supports and teacher knowledge together (Ganimian, Mbiti, and Mishra, 2025). Adding workers can improve outcomes when their roles complement existing staff (Duflo, Dupas, and Kremer, 2015; Ganimian, Muralidharan, and Walters, 2024). We show that complementarities also exist within a single teacher's task portfolio. The teacher's classroom instruction and parental engagement share an information base that is destroyed when the tasks are separated. This extends the education specialization literature (Fryer Jr., 2018) from subject-matter specialization — where complementarity costs may be modest — to functional specialization involving relational, non-routine tasks (Acemoglu and Autor, 2011), where the costs are substantial.

Finally, we also contribute to the literature on teacher professional development in early childhood settings by providing large-scale evidence from a government school system. A 15-day training program embedded within the public system can shift teacher-student interactions, reducing punitive discipline and increasing warmth and engagement, even in resource-constrained schools where prior evidence is limited (Araujo et al., 2016; Kraft, Blazar, and Hogan, 2018; Wolf et al., 2019).

Our findings speak directly to policy debates about scaling early childhood services without sacrificing quality. Evidence from sub-Saharan Africa illustrates the risks of expanding access to low-quality provision (Blimpo et al., 2022). Successful programs in Colombia and India highlight the importance of teacher training and teaching quality for translating access into learning gains (Andrew et al., 2024). We contribute to this debate by showing that institutional design choices shape program effectiveness as much as training content or curriculum design. Whether teachers

remain central to both classroom instruction and parental engagement, or whether these functions are divided, matters greatly. Support that reduces routine burden while preserving the teacher's relational role enhances or maintains effectiveness. Support that reassigns a relational function to a specialist risks destroying the coordination between school and home that makes integrated programs productive. The organization of labor at the frontline — which tasks go together, which can be separated, and at what coordination cost — is itself a first-order policy lever for program effectiveness (Becker and Murphy, 1992; Garicano, 2000).

2 Background

2.1 Context

This study takes place in Jhapa District in southeastern Nepal, the country's easternmost district (see Figure A1), with a population of approximately one million.[2] The district has a young population: over 25 percent of residents are children under age 15, making early childhood investments particularly relevant as this cohort enters school and eventually the labor force. We work across fourteen municipalities within the district, partnering with the local government education system to implement the intervention in 201 government preschools.

Nepal has substantially expanded early childhood education access over the past two decades. ECD enrollment rose from 12 percent in 2000 to 86 percent in 2017, with over 36,000 centers serving nearly one million children (Poudyal, Banskota, and Khadka, 2019; World Bank, 2020). The share of first-graders with prior ECD experience doubled from 33 to 66 percent between 2007 and 2017. Despite constitutional guarantees of early childhood development rights and government commitments to universal access by 2030, this expansion has occurred with minimal resources: ECD spending remains below 0.1 percent of GDP and accounts for only 1–3 percent of the education budget. Quality outcomes reflect this under-investment. In 2014, fewer than one-third of Nepali children aged 36–59 months were developmentally on track in literacy and numeracy, and only 64

[2]Central Bureau of Statistics Nepal (2021).

percent met benchmarks in socio-emotional development (Poudyal, Banskota, and Khadka, 2019).

This gap between expanding access and lagging quality is the central challenge our intervention confronts, and it stems from constraints on the two inputs that most directly shape what children experience in the classroom: the preparation of the teachers who deliver early education, and the resources available to support them. We describe each of these below.

2.2 The Training Gap

Nepal's rapid expansion of ECD access has not been matched by a corresponding investment in the teachers (called *facilitators*) who deliver it. Facilitators are required to have completed only grade 10, and the intensity of their preparation remains low: as of 2018/19, only about 10 percent of ECD facilitators had received the full one-month training, with most having attended only short courses (World Bank and UNICEF, 2020b).

Contexually closer to our study sample, our own prior survey of educators in eastern Nepal (Buck et al., 2020) documents the depth of this gap: only 14 percent of teachers held a bachelor's degree, fewer than 7 percent had any ECD certification, and only about 37 percent had received more than five days of ECD training. Crucially, we also found that the barrier was not teacher motivation but a lack of resources and training. As a result, many facilitators do not place enough weight on the behaviors that matter most for early cognitive, language, and social development, even as they remain the primary adults responsible for promoting it.

Nepal's existing training system has largely struggled to close this gap. A large-scale evaluation of a government teacher training program in Nepal found little or no effect on student learning, underscoring that simply providing more of the same training is insufficient (Schaffner, Glewwe, and Sharma, 2025). Many parents, in turn, also have limited knowledge of early stimulation practices, compounding the information gap between school and home.

2.3 The Staffing and Workload Gap

Low- and middle-income countries often face tight resource constraints not only in terms of educational facilities and infrastructure but also in terms of human resources—the teachers who deliver the educational content. Student enrollment expansions have outpaced workforce investments (Bold et al., 2018), and teachers in these regions often juggle instruction with many non-teaching tasks, spreading staff thin across responsibilities (World Bank and UNICEF, 2020a). This constraint becomes tighter in the context of ECE, where the recommended class size for the most effective functioning of programs is quite low. E.g., the National Association for the Education of Young Children (NAEYC), recommends a student-teacher ratio of 10:1 and in-class support for teachers of four- and five-year-olds (National Association for the Education of Young Children, 2025). Additionally, a meta-analysis of child-staff ratios finds that effective early childhood programs average roughly nine children per adult (Vermeer et al., 2016).[3]

The Government of Nepal, by contrast, sets a target of 20:1 as the maximum student-teacher ratio in ECE centers (World Bank and UNICEF, 2020b), which is double the ideal ratio, even accounting for contextual differences. In reality, this target is often not met: 41 percent of ECE centers in Nepal exceed the 20:1 ratio, with an average of about 27 students per center, almost always supported by a single teacher (World Bank and UNICEF, 2020b). Our own study sample in the Jhapa district corroborates this pattern. Using the enrollment data we collected for the baseline survey, we find an average of 19 students enrolled across the 201 ECE centers, with 92 percent of schools staffed by a single teacher. More importantly, roughly one-third of the sample schools exceeded the 20:1 ratio with an average of nearly 34 students per teacher. Although enrollment can change after the initial roster is recorded, the very low attrition at endline indicates that the number of students did not fall—and since it is common for enrollment to swell as the school year progresses, our reported figures are best interpreted as a lower bound on the true class sizes managed by these facilitators.

Compounding this staffing constraint, the single facilitator carrying this load is also poorly compensated. The monthly salary of an ECD facilitator is about NPR 6,000—roughly 23 percent

[3]Nepal's teachers also face documented occupational health consequences of these working conditions, including high rates of burnout (Paudel et al., 2024) and voice disorders linked to large class sizes (Phuyal, Shrestha, and Bhattarai, 2024).

of a primary school teacher's pay and less than half the prevailing minimum wage—and annual turnover among facilitators is around 20 percent (World Bank and UNICEF, 2020b). These under-paid, under-trained facilitators are simultaneously responsible for classroom instruction across all developmental domains, parental engagement, and routine administrative tasks, leaving little capacity to do all of them well.

Taken together, these two constraints—a shortage of trained teachers and a workforce stretched thin across too many students and too many tasks—motivate our study design that intervenes on both margins. Our intervention provides structured teacher training to address the preparation gap while embedding parental education within the school system, leveraging teachers' position as trusted community figures to bridge information gaps between classroom and home. At the same time, it tests two distinct models for relieving the teacher's workload—adding support inside the classroom versus offloading a function to a separate worker outside it—allowing us to ask not only whether additional support helps, but how that support is best organized.

We implemented this design in partnership with Seto Gurans, a local nonprofit specializing in early childhood development, and district education authorities in Jhapa. Seto Gurans recruited and trained facilitators, coordinated with schools, and managed program logistics, while we maintained control over random assignment and data collection. This partnership enabled us to test the intervention within Nepal's government school system while ensuring implementation fidelity across treatment arms.

3 Experimental Design and Method

3.1 Intervention Design

The intervention provides teacher training and parental education to strengthen both classroom instruction and home learning environments, following a two-generation approach (Shonkoff and Fisher, 2013). We partnered with Seto Gurans, a local nonprofit specializing in early childhood

development, to design and deliver the curriculum through Nepal's Ministry of Education.[4] Unlike intensive home-visiting models such as Reach Up and Learn (Attanasio et al., 2020; Gertler et al., 2014; Yousafzai et al., 2014), our approach delivers parental education through school-based group sessions led by teachers or school-based facilitators, reducing per-child implementation costs while building on teachers' existing relationships with families.

All teachers received 15 days of training: ten days on developmentally appropriate pedagogy, classroom management, and national ECD standards, followed by five days preparing them to lead parental education sessions. The parental education curriculum comprised 30 topics organized under five components: the role of ECD caregivers, stages of child development, health and nutrition, child protection and care, and early learning. From this curriculum, we selected 18 topics jointly with Seto Gurans as part of the Support for Nepal Early Childhood Development Study. We designed and added two sessions focused specifically on disability awareness and inclusive care. In total, 20 two-hour sessions were delivered over six months in community settings, covering health, nutrition, hygiene, disability awareness, and cognitive and socio-emotional stimulation through play-based activities. To test how implementation models affect program effectiveness, we designed three treatment arms that vary the organization of responsibilities between teachers and support staff while holding training content constant. Full curriculum details appear in Appendix C.

3.2 Treatment Arms

We randomized a sample of 201 schools into the following four arms:

3.2.1 Treatment Arm 1: Teachers Training

In T1 (50 schools), teachers completed the 15-day training program and then delivered all 20 parental education sessions themselves. Sessions were two hours each, held at the school over six months, and attended by primary caregivers of enrolled children. This model places full responsibility for both classroom instruction and parental engagement on teachers, preserving continuity between

[4]Seto Gurans (www.setogurans.org) is a Nepali organization with extensive experience implementing early childhood development programs.

the two domains but risking overload. It tests whether a single teacher can effectively manage both without additional support.

3.2.2 Treatment Arm 2: Teachers Training + Inside Classroom Helper

In T2 (50 schools), teachers received the same training and delivered all 20 parental education sessions, but were supported by an in-class helper who assisted with daily classroom management. The helper absorbed routine tasks, freeing teachers to maintain instructional quality while still leading parental engagement. This model tests whether workload relief improves outcomes without severing the teacher-family connection.

3.2.3 Treatment Arm 3: Teachers Training + Outside Classroom Helper

In T3 (50 schools), teachers received the same training but an external facilitator delivered all 20 parental education sessions. Facilitators were recruited and trained specifically for this parental education role, allowing teachers to focus exclusively on classroom instruction. This model introduces specialization: teachers concentrate on pedagogy while facilitators manage parental outreach. It tests whether dividing responsibilities improves efficiency or weakens coordination between what children experience at school and what parents do at home.

3.2.4 Control Group

The 51 schools in this arm served as the control group and did not receive any intervention in the form of teacher training or parental-education sessions - this group was business as usual.

3.3 Sample Selection, Randomization, Training, and Timeline

Sample Selection and Randomization: This study was conducted in Jhapa district, covering schools from all fifteen municipalities. Randomization was done at the school level: 201 schools with Early Childhood Development (ECD) sections were selected and randomly assigned to one of three treatment arms (50 schools each) or to the control group (51 schools). Figure A2 displays the

geographic distribution of study schools across the district. All children enrolled in the selected schools were part of the intervention. For data collection, however, budget constraints required sampling a subset of students. Using systematic random sampling, we selected 3,131 children across the study schools. Specifically, every fifth child on the admission list was chosen until 20 children were enrolled per school. In schools with fewer than 20 enrolled children, all ECD students were included in the sample. This approach ensured consistent representation across schools of varying sizes while maintaining feasibility for the endline.

Training: All treatment teachers attended a centralized 15-day training program at the Provincial Education Training Center in Koshi Province. The training was scheduled in two windows reflecting teacher availability. The core pedagogical training (10 days, April 3–12, 2022) covered evidence-based instructional strategies, classroom management, and national ECD standards, and was conducted before the start of the new school session. This was followed by parental education delivery training (5 days, May 29–June 2, 2022), which prepared teachers to lead the 20 caregiver sessions in their communities. Teachers in all three treatment arms received identical training content; arm-specific assignments were revealed only after training concluded. Separately, facilitators for T3 schools received dedicated parental education training (June 4–10, 2022) and classroom helpers for T2 schools received orientation training (June 12–14, 2022). All training activities were conducted off-site at centralized venues and did not involve caregivers or children.

Timeline: Figure 1 summarizes the study timeline. Due to post-COVID disruptions, the 2022 school year followed an unusual schedule: examinations and evaluations of the prior-year student cohort continued through end of April, and enrollment and registration drives for new ECE students — the classes targeted by our intervention — extended through end of May 2022. The period from May through June 2022 was primarily administrative — enrollment drives, and registration occupied this window, and regular classroom instruction for the new ECE cohort had not yet commenced. These COVID-related delays in completing prior-year evaluations and enrolling new students meant that actual ECE classes for our study children could only reasonably begin towards the end of June

2022. Survey teams received finalized enrollment lists only by end of May, at which point baseline data collection commenced. Baseline interviews were conducted from June 3 to June 22, 2022, with 3,131 primary caregivers across all 201 schools — during the administrative period before classroom instruction began. The intervention comprised two components: classroom instruction by trained teachers, which began in late June 2022, and 20 parental education sessions for caregivers, which ran from July 2022 through January 2023. Endline data collection took place from February 9 to March 4, 2023, approximately eight months after baseline.

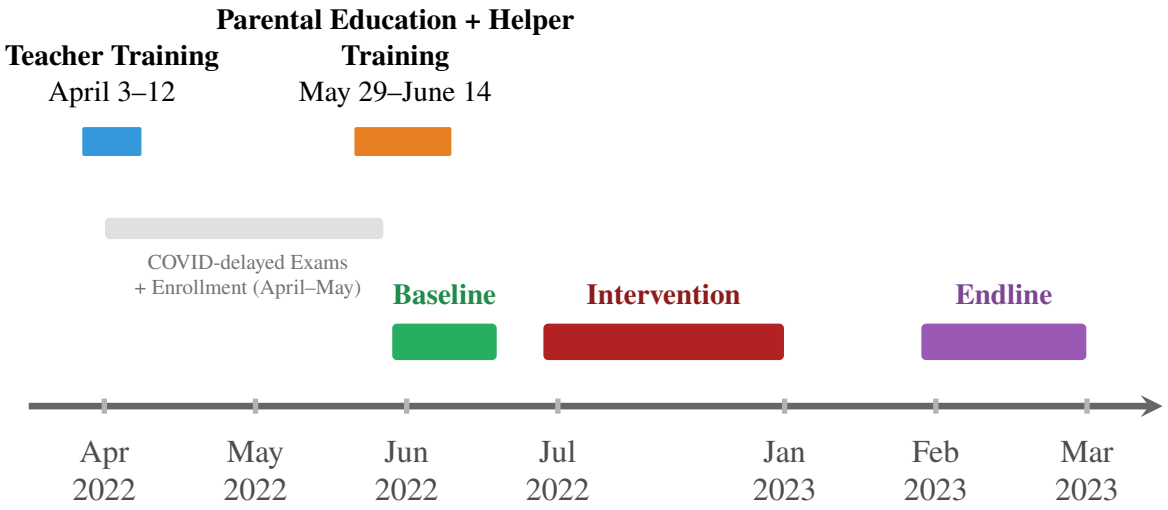


Figure 1: Study Timeline

Notes: This figure shows the study timeline from April 2022 to March 2023. Teacher pedagogical training (TPD) took place April 3–12, 2022, at a centralized off-site venue, scheduled before the start of the new school session. Parental education delivery training for teachers, facilitator training, and helper orientation occurred May 29–June 14, 2022. Between these training windows, schools conducted prior-year examinations and new ECE enrollment drives (April–May 2022); enrollment lists were finalized by end of May. Baseline data were collected in June 2022, during the administrative period before classroom instruction began. Due to post-COVID delays in examinations and enrollment, the regular ECE instruction for study children commenced late-June 2022 and the parental education sessions in July 2022. Endline data collection occurred February–March 2023.

4 Data Collection, Outcome Measures, and Validity of the Experimental Design

4.1 Baseline Survey

Baseline data were collected from June to July 2022 through in-person interviews with 3,131 primary caregivers across all 201 schools. Interviews lasted approximately 33 minutes and were conducted by trained enumerators in the local language. Survey modules covered demographic information, household amenities, health status and beliefs (including disability), parental stress, and caregiving practices. From these data we constructed baseline indices for parental engagement (PESL), child development (ASQ), and caregiver stress, which we use to examine heterogeneity in treatment effects by baseline characteristics.

4.2 Learning Environment and Practices Check

At endline, we assessed classroom learning environments across all 201 schools. Teachers completed in-person interviews covering water and sanitation facilities, lesson preparation, pedagogical practices, physical infrastructure, and classroom organization. From these data we constructed the Learning Environment and Practices (LEP) Index, aggregating indicators of teacher planning, instructional quality, infrastructure, and learning area organization.[5] The LEP Index serves as a first-stage outcome measuring how the intervention affected classroom quality.

4.3 Teacher-Student Interaction

At endline, we complemented survey data with a structured classroom observation to assess teacher–student interactions. Trained enumerators conducted twenty one-minute observations within a two-hour classroom window, each followed by a four-minute period to record assessments. The protocol focused on established best practices for early childhood education. To minimize bias,

[5]Appendix G.2 provides details on the items used in constructing the Learning Environment and Practices Index.

teachers were not informed about the specific aspects being observed or the evaluation criteria. These observations provide a direct measure of the quality of classroom interactions and the implementation of ECD practices in study schools.

4.4 Parental Education Session Attendance

We recorded attendance at all 20 parental education sessions conducted during the intervention period. Attendance data captures which caregivers participated in each session, allowing us to examine participation patterns across treatment arms and explore attendance as a mechanism linking treatment assignment to outcomes.

4.5 Endline Survey

The endline survey was conducted from February to March 2023 with 3,004 primary caregivers (96% of the baseline sample). Interviews lasted approximately 29 minutes. The questionnaire mirrored the baseline design, covering demographic characteristics, household amenities, health status and beliefs (including disability), parental stress, and caregiving practices. From the endline survey we constructed our primary outcome measures: the Ages and Stages Questionnaire (ASQ) Index for child development, the Practice of Early Stimulation and Learning (PESL) Index for parental engagement practices, and the Caregiver Stress Index.

4.6 Outcome Measures

Primary Outcomes. Our three primary outcomes are child development, parental engagement, and caregiver stress.

Child Development: We measure child development using the Ages and Stages Questionnaire, a parent-completed developmental screening instrument originally developed by Squires, Bricker, and Potter (1997); we use the third edition (ASQ-3) documented in Squires and Bricker (2009). The ASQ is designed to assess whether children are meeting age-appropriate developmental milestones

through concrete behavioral items — specific, observable actions that the child either performs or does not (e.g., "Does your child kick a ball by swinging his leg forward?" "After watching you draw a circle, can your child make one?"). The administration protocol instructs enumerators to have the child attempt each activity during the assessment: "Try each activity with your baby before marking a response." Many items contain explicit test instructions ("Ask your child to...," "Do not repeat the numbers," "Give your child two or three tries before marking") that structurally require real-time demonstration. Items are scored on a three-point scale — "Yes" (10 points, performs regularly), "Sometimes" (5 points, emerging), or "Not Yet" (0 points) — anchoring responses to observed behavior rather than subjective impressions.

The ASQ assesses five domains of child development: Communication, Gross Motor, Fine Motor, Problem-Solving, and Personal-Social. Each domain contains six items calibrated to the child's age group (36, 48, or 60 months), targeting skills that are *emerging* at that stage. We construct an aggregate ASQ score summing across all five domains, plus individual domain scores, all standardized relative to the control group mean. Appendix G.3 provides the complete item-by-item content for each domain and age group.

The ASQ is one of the most widely used developmental screening instruments globally, translated into over 20 languages and validated in more than 60 countries across diverse socioeconomic and cultural contexts. Its psychometric properties in the US standardization sample include test-retest agreement of 92%, inter-observer agreement of 93% between parents and trained professional examiners, and concurrent validity of 85% with clinician-administered Bayley Scales (Squires and Bricker, 2009). These properties have been validated across diverse settings including South Asian, Latin American, and East Asian populations (Appendix E). It has been employed as a primary outcome in rigorous economic evaluations of early childhood programs across varied settings including Peru, Rwanda, and India (Araujo, Dormal, and Schady, 2019; Justino et al., 2023; Weaver et al., 2025).

The primary alternative to ASQ — clinician-administered direct assessment such as the Bayley Scales — requires trained examiners and individual sessions of 30–90 minutes per child, costs that

are prohibitive at our sample size and in our setting (Bayley, 2006). The ASQ's parent-completed format captures development across multiple contexts (home, community, school) at a fraction of this cost and without specialized clinical training, while achieving 85% concordance with the clinical gold standard — making it both valid and feasible for large-scale implementation in resource-constrained settings like ours.

Although caregiver-completed, the ASQ's concrete behavioral format and real-time administration protocol make it structurally distinct from attitudinal self-reports. In Section 6.4.1, we directly test whether treatment effects persist on items where reporting bias is structurally precluded.

Parental Engagement: The Practice of Early Stimulation and Learning (PESL) Index captures whether caregivers engage in developmentally supportive activities with their children, including: reading books, telling stories, singing songs, taking the child outside, playing together, and practicing early learning activities such as naming, counting, or drawing. Each activity is coded as a binary indicator (1 if practiced, 0 otherwise), summed, and standardized relative to the control group. We also construct a PESL Intensity Index that measures the frequency of these activities per week (1–3 days, 4–5 days, or 6+ days), which is similarly standardized.[6]

Caregiver Stress: The Caregiver Stress Index is constructed from a seven-item scale assessing: feeling reinforced or supported by the child, emotional connection with the child, alignment between caregiving expectations and reality, interpretation of the child's behavior, and overall confidence in caregiving abilities. Caregivers respond using a five-point Likert scale (1 = Strongly Disagree to 5 = Strongly Agree), with responses summed and standardized relative to the control group.

Intermediate Outcomes. We also measure a set of intermediate outcomes used to trace the channels behind the primary effects, namely classroom quality and caregiver attitudes and beliefs.

Classroom Quality: We assess the intervention's effect on classroom quality through two measures, the Learning Environment and Practices (LEP) Index and the Teacher-Student Interaction Index.

The LEP Index captures classroom features and teacher behaviors reflecting the overall quality of

[6]Appendix G.4 details the variables used in the construction of the PESL and Stress Indices. Appendix G.5 provides details on the PESL Intensity (Intensity of Early Stimulation and Learning) Index.

early learning environments. It combines information across four domains: (i) teacher planning and routines, including preparation of lesson plans, following a daily timetable, taking attendance, and maintaining student records; (ii) pedagogical practices, including regular use of stories, poems, songs, learning materials, and inclusive activities for children with disabilities; (iii) physical environment, including cleanliness, safe drinking water, child-friendly toilets, and adequate seating; and (iv) organization of structured learning areas, including designated spaces for language, math, science, creativity, role play, and construction, use of the ECD kit box, and regular supervision by the head teacher. Each component is measured as a binary indicator, with the total score standardized relative to the control group mean to form the index.[7]

To assess teacher-student interactions, we conducted 20 structured one-minute classroom observations within a two-hour window, followed by four-minute recording periods. From these observations we constructed two indices: the Positive Interaction Index and the Negative Interaction Index.

The Positive Interaction Index captures behaviors that foster an engaging and supportive classroom environment, including: greeting children warmly, reading to groups, facilitating discussions, encouraging student participation, providing individual attention, using positive reinforcement, showing enthusiasm and positive affect, and creating opportunities for peer interaction. The Negative Interaction Index measures behaviors that undermine the learning environment, including: harsh or punitive discipline, yelling or threatening, ignoring student questions or needs, showing frustration or impatience, and failing to intervene in conflicts. Both indices are constructed by aggregating observations across the 20 snapshots per classroom and standardizing relative to the control group mean.[8]

Attitudes and Beliefs: We measure caregiver attitudes toward disability through four questions covering acceptance, inclusion, and understanding of children with disabilities, summed and standardized to form the Attitude to Disability Index. We also assess caregiver confidence in supporting children with disabilities through Likert-scale questions on knowledge and ability,

[7]Appendix G.2 provides details on the items used in constructing the Learning Environment and Practices Index.
[8]Appendix G.2 provides details on the variables used in the construction of the Positive and Negative Interaction Indices.

forming the Confidence Disability Index. Finally, we measure beliefs about physical discipline through questions on whether caregivers view spanking as necessary or normal and whether they reported using spanking in the previous 14 days.[9]

4.7 Validity of the Experimental Design

4.7.1 Sample Characteristics and Baseline Balance

Table A2 presents descriptive statistics and baseline balance across treatment arms. Column 1 reports overall sample characteristics. Children are on average 48 months old (SD 12), with roughly equal shares of boys and girls. Households average three adults and 1.3 siblings. The primary caregiver is female in 84 percent of cases, with average educational attainment equivalent to six to eight years of schooling. Baseline developmental and caregiver practice measures are standardized around zero with considerable variation: the aggregate ASQ score averages 0.04 (SD 0.95), the PESL Index averages 0.04 (SD 0.99), and the Caregiver Stress Index averages 0.07 (SD 1.07). Attitudes toward disability average slightly negative (–0.10, SD 0.93).

Columns 2–5 report means by treatment arm, while columns 6–8 present tests of mean differences relative to the control group. Demographic characteristics are well balanced across treatment arms. Among baseline outcome measures, we observe small differences in attitudes toward disability and the PESL Index. However, there is no consistent pattern across related measures, and the magnitude of these differences is small relative to overall variation. Given the number of comparisons, such differences likely reflect chance rather than systematic imbalance. Overall, the results are consistent with successful randomization.

4.7.2 Attrition and Compliance

Approximately 4 percent of the baseline sample (127 out of 3,131 children) did not participate in the endline survey. Although attrition is modest, we test whether it is systematically associated with treatment assignment or baseline characteristics in Section 6.4.2. Additionally, 20 of the

[9]Appendix G.5 provides details on the variables used in the construction of the Attitude to Disability and Confidence Disability Indices.

150 treatment schools (13 percent) did not implement parental education sessions, resulting in noncompliance. We report intent-to-treat estimates in the main analysis, which are unaffected by compliance rates. In Section 6.4.2, we examine baseline characteristics of noncompliant schools to assess whether noncompliance influences the interpretation of our findings.

4.8 Deviations from the Pre-Analysis Plan

Our analysis follows the Pre-Analysis Plan (PAP) registered at the AEA RCT Registry (AEARCTR-0011318), which specified primary outcomes, index construction, multiple-hypothesis corrections, empirical specifications, and attrition protocols, with minor modifications necessitated by field conditions. Most notably, we exclude baseline controls from main specifications because the COVID-19 pandemic disrupted baseline data collection, leaving substantial missing data that would reduce the effective sample by about 30%. Following McKenzie (2012) and recent guidance from Kerwin, Rostom, and Sterck (2024), we rely on randomization alone to preserve statistical power, as including baseline controls is not required for identification and conventional balance tests are prone to misleading over-rejection. Robustness checks with baseline controls for primary outcomes are reported in Appendix Table A3. We also extend the PAP by disaggregating treatment effects across implementation models ("training only" versus "training plus additional support"), which reveals meaningful heterogeneity (Banerjee et al., 2020), by standardizing variable coding so that positive coefficients uniformly indicate improvements, and by adapting ASQ scoring from narrow age-specific cutoffs to broader developmental categories, enhancing comparability across our heterogeneous sample. Finally, we examine outcomes of direct policy relevance identified during fieldwork, including heterogeneity by age and baseline status. These deviations reflect challenges during the pandemic and field insights while maintaining the integrity of the experimental design.

5 Empirical Specification

We estimate the intent-to-treat (ITT) effect of assigning a trained teacher to ECD classrooms, first by pooling all treatment arms together (Teacher Training, Teacher Training with an Inside Classroom Helper, and Teacher Training with a Facilitator, relative to the control) and then separately by treatment arm. The pooled specification is:

$$Y_{i,s,t=1} = \beta_0 + \beta_1 T_s + Y_{i,s,t=0} + \epsilon_{i,s} \tag{5.1}$$

where $Y_{i,s,t=1}$ denotes the outcome for individual (student or parent) i in school s measured post-treatment, and $Y_{i,s,t=0}$ is the corresponding baseline value. T_s indicates the treatment assignment of school s: schools were randomly allocated to one of three treatment arms or to a control group. In this specification, $T_s = 1$ if school s was assigned to any of the three treatment arms and $T_s = 0$ if assigned to the control group. $\epsilon_{i,s}$ is the error term. Standard errors are clustered at the school level to account for within-school correlation and because treatment was assigned at that level.[10] The coefficient of interest, β_1, estimates the ITT effect of access to a trained teacher in the ECD classroom.

We next estimate ITT effects separately by treatment arm, relative to the control. In this specification (Equation 5.2) , indicators are included for assignment to each of the three treatment arms: Teacher Training (T1), Teacher Training with an Inside Classroom Helper (T2), and Teacher Training with a Facilitator (T3). All other covariates are defined as in Equation 5.1, including baseline outcomes where available. The coefficients α_1, α_2, and α_3 estimate the ITT effects for each treatment arm relative to the control group.

$$Y_{i,s,t=1} = \alpha_0 + \alpha_1 T1_s + \alpha_2 T2_s + \alpha_3 T3_s + Y_{i,s,t=0} + \eta_{i,s} \tag{5.2}$$

Finally, to directly assess the marginal contribution of additional staffing support over teacher

[10] We report the estimates from Equation 5.1 with baseline controls $Y_{i,s,t=0}$ in Appendix Table A3.

training alone, we estimate a nested interacted specification:

$$Y_{i,s,t=1} = \gamma_0 + \gamma_1 \text{Training}_s + \gamma_2(\text{Training}_s \times \text{Helper}_s) + \gamma_3(\text{Training}_s \times \text{Facilitator}_s) + \mu_{i,s} \quad (5.3)$$

where $\text{Training}_s = 1$ for all treatment schools (since all received the 15-day TPD), $\text{Helper}_s = 1$ for T2 schools, and $\text{Facilitator}_s = 1$ for T3 schools. The coefficient γ_1 captures the Training Only effect (T1 versus control). The coefficients γ_2 and γ_3 estimate the *marginal* effect of adding a helper or facilitator over training alone, making the magnitude and statistical significance of between-arm comparisons transparent. We use this specification to test whether additional staffing support enhances, maintains, or undermines the gains from teacher training. All specifications include municipality fixed effects and cluster standard errors at the school level.

6 Results

We first report effects on our three primary outcomes: child development (ASQ), parental engagement practices (PESL), and caregiver stress. We then assess changes in caregiver beliefs and attitudes toward discipline and disability. We present results for pooled treatment effects and separately by treatment arm to compare implementation models. The intervention's effects on classroom environments and teacher-student interactions, which capture how the intervention shaped the learning environment, are examined in Section 7 as channels through which these outcomes were generated.

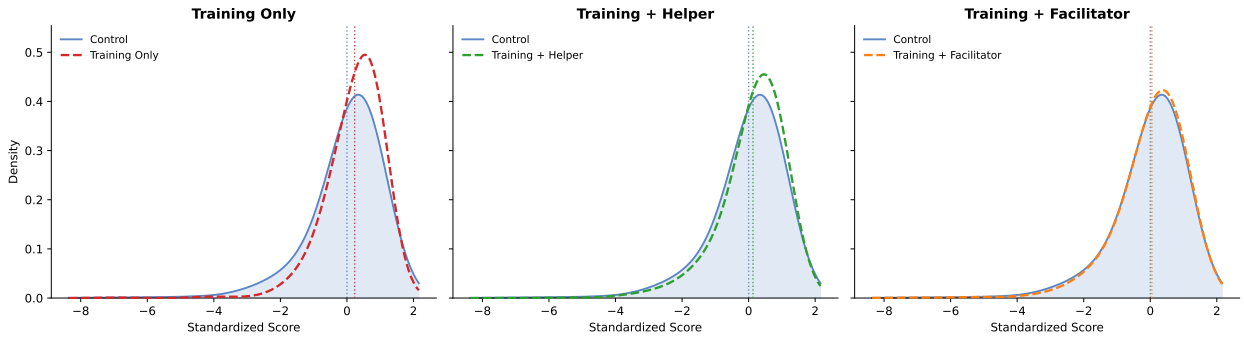
Because we test multiple hypotheses across outcome domains, we report Benjamini-Hochberg False Discovery Rate (FDR) adjusted p-values alongside unadjusted p-values (Benjamini and Hochberg, 1995) in the main outcomes table (Table 1).

6.1 Effects on Children's Development and Parental Engagement

Before turning to the regression estimates, we examine the endline distributions of each outcome by arm. Figures 2–4 plot, for each treatment arm, the distribution of the outcome against the control distribution (shaded), with dashed vertical lines marking the group means. These distributional comparisons preview the patterns that the regressions in Table 1 then quantify.

Figure 2 shows child development (ASQ). In the Training Only and Training + Helper arms, the distribution shifts visibly to the right of control, and the arm mean lies clearly above the control mean. In the Training + Facilitator arm, by contrast, the treatment and control distributions are nearly indistinguishable and their means almost coincide. The visual impression is that the first two models moved children's development while the facilitator model did not.

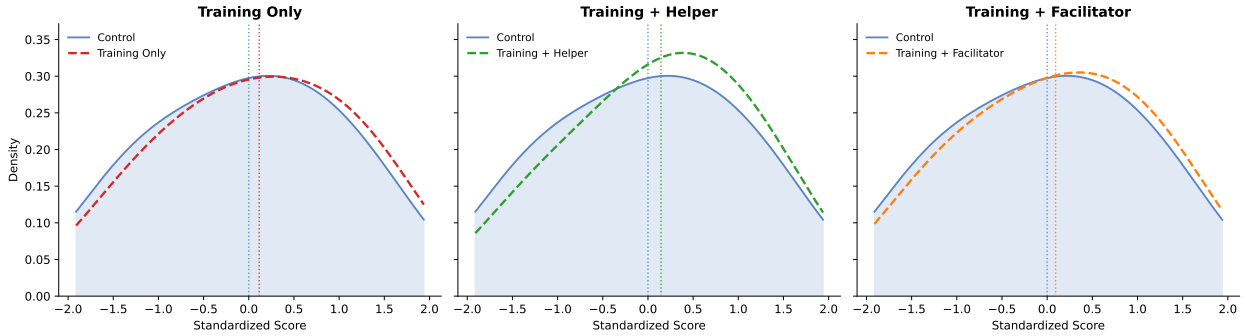
Figure 2: Distribution of Child Development (ASQ) by Treatment Arm



Notes: Each panel plots the endline distribution of the standardized ASQ index for one treatment arm (dashed) against the control distribution (shaded). Dashed vertical lines mark group means. Scores are standardized relative to the control group. Densities are Gaussian kernel density estimates (Scott's rule scaled by a factor of 0.65), applied uniformly across the three outcomes. *Source:* Endline Survey.

Figure 3 shows parental engagement (PESL). All three arms shift the distribution rightward relative to control, but the shift is most pronounced in the Training + Helper arm, where mass accumulates in the upper range of the score. The Training Only and Training + Facilitator arms show smaller, broadly similar rightward movements.

Figure 3: Distribution of Parental Engagement (PESL) by Treatment Arm



Notes: Each panel plots the endline distribution of the standardized PESL index for one treatment arm (dashed) against the control distribution (shaded). Dashed vertical lines mark group means. Scores are standardized relative to the control group. Densities are Gaussian kernel density estimates (Scott's rule scaled by a factor of 0.65), applied uniformly across the three outcomes. *Source:* Endline Survey.

Figure 4 shows caregiver stress. In the Training Only and Training + Helper arms, the distributions track control closely and the means sit at or slightly below the control mean. In the Training + Facilitator arm, the arm mean lies to the right of control, with somewhat more mass in the upper tail, hinting at a modest increase in stress in the arm where an external worker delivered the parental education sessions.

Figure 4: Distribution of Caregiver Stress by Treatment Arm



Notes: Each panel plots the endline distribution of the standardized Caregiver Stress index for one treatment arm (dashed) against the control distribution (shaded). Dashed vertical lines mark group means. Scores are standardized relative to the control group. Densities are Gaussian kernel density estimates (Scott's rule scaled by a factor of 0.65), applied uniformly across the three outcomes. *Source:* Endline Survey.

Table 1 quantifies these patterns. It reports treatment effects on our three primary outcomes: child development (ASQ), parental engagement (PESL), and caregiver stress. Panel A presents pooled treatment effects, Panel B compares training alone to training with support, and Panel C disaggregates by type of support.

Table 1: Treatment Effects on Child Development and Caregiver Outcomes

	ASQ Index	PESL Index	Caregiver Stress Index
	(1)	(2)	(3)
Panel A: Pooled Arms			
Any Treatment	0.15***	0.13**	0.03
	(0.05)	(0.05)	(0.05)
BH-Corrected P-Values	[0.04]	[0.04]	[0.64]
Observations	3,004	3,004	3,004
R^2	0.12	0.08	0.36
Panel B: Training + Helper			
Training Only	0.21***	0.11*	-0.01
	(0.06)	(0.06)	(0.06)
BH-Corrected P-Values	[0.00]	[0.13]	[0.86]
Training + Any Help	0.11**	0.13**	0.05
	(0.06)	(0.05)	(0.05)
BH-Corrected P-Values	[0.11]	[0.04]	[0.41]
Observations	3,004	3,004	3,004
R^2	0.12	0.08	0.36
Panel C: Individual Arms			
Training Only	0.21***	0.11*	-0.01
	(0.06)	(0.06)	(0.06)
BH-Corrected P-Values	[0.00]	[0.13]	[0.86]
Training + Helper	0.16**	0.16***	-0.01
	(0.07)	(0.06)	(0.06)
BH-Corrected P-Values	[0.04]	[0.04]	[0.86]
Training + Facilitator	0.08	0.11*	0.11*
	(0.06)	(0.06)	(0.06)
BH-Corrected P-Values	[0.27]	[0.13]	[0.12]
Observations	3,004	3,004	3,004
R^2	0.12	0.08	0.36
Helper = Facilitator (p-value)	0.18	0.44	0.05

Notes: The table presents results for the main outcomes estimated using regression equation (5.1). Each outcome variable is a standardized score relative to the control group: the *ASQ Index* (Ages and Stages Questionnaire, a measure of child development), the *PESL Index* (Practices of Early Stimulation and Learning, capturing parental engagement), and the *Caregiver Stress Index*. All regressions include municipality fixed effects. *Any Treatment* equals one if a school received *Training Only*, *Training + Helper*, or *Training + Facilitator*, and zero otherwise. Standard errors are clustered at the school level and reported in parentheses. The data comes from the Endline Survey, which includes all 3,004 student respondents. P-values were also adjusted for multiple hypothesis testing using the False Discovery Rate (FDR) procedure Benjamini and Hochberg (1995). These adjusted p-values are reported in brackets under the standard errors. Raw p-values were first ranked in ascending order, then each was multiplied by the total number of hypotheses and divided by its rank. To ensure consistency, adjusted p-values were monotonically corrected by replacing each with the minimum of its own value and all larger-ranked values. The final adjusted p-values therefore increase weakly with rank and represent the standard Benjamini–Hochberg FDR correction. ***, **, and * denote statistical significance at the 1, 5, and 10 percent levels, respectively.

Pooled treatment improved child development by 0.15 SD ($p = 0.01$, BH-adjusted $p = 0.04$) and parental engagement by 0.13 SD ($p = 0.01$, BH-adjusted $p = 0.04$). The intervention did not increase caregiver stress (0.03 SD, $p = 0.55$, BH-adjusted $p = 0.64$), indicating that the program fostered new developmental activities without overburdening families.

Panel C reveals that child development effects differ across implementation models. Both the training-only arm (0.21 SD, $p < 0.01$, BH-adjusted $p = 0.00$) and the helper arm (0.16 SD, $p = 0.02$, BH-adjusted $p = 0.04$) produced significant gains relative to controls. The facilitator arm showed no detectable effect on child development (0.08 SD, $p = 0.20$, BH-adjusted $p = 0.27$), indicating that delegating parental engagement to an external worker eliminated developmental gains despite teachers focusing exclusively on instruction. We assess the precision of between-arm differences formally in Section 6.1.1.

A natural concern is whether treatment effects on caregiver-reported ASQ reflect genuine developmental gains or motivated reporting. We address this in Section 6.4.1, where we decompose the ASQ into indices isolating items requiring physical child demonstration and find that effects persist where reporting bias is structurally precluded.

Parental engagement improved across all three arms. The helper arm produced gains of 0.16 SD ($p < 0.01$, BH-adjusted $p = 0.04$), while both the training-only arm (0.11 SD, $p = 0.07$) and the facilitator arm (0.11 SD, $p = 0.08$) produced modest, statistically similar effects. Despite facilitators specializing in parental education sessions, they were no more effective at improving parental engagement than capacity-constrained teachers working alone, suggesting that classroom context and teacher credibility matter for family outreach.

Caregiver stress showed no significant changes in the training-only (-0.01 SD, $p = 0.82$) or helper arms (-0.01 SD, $p = 0.86$), but increased marginally in the facilitator arm (0.11 SD, $p = 0.06$, BH-adjusted $p = 0.12$). The difference between helper and facilitator arms on stress is marginally significant ($p = 0.05$), suggesting that external facilitators introducing expectations without established teacher relationships may create stress rather than support.

The pattern across arms points to the importance of maintaining teacher involvement in parental

engagement. The two models that preserve this connection (T1 and T2) both produce significant child development gains without raising stress, while the model that severs it (T3) produces null developmental effects and raises stress. The facilitator model's failure is notable because it improved classroom environments (0.60 SD on the LEP Index) and produced some parental engagement gains (0.11 SD) — suggesting that both inputs were present, but their coordination was broken when delivered by separate actors. Section 6.1.1 further isolates the marginal effect of adding a helper versus a facilitator on each outcome, formalizing the between-arm comparisons.

Our pooled child development effect (0.15 SD) and training-only effect (0.21 SD) are consistent with effect sizes documented in other early childhood interventions combining teacher training with parental engagement. The absence of stress increases distinguishes our intervention from some remote parenting programs that inadvertently raised caregiver anxiety (Arteaga, de Barros, and Ganimian, 2025), indicating that school-based, teacher-led models can enhance parental engagement without overburdening families. Detailed cost-effectiveness analysis and comprehensive benchmarking against comparable programs appear in Section 8.

6.1.1 Comparing Implementation Models: Marginal Effects and Statistical Power

To directly assess whether implementation models differ from each other — and how precisely these differences are estimated — we report the nested interacted specification from Equation 5.3. Table 2 presents the results. The Training coefficient captures the Training Only effect (T1 versus control); the interaction terms show the marginal contribution of adding a helper or facilitator over training alone.

The results reveal a clear asymmetry. Adding an in-class helper produces no measurable change in any outcome: the addon coefficient is −0.053 SD on child development ($p = 0.38$), 0.047 SD on parental engagement ($p = 0.43$), and 0.000 SD on stress ($p = 1.00$). The helper neither helps nor hurts — capacity augmentation that preserves the teacher's role in parental engagement is neutral. By contrast, adding a facilitator who takes over parental engagement significantly reduces child development gains (−0.139 SD, $p = 0.02$) and significantly raises caregiver stress (0.117

Table 2: Marginal Effects of Staffing Support Over Training Alone

	ASQ Index	PESL Index	Stress Index
	(1)	(2)	(3)
Training (all arms vs Control)	0.211***	0.109*	−0.011
	(0.060)	(0.061)	(0.055)
Training × Helper (addon)	−0.053	0.047	0.000
	(0.060)	(0.059)	(0.059)
Training × Facilitator (addon)	−0.139**	−0.005	0.117**
	(0.058)	(0.062)	(0.054)
Observations	3,014	3,014	3,014
R^2	0.22	0.08	0.36

Notes: The table reports results from the nested interacted specification (Equation 5.3). Training equals one for all treatment schools. Training × Helper equals one for the Helper arm only; its coefficient estimates the marginal effect of adding a classroom helper over training alone. Training × Facilitator equals one for the Facilitator arm only; its coefficient estimates the marginal effect of adding a facilitator. All regressions include municipality fixed effects. Standard errors clustered at the school level in parentheses. ***, **, * denote significance at 1, 5, and 10 percent levels.

SD, $p = 0.03$). Delegating the parental engagement function to a worker disconnected from the classroom actively worsens outcomes relative to teachers working alone.

These contrasting patterns establish that the binding distinction is not between "training alone" and "training with support" but between models that preserve teacher involvement in family engagement (T1 and T2) and the model that severs it (T3). Table A5 (Appendix) reports pairwise differences and ex-post power calculations that formalize this interpretation. Training Only and Training plus Helper are statistically indistinguishable on all outcomes ($p > 0.38$), with minimum detectable effects of approximately 0.17 SD — substantially exceeding the observed 0.05 SD difference. The study was not powered to detect differences of this magnitude between arms (power = 14%). By contrast, Training Only significantly outperforms the Facilitator model on child development (difference = 0.14 SD, $p = 0.02$) and caregiver stress (difference = 0.12 SD, $p = 0.03$), with power of 67% and 58% respectively. These power constraints mean that the T1–T2 comparison is less informative in our study: we cannot distinguish between the null of equal effects and meaningful differences below 0.17 SD. The T1–T3 comparison, however, is informative: the study was adequately powered and detected a significant difference.

In sum, the evidence supports a clear interpretation: models preserving teacher involvement in parental engagement (T1 and T2) produce equivalent child development gains that are genuine —

persisting on ASQ items where reporting bias is structurally precluded. The model that delegates parental engagement to an external worker (T3) produces significantly worse outcomes on both child development and caregiver stress, consistent with the loss of informational complementarities when classroom observation and parental advice are performed by different actors.

6.2 Effects on Caregiver Attitudes Toward Disability

The primary outcomes capture behavioral changes targeted by the bulk of the parental education curriculum — parental engagement practices (PESL) and child developmental skills (ASQ). Two additional curriculum components targeted caregiver beliefs rather than behaviors: positive discipline (Session 12) and disability awareness (Sessions 19–20). We examine whether these content-specific modules shifted corresponding attitudes. Table 3 reports treatment effects on caregiver beliefs about discipline and attitudes toward disability. Panel A presents pooled effects, Panel B compares training to training with support, and Panel C disaggregates by support type.

The intervention did not change caregiver views on physical discipline (0.00 SD, $p > 0.10$). Caregivers in both treatment and control groups maintained similar beliefs about whether spanking is necessary or acceptable. By contrast, attitudes toward disability improved modestly. The Attitude to Disability Index, which captures views on integrating children with disabilities in school and community settings, increased by 0.09 SD ($p < 0.10$). Caregivers also reported marginally higher confidence in supporting children with disabilities (0.08 SD, $p > 0.10$), though this effect is not statistically significant.

Panel C reveals that the helper model likely drove these attitudinal changes. The training-only arm shows minimal effects in attitudes toward disability (0.07 SD, $p > 0.10$), while the helper arm generates the strongest improvements in both attitudes toward disability (0.14 SD, $p < 0.05$) and confidence in supporting children with disabilities (0.13 SD, $p < 0.10$). The facilitator arm produces small, statistically insignificant effects (0.05 SD and 0.10 SD respectively). This pattern mirrors the parental engagement results: in-class support enabled teachers to deliver more effective parental education, while capacity-constrained teachers (training-only) and external facilitators both

Table 3: Effects on Caregiver Beliefs and Attitudes

	Attitude to Discipline	Attitude to Disability Index	Confidence Disability Index
	(1)	(2)	(3)
Panel A: Pooled Treatment Arms			
[0.5em] Any Treatment	0.00	0.09*	0.08
	(0.03)	(0.05)	(0.06)
Observations	3,004	3,004	3,003
R^2	0.06	0.10	0.09
Panel B: Combining Training and Help			
[0.5em] Training Only	-0.00	0.07	0.01
	(0.03)	(0.06)	(0.07)
Training + Any Help	0.00	0.10*	0.11*
	(0.03)	(0.05)	(0.06)
Observations	3,004	3,004	3,003
R^2	0.06	0.10	0.09
Panel C: Individual Treatment Arms			
[0.5em] Training Only	-0.00	0.07	0.01
	(0.03)	(0.06)	(0.07)
Training + Helper	-0.03	0.14**	0.13*
	(0.03)	(0.06)	(0.06)
Training + Facilitator	0.03	0.05	0.10
	(0.03)	(0.06)	(0.07)
Observations	3,004	3,004	3,003
R^2	0.06	0.11	0.09
Helper = Facilitator (p-value)	0.08	0.14	0.62

Notes: The table presents results on caregiver beliefs estimated using regression equation (5.1). The outcome variable *Attitude to Discipline* is a dummy equal to 1 if the parent considers it necessary or normal to spank a child for discipline. The outcome variable *Attitude to Disability Index* is a standardized index measuring caregiver attitudes toward disability, constructed as the additive sum of responses to whether the caregiver believes that children with disabilities should be separated from their parents, whether schools should accommodate children with disabilities, whether disabilities can be helped, and whether children with disabilities deserve access to educational opportunities. The outcome variable *Confidence Disability Index* measures the caregiver's confidence in supporting children with disabilities, based on whether they believe they have the necessary knowledge and skills and whether they feel able to provide such support. All regressions include municipality fixed effects. *Any Treatment* equals one if a school received *Training Only*, *Training + Helper*, or *Training + Facilitator*, and zero otherwise. Standard errors are clustered at the school level and reported in parentheses. The data comes from the Endline Survey, which includes all 3,004 student respondents. ***, **, and * denote statistical significance at the 1, 5, and 10 percent levels, respectively.

produced weaker attitudinal shifts.

The null effect on discipline beliefs alongside modest improvements in disability attitudes suggests the intervention was more effective at introducing new information (disability inclusion) than overturning established norms (physical discipline). This is consistent with evidence that intensive two-generation programs can reduce harsh disciplinary beliefs (Love et al., 2005), whereas parenting programs in low- and middle-income countries have shown mixed success in shifting deeply entrenched practices (Knerr, Gardner, and Cluver, 2013). Our disability attitude improvements

(0.09–0.14 SD) align with evidence that parenting interventions shift beliefs most readily in domains where new information is being introduced through trusted community figures (Yousafzai et al., 2014). The helper model's advantage in producing these shifts reinforces that workload relief enables teachers to engage parents more effectively on complex topics requiring sustained discussion and trust-building.

6.3 Heterogeneity in Program Effects

Beyond average effects, characterizing who the program helps most, and who it may burden, carries direct welfare and policy weight. Returns to early childhood investment depend on where that investment lands. Because early inputs and a child's existing endowments are complementary in producing later human capital, the same program may yield different developmental returns for children who begin at different points (Aizer and Cunha, 2012; Johnson and Jackson, 2019). Additionally, these early gains carry through to schooling and adult earnings (Gertler et al., 2014; Heckman et al., 2010), and because which children a program reaches shapes its measured return (Kline and Walters, 2016), identifying the baseline characteristics that predict larger or smaller effects tells us where the program's return on investment is greatest and where it may come at a cost.

Following our pre-analysis plan, we examine whether effects vary by baseline characteristics - specifically baseline parental engagement as the primary dimension of heterogeneity. We report treatment effects by baseline quartile for each outcome and each arm in Appendix Figure A3. Child development gains are largest for children in the lowest baseline ASQ quartile (0.31 SD for Training Only) and smaller in higher quartiles. Parental engagement shows the reverse: gains are largest for families in the highest baseline PESL quartile. Both patterns appear consistent: children with lower baseline scores have more room to improve, while families already engaged at baseline are the most responsive to the program. The pattern does not differ across arms; interacting each arm with baseline quartile leaves the interaction terms jointly insignificant for every outcome (F-test $p > 0.65$).

Caregiver stress differs across arms on the pre-specified engagement split. For families with

below-median baseline engagement, the Facilitator arm raised stress by 0.20 SD ($p < 0.01$), while the two arms that keep the teacher engaged with families did not. For families with above-median baseline engagement, Training Only reduced stress by 0.14 SD ($p < 0.05$). This matches the arm-level stress result in Table 1: a facilitator with no prior relationship with the family introduces new expectations without the familiarity and trust a child's own teacher brings. These subgroup estimates are imprecise, so we read them as indicative, but the direction is the one we pre-specified.

The splits above examine one baseline characteristic at a time, and the choice of which characteristic to split on is somewhat arbitrary. In addition, if the families and children who respond most to the program differ along several characteristics at once, such splits may mask the heterogeneity and may point to the wrong characteristic. We therefore estimate a conditional average treatment effect (CATE) for each child using a causal forest (Athey, Tibshirani, and Wager, 2019; Wager and Athey, 2018). The CATE is the program's expected effect for a child with a given set of baseline characteristics. We employ a causal forest to estimate it using a set of twelve baseline child and household characteristics jointly and let the data determine which of them predict larger effects.[11]

Figure 5 summarizes the estimates for each arm and outcome. Panel A plots the distribution of estimated CATEs across children; Panel B reports the share of forest splits that use each baseline characteristic, a measure of its contribution to the heterogeneity. For child development, the predicted effects are positive for essentially all children under Training Only and the Helper arm; under the Facilitator arm they are more dispersed and shifted toward zero. For caregiver stress, the Facilitator distribution is the only one that places substantial mass in the harmful direction. In every arm, the child's baseline development and age account for most of the heterogeneity with a few other baseline characteristics varying in presence and importance across the three arms and

[11] We estimate the CATE with a causal forest (`CausalForestDML`, Athey, Tibshirani, and Wager, 2019) separately for each arm relative to control. Estimation proceeds in three steps. First, following the double machine-learning approach (Chernozhukov et al., 2018a), we partial municipality fixed effects out of both the outcome and the treatment indicator using cross-fitted LASSO, so that the forest operates on residualized outcomes and treatment, matching the fixed-effects structure of our main specification. Second, the forest is grown on these residuals. We populate 2,000 trees to fit on a random subsample of the data, and at each node the splitting variable is chosen from a randomly drawn subset of the twelve characteristics, with the split placed to maximize the difference in estimated treatment effects across the resulting groups; this randomization decorrelates the trees, and averaging over them reduces the variance of the estimated CATE. We require a minimum of 30 observations per leaf. Third, each tree uses *honest estimation*—the observations used to choose splits are held separate from those used to estimate the effect within each leaf—which prevents the same data from both selecting and estimating the heterogeneity and yields valid confidence intervals (Athey, Tibshirani, and Wager, 2019).

outcomes. Finally, we conduct a battery of tests that confirm the CATE predictions align with real signal about the underlying heterogeneity and are not a product of potential *overfitting* arising from the non-parametric nature of the forests.[12]

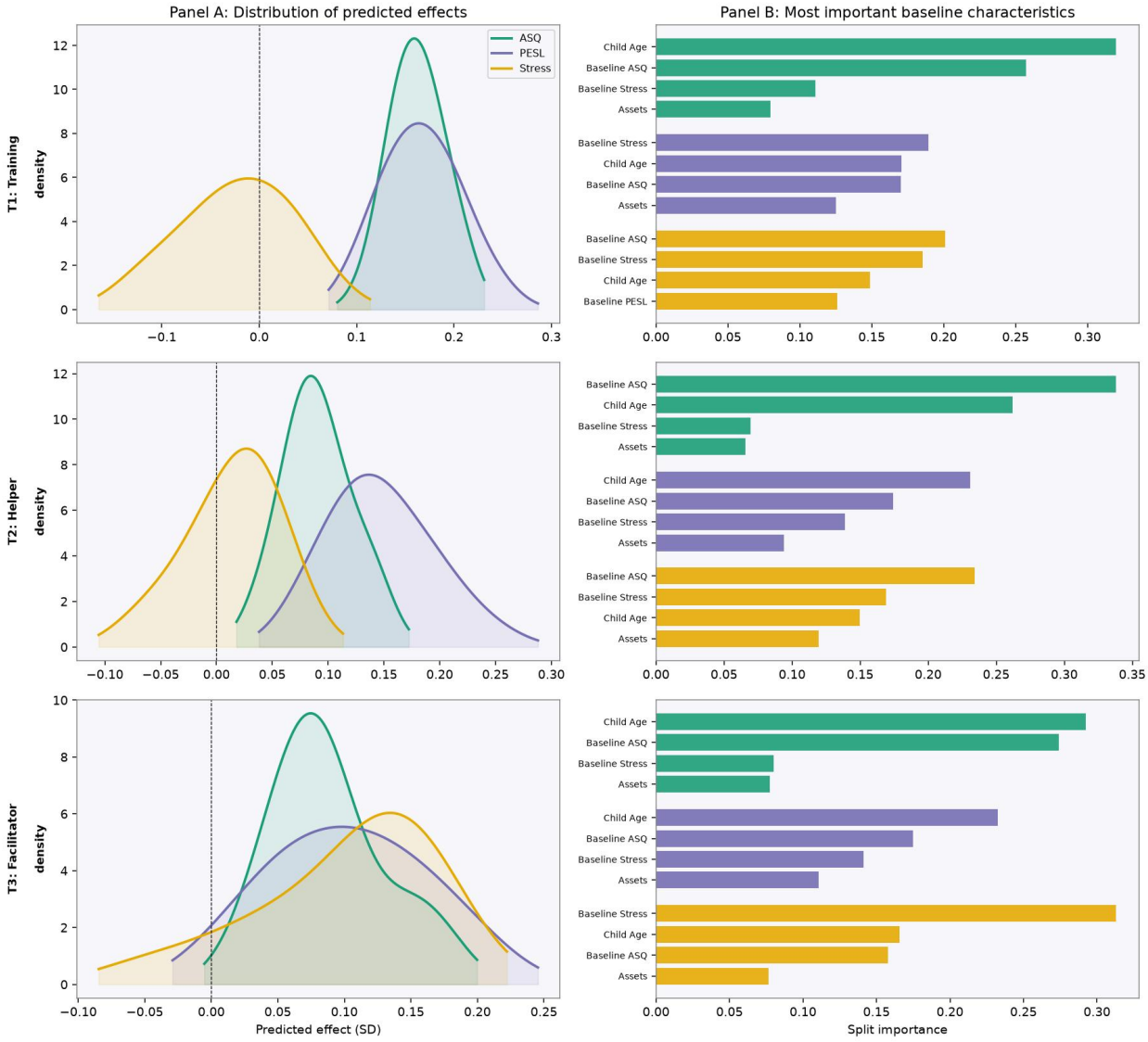
Figure A4 reports, for each arm, the importance of each baseline characteristic and the signed direction in which it predicts larger effects. For child development, baseline ASQ and child age are the two leading characteristics in every arm, together accounting for over half of the split importance, and both enter negatively: children with lower baseline ASQ scores and younger children have larger predicted effects. This is the compensatory pattern observed in Figure A3, recovered by the forests' predictions without being imposed. However, the importance of these characteristics varies by arm—child age is the leading characteristic under Training Only (32 percent of splits), while baseline ASQ leads under the Helper arm (34 percent)—but household characteristics such as assets and caregiver education contribute little in any arm, indicating that who gains in child development is governed by the child's own baseline standing rather than by family circumstances.

Parental engagement follows a different ordering. Child age is the leading characteristic and enters positively, so engagement gains concentrate among families of older children, while baseline parental engagement (baseline PESL) itself contributes little; baseline ASQ remains a secondary but consistent predictor. This is the reinforcing pattern, in which families already positioned to act respond most.

Caregiver stress is where the arms diverge. Under Training Only and the Helper arm, baseline ASQ is the leading characteristic and enters negatively, so these arms relieve stress most for families of children with lower baseline ASQ. Under the Facilitator arm the picture changes: baseline caregiver stress becomes the single most important characteristic (31 percent of splits) and enters positively, so families already reporting higher baseline stress see the largest further increases. This is an indication that the Facilitator arm potentially depressed the gains for a specific group, which

[12]We subject these predictions to three checks, reported in Appendix Figures A5–A7. First, a calibration test confirms that the forest's ranking carries real signal: sorting children by predicted effect and re-estimating actual effects recovers the predicted ordering, and the coefficient on the average predicted effect is statistically indistinguishable from one for child development and engagement. Second, because sorting on an in-sample prediction and re-estimating within groups can mechanically manufacture differences, we re-estimate the group effects using cross-fitting, forming the groups from a prediction trained on a separate part of the sample; the pattern is unchanged. Third, partial-dependence plots show that each arm's leading characteristic moves the predicted effect smoothly and in a consistent direction across children, indicating that the heterogeneity reflects stable relationships rather than a few influential cases. We also confirm that the results are not sensitive to the minimum leaf size in the estimated forests.

Figure 5: Estimated Treatment-Effect Heterogeneity by Arm



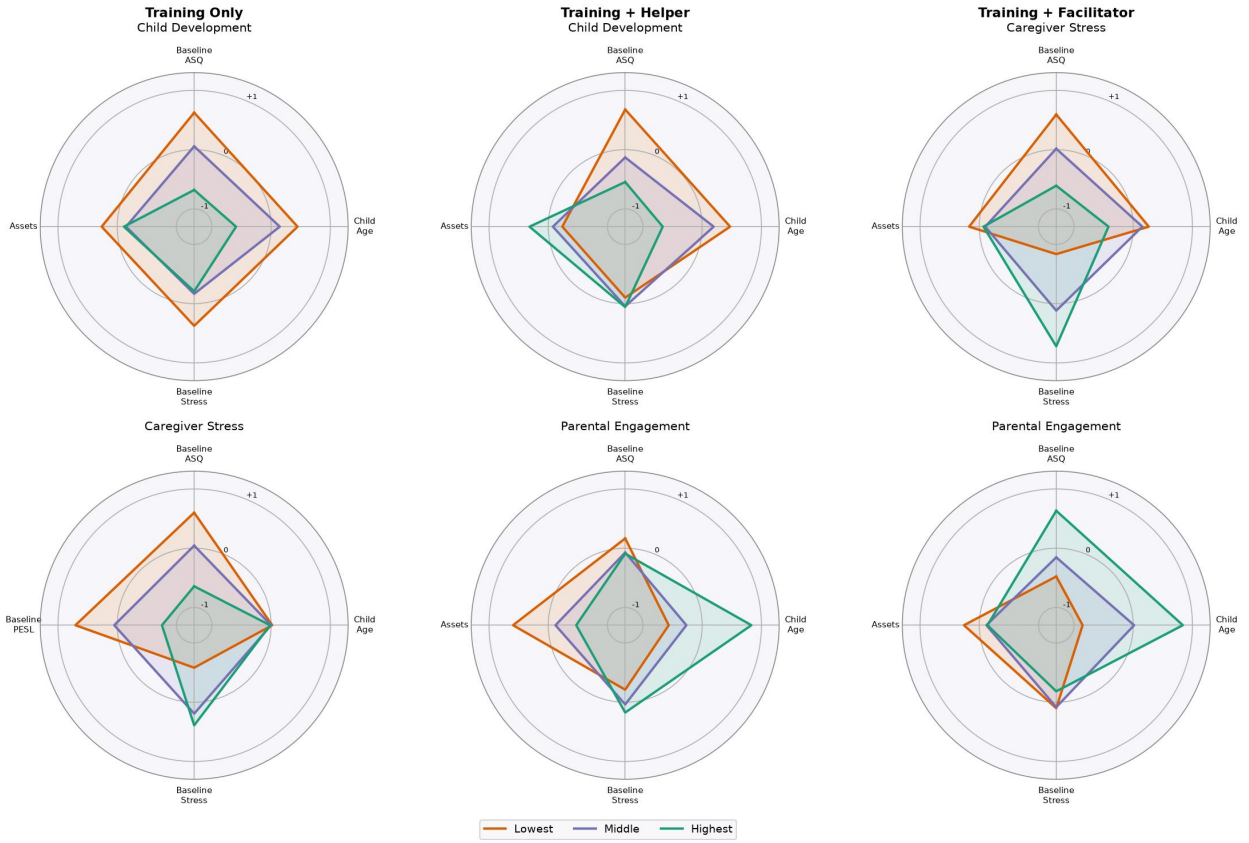
Notes: Each row corresponds to one treatment arm relative to control (Training Only, Training + Helper, Training + Facilitator). Panel A plots the distribution of estimated conditional average treatment effects (CATEs) for the three outcomes—child development (ASQ), parental engagement (PESL), and caregiver stress—as smoothed densities; the vertical line marks a zero effect. For child development and parental engagement a positive effect is beneficial, whereas for caregiver stress a positive effect denotes an increase in stress. Panel B reports variable importance, defined as the share of forest splits that use each baseline characteristic, for each outcome's four most important characteristics. Estimation follows the causal-forest procedure described in the text.

we characterize below.

The forest's ranking of children by predicted effect lets us characterize who the program helps most and least in terms of their baseline characteristics. Figure 6 shows the baseline profile of the lowest, middle, and highest predicted-effect groups for the most informative arm–outcome combinations; the remaining combinations, which follow the same patterns, appear in Appendix

Figure A8.

Figure 6: Baseline Profile of Predicted Groups for Selected Arm–Outcome Combinations



Notes: Each panel corresponds to one treatment arm (columns) and one outcome, showing the average baseline profile of children in the lowest, middle, and highest tercile of the forest-predicted effect for that arm and outcome. Panels display the most informative arm–outcome combinations; the remaining combinations appear in Appendix Figure A8. Each axis is one baseline characteristic, with the three characteristics common across panels (baseline ASQ, child age, and baseline stress) held in fixed positions and the fourth position occupied by that panel's next most important characteristic. Values are standardized across the estimation sample, so a value of +1 indicates the group averages one standard deviation above the sample mean and a larger radius indicates a higher value. For caregiver stress, a higher predicted effect denotes a larger increase in stress.

For child development, the children who gain most have lower baseline ASQ and are younger, and this profile is common across all three arms. What separates the arms is the household. Under Training Only, the largest gains accrue to less advantaged households, with lower assets and no more education than average. Under the Helper arm, the largest gains shift toward better-resourced households: the children who gain most have more educated caregivers (0.33 SD above the sample mean) and more household assets. Relieving the teacher of routine classroom work appears to let the program reach children whose home environments can act on what the teacher conveys.

Caregiver stress identifies who bears the cost of the Facilitator arm. The families whose stress

rises most are those already reporting high baseline stress (0.72 SD above the mean), with lower baseline ASQ, younger children, and somewhat fewer resources. The families whose stress falls are the mirror image: low baseline stress and higher baseline ASQ. What distinguishes the Facilitator arm is not the direction of this pattern but which characteristic drives it: under Training Only the group whose stress changes most is defined primarily by baseline ASQ, whereas under the Facilitator arm baseline stress itself becomes the dominant characteristic, so the largest increases concentrate among families already reporting the most stress. The cost of moving parental engagement to an external worker therefore falls on the households least able to absorb an additional, unfamiliar demand—the households the program is most intended to reach.

This heterogeneity analysis allows us to identify where the intervention has most bite. Developmental gains are largest for children with lower baseline ASQ, and caregiver stress increases only under the Facilitator arm, and only among families already reporting higher baseline stress. Identifying which families and children benefit, which are unaffected, and which are made worse off can help target programs of this kind and match the delivery model to the populations they are intended to serve.

6.4 Robustness

6.4.1 Experimenter Demand Effects and ASQ Decomposition

A concern with caregiver-reported developmental outcomes is that treatment may improve parents' awareness of milestones — what de Quidt, Vesterlund, and Wilson (2019) term "motivated reporting" — rather than children's actual skills. We address this directly.

The empirical literature on experimenter demand effects provides a quantitative bound. de Quidt, Haushofer, and Roth (2018) estimate demand effects averaging 0.13 standard deviations across eleven experimental tasks using explicit "demand treatments." Critically, demand effects are strongest for attitudinal or hypothetical questions and substantially weaker for concrete behavioral reports (de Quidt, Vesterlund, and Wilson, 2019; Winichakul et al., 2024). The ASQ falls squarely in the latter category: items describe specific physical actions ("Does your child kick a ball by swinging

his leg forward?") rather than opinions or attitudes. The administration protocol instructs caregivers to have the child attempt each activity in real time ("Try each activity with your baby before marking a response"), grounding responses in observed performance rather than recall. Items are calibrated to skills that are *emerging* at each age group — not already mastered — reducing the scope for over-reporting, and the three-point response scale (Yes/Sometimes/Not Yet) provides an intermediate option that absorbs uncertainty without forcing binary inflation. Moreover, demand effects require participants to infer and conform to researcher expectations — a cognitive step more plausible among Western, survey-experienced populations. Yet even in such settings, Mummolo and Peterson (2019) find that participants exhibit limited ability to adjust responses to align with researcher expectations, even when given financial incentives to do so. In our rural Nepali setting — where caregivers have limited prior exposure to surveys, data were collected by local enumerators rather than program implementers, and social desirability has been shown not to predict over-reporting of health behaviors (Katz et al., 2023) — the scope for demand-driven distortion is further attenuated.

Nevertheless, to provide a direct test, we decompose the ASQ following a simple tenet, that we retain only items whose performance the child can be asked to demonstrate and that the enumerator can directly observe during the assessment rather than items reported from memory by the caregiver. Items that pass this screen form the basis of progressively stricter indices, while items that rely on caregiver recall of behavior outside the visit are set aside.[13] We construct three progressively stricter indices, the *ASQ Observed* index (70–77% of items) retains only items where behavior can be elicited during the assessment visit, excluding home routines and habitual speech requiring caregiver recall; the *ASQ Strict Test* index (37–67%) retains only items where the questionnaire mandates a formal performance protocol with explicit test instructions; and the *ASQ Performance* index (20–40%) retains only items requiring physical motor output — drawing, jumping, catching — where no verbal report or caregiver interpretation is involved. Each is a strict subset of the one above. Appendix F provides the complete item-by-item classification for each age group.

[13]This is why even our loosest index, ASQ Observed, is a strict subset of the Original ASQ rather than equal to it. The Original index includes items such as daily routines (using a spoon, dressing), habitual speech, and playground or social behavior, which the enumerator cannot elicit on the spot and which we therefore drop. The three indices are nested by construction, with ASQ Performance $\subset$ ASQ Strict Test $\subset$ ASQ Observed $\subset$ Original, so an item excluded from ASQ Observed is necessarily excluded from the stricter two.

Table 4 reports the results. Training Only effects remain large and statistically significant across progressively stricter indices: 0.221 SD on the observed index ($p < 0.01$), 0.185 SD on strict-test items ($p < 0.05$), and 0.142 SD on performance-only items ($p < 0.10$). These magnitudes are economically meaningful and statistically distinguishable from zero even when restricting to items where the child must physically demonstrate a skill — leaving less room for reporting bias. The observed attenuation across indices are more consistent with loss of precision from using fewer items (standard errors grow from 0.060 to 0.077) rather than removal of demand-driven signal. The Facilitator arm remains null across all indices (0.028–0.031 SD), confirming the absence of detectable developmental effects regardless of measurement approach. The differential pattern across arms also indicates against uniform demand: if all treated parents simply became better reporters, effects should remain proportionally similar across arms and indices. Instead, Training Only produces significant effects even on the strictest items, while the Facilitator produces null effects — a pattern consistent with genuine developmental gains in T1 that do not occur in T3.

Table 4: ASQ Decomposition: Treatment Effects on Progressively Stricter Indices

	ASQ Index			
	Original	Observed	Strict Test	Performance
	(1)	(2)	(3)	(4)
Training Only	0.211***	0.221***	0.185**	0.142*
	(0.060)	(0.069)	(0.086)	(0.077)
Training + Helper	0.158**	0.164**	0.137	0.079
	(0.066)	(0.074)	(0.090)	(0.082)
Training + Facilitator	0.083	0.079	0.028	0.031
	(0.064)	(0.069)	(0.082)	(0.075)
Observations	3,004	3,004	3,004	3,004
R^2	0.23	0.18	0.06	0.04

Notes: Each column reports treatment effects from a separate regression of the specified ASQ index on treatment arm indicators with municipality fixed effects. Column (1) uses the original ASQ index (all 30 items per age group). Column (2) uses the *Observed* index (70–77% of items), retaining only items where behavior is elicited during the assessment visit. Column (3) uses the *Strict Test* index (37–67%), retaining only items with explicit test protocol instructions. Column (4) uses the *Performance* index (20–40%), retaining only items requiring physical motor output. All indices are standardized relative to the control group. Standard errors clustered at the school level in parentheses. See Appendix F for complete item classification. ***, **, * denote significance at 1, 5, and 10 percent levels.

6.4.2 Attrition and Compliance

As noted in Section 4.7.2, 127 of 3,131 children (4%) did not participate in the endline survey. We examine whether attrition threatens internal validity by differing systematically between treatment and control groups or by correlating with baseline characteristics.

Table A6 Panel A reports attrition rates by treatment arm. Attrition in training-only schools (2.9%) and helper schools (2.2%) is slightly higher than in control schools, but differences are not statistically significant ($p = 0.20$ and $p = 0.25$, respectively). Facilitator schools show minimal differential attrition (0.6%, $p = 0.52$). Panel B examines baseline predictors of attrition. Children from lower-asset households were more likely to attrit (-0.003, $p = 0.01$), as were children with less-educated caregivers (-0.004, $p = 0.07$). Female children were less likely to attrit (-0.010, $p = 0.09$). Baseline outcome measures (ASQ, PESL, Stress) do not significantly predict attrition.

Table A7 presents Lee bounds addressing potential selection bias. For child development (ASQ), the 95% confidence interval ranges from 0.066 to 0.229 SD, with the lower bound remaining positive. This indicates that even under worst-case assumptions about selective attrition, treatment effects on child development remain positive. For parental engagement (PESL), bounds range from -0.116 to 0.121 SD, crossing zero. However, the low overall attrition rate (4%) and absence of baseline outcome differences between attritors and non-attritors suggest limited scope for substantial bias. For caregiver stress, bounds range from -0.119 to 0.051 SD.

We also examine non-compliance in program implementation. Table A8 shows that 20 of 150 treatment schools (13%) did not conduct parental education sessions: 6 training-only, 9 helper, and 5 facilitator schools. Children in non-compliant schools lived in larger households (0.313 additional adults, $p = 0.03$), were less likely to have female caregivers (-0.067, $p = 0.05$), and had less-educated caregivers (-0.336 years, $p = 0.01$). Importantly, baseline outcome measures (ASQ, PESL, Stress) did not differ significantly between compliant and non-compliant schools ($p > 0.67$ for all outcomes). Our intent-to-treat estimates include all treatment schools regardless of implementation, providing conservative estimates of program effectiveness.

7 Potential Mechanisms

We examine potential mechanisms through which the intervention generated effects: how the intervention reshaped classroom environments and teacher-student interactions, compensatory and reinforcement pathways, differential parental attendance across treatment arms, and shifts in how home and school environments interact.

7.1 Effect on Classroom Environments

Table 5 reports treatment effects on the Learning Environment and Practices (LEP) Index. Panel A presents pooled treatment effects, Panel B compares teacher training alone to training combined with any support staff, and Panel C disaggregates by type of support (in-classroom helper versus outside classroom facilitator).

Pooled treatment schools improved classroom environments by 0.91 SD relative to control schools ($p < 0.01$). This substantial effect indicates that teacher training combined with structured parental education significantly strengthened classroom organization, resources, and instructional practices.

Panel C reveals important heterogeneity across implementation models. The helper model produced the largest gains (1.14 SD, $p < 0.01$), significantly exceeding both the training-only model (0.97 SD, $p < 0.01$) and the facilitator model (0.60 SD, $p < 0.01$). The difference between helper and facilitator models is statistically significant ($p = 0.03$), suggesting that in-class support is particularly effective at improving classroom quality. Helpers assist with classroom management, freeing teachers to focus on instructional organization and child-centered activities while preserving the teacher's engagement with families. By contrast, facilitators — who assume the teacher's parental engagement role — produced smaller classroom improvements, suggesting that the teacher's dual function across classroom and home domains may reinforce both.

Panel B shows that training combined with any form of support (0.87 SD) produces similar effects to training alone (0.97 SD), though both models generate substantial improvements. The

Table 5: Effect on Learning Environment and Practices Index

	Learning Environment & Practices Index
Panel A: Pooled Treatment	
Any Treatment	0.91***
	(0.18)
Observations	201
R^2	0.27
Panel B: Training + Helper	
Training Only	0.97***
	(0.23)
Training + Any Help	0.87***
	(0.20)
Observations	201
R^2	0.27
Panel C: Individual Arms	
Training Only	0.97***
	(0.23)
Training + Helper	1.14***
	(0.24)
Training + Facilitator	0.60***
	(0.23)
Observations	201
R^2	0.29
Helper = Facilitator (p-value)	0.03

Notes: The sample comprises all schools that participated in the intervention. The table reports treatment effects on the *Learning Environment and Practices Index* (see Appendix Section G.2). This index aggregates 20 checklist items capturing lesson planning, pedagogical practices, physical infrastructure, and organization of structured learning areas. Each item was coded 1 if present and 0 if absent; totals are standardized relative to the control group. All regressions include municipality fixed effects. *Any Treatment* equals one if a school received *Training Only*, *Training + Helper*, or *Training + Facilitator*, and zero otherwise. Standard errors are clustered at the school level and reported in parentheses. ***, **, * denote statistical significance at the 1, 5, and 10 percent levels, respectively.

helper model's larger effect in Panel C reflects that in-class support specifically enhances classroom quality, while external facilitators do not.

7.2 Improvements in Teacher-Student Interactions

Table 6 reports treatment effects on teacher-student interactions. Panel A shows that treatment classrooms experienced a 0.49 SD increase in positive interactions ($p < 0.01$) and a 0.57 SD decline in negative interactions ($p < 0.01$) relative to controls. Teachers in treatment schools were more likely to read to groups, facilitate conversations, and display warmth, while relying less on harsh

discipline and showing reduced disengagement.

Table 6: Effects on Teacher–Student Interaction Quality

	Positive Interaction Index	Negative Interaction Index
	(1)	(2)
Panel A: Pooled Arms		
[0.5em] Any Treatment	0.49***	-0.57***
	(0.17)	(0.15)
Observations	3,004	3,004
R^2	0.15	0.20
Panel B: Training + Helper		
[0.5em] Training Only	0.55***	-0.59***
	(0.21)	(0.17)
Training + Any Help	0.46**	-0.56***
	(0.18)	(0.16)
Observations	3,004	3,004
R^2	0.15	0.20
Panel C: Individual Arms		
[0.5em] Training Only	0.55***	-0.59***
	(0.21)	(0.17)
Training + Helper	0.44**	-0.53***
	(0.19)	(0.18)
Training + Facilitator	0.48**	-0.58***
	(0.20)	(0.17)
Observations	3,004	3,004
R^2	0.15	0.20
Helper = Facilitator (p-value)	0.84	0.71

Notes: The table reports results on teacher–student interactions estimated from regression equation (5.1). The outcome variables are the *Positive Interaction Index* and the *Negative Interaction Index* (see Appendix Section G.2), each measured as the frequency of interactions observed across 20 structured rounds. Positive interactions include reading to a group, teacher or student singing, circle-time activities, back-and-forth discussions, engaging in pretend play, telling children what to do, calmly negotiating peer conflict, and smiling. Negative interactions capture negative discipline, yelling, not engaging with children, talking with another adult, or being out of the classroom. A higher score reflects greater frequency of the specified type of interaction. All regressions include municipality fixed effects. *Any Treatment* equals one if a school received *Training Only*, *Training + Helper*, or *Training + Facilitator*, and zero otherwise. Standard errors are clustered at the school level and reported in parentheses. The data comes from the Endline Survey, which includes all 3,004 student respondents. ***, **, and * denote statistical significance at the 1, 5, and 10 percent levels, respectively.

Panels B and C reveal that improvements in teacher-student interactions were consistent across all three treatment arms. The uniformity of effects is striking: coefficients range narrowly from 0.44 to 0.53 SD for positive interactions and -0.53 to -0.59 SD for negative interactions across arms, with no statistically significant differences ($p > 0.7$). This pattern contrasts sharply with the LEP Index results, where the helper model produced significantly larger gains than other arms. The consistency of effects across arms, despite their substantial differences in classroom

organization (Table 5), demonstrates that the core teacher training curriculum successfully shifted pedagogical interactions independent of implementation model. The training changed teacher behaviors toward developmentally appropriate interactions regardless of whether teachers received additional classroom support or delegated parental engagement to facilitators.

Professional training is akin to increasing teacher quality, and that has been associated with better student outcomes, even in resource constrained environments (Araujo et al., 2016). Our findings reinforce the channel that professional development focused on child-centered pedagogy can meaningfully improve teacher-student interactions, with effects concentrated in the domains directly targeted by training.

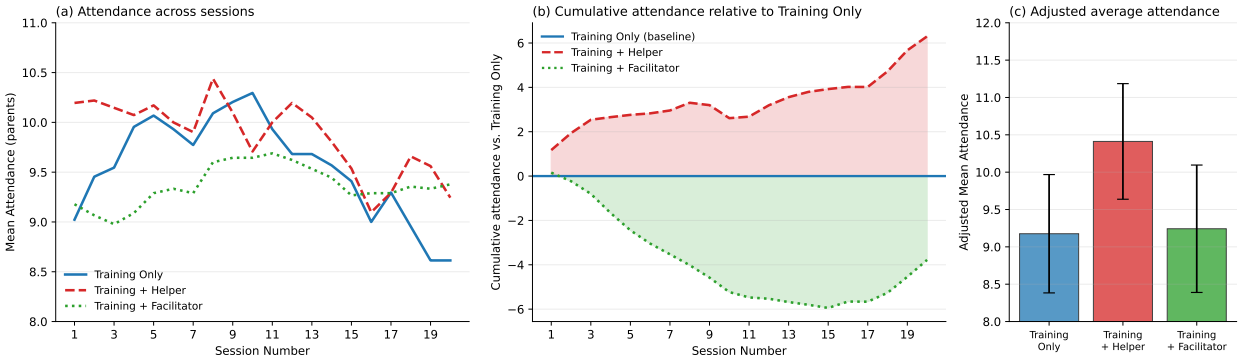
This uniformity is itself informative about mechanism. Directly-observed classroom interactions improved substantially and uniformly across all three treatment arms, yet child development effects diverge sharply: T1 and T2 produce significant gains while T3 does not. This pattern suggests that improved classroom interactions may be necessary but not sufficient for child development — they appear to translate into developmental gains only when teachers also maintain the connection between classroom instruction and family engagement. The facilitator model achieves comparable interaction improvements but severs this connection, and the developmental gains do not materialize.

7.3 Parental Attendance, Engagement Dosage, and Implementation Fidelity

Figure 7 summarizes parental attendance at education sessions across treatment arms. Schools with classroom helpers recorded higher parental attendance, averaging 1.24 additional parents per session compared to training-only schools ($p < 0.10$; Appendix Table A1). This attendance advantage persisted throughout the program despite overall declines across all arms (panel (a); coefficient on session number: -0.03, $p < 0.05$). By contrast, schools with facilitators showed no meaningful difference from training-only schools in average attendance (0.07, $p > 0.10$). These patterns indicate that helpers enabled teachers to mobilize parents more effectively by easing classroom workload, while facilitators operating outside the classroom provided no recruitment advantage.

The higher attendance in the helper arm reflects stronger initial mobilization rather than better

Figure 7: Parental Education Session Attendance by Treatment Arm



Notes: The figure summarizes parental education session attendance by treatment arm across the twenty sessions. Panel (a) plots unadjusted mean attendance at each session. Panel (b) plots cumulative attendance per school relative to the *Training Only* arm, showing each support arm's running total attendance advantage or shortfall over the reference arm. Panel (c) reports adjusted mean attendance per arm from a regression that controls for enrollment and municipality fixed effects, with 90 percent confidence intervals. The sample includes 130 treatment schools that implemented parental education sessions and for which attendance data were collected. Out of 150 treatment schools, 12 did not implement parental education sessions (9 had not sent teachers for parental education training, and 3 had facilitators who opted out), and attendance data were not collected for 8 additional schools. *Source:* Parental education session data.

retention. Attendance consistency, measured as the within-school coefficient of variation, did not differ across arms (differences of -0.02 relative to training-only for both support arms, $p > 0.10$; Appendix Table A1). Once parents began attending, participation regularity was similar everywhere, so the arm differences in total turnout were driven by how many parents were drawn in, not by how steadily they returned.

The helper arm's higher attendance is mirrored in its parental engagement gains: it produced the largest PESL improvement of any arm (0.16 SD), consistent with the additional turnout it generated. These developmental effects, however, did not scale one-for-one with engagement. Child development gains in the helper arm (0.16 SD) were not statistically distinguishable from training-only (0.21 SD, $p = 0.38$), even though the helper arm mobilized more parents and raised engagement more. The pattern suggests that parental engagement is a real pathway through which the helper operates, but that the teacher's coordinating role across classroom and home contributes to child development beyond engagement dosage alone.

Panel (b) of Figure 7 makes the engagement channel concrete: helper schools accumulated a steadily widening lead in cumulative attendance over the twenty sessions, tracking their larger measured PESL gains. Although the between-arm differences are not precisely estimated, the attendance and engagement evidence points in a consistent direction — the helper channel plausibly

strengthened parental engagement. That this did not produce a commensurate child development advantage over training alone indicates that engagement gains and developmental gains may not move together, not that the helper had no effect on engagement.[14]

Implementation fidelity also varied across arms: 9 of 20 non-compliant schools (those not conducting parental education sessions) were in the helper arm, compared to 6 in training-only and 5 in the facilitator arm. Despite this higher non-compliance, T2 still produces significant effects on child development (0.16 SD, $p = 0.02$) and parental engagement (0.16 SD, $p < 0.01$), suggesting that our intent-to-treat estimates for the helper model are conservative — compliant schools were likely producing even larger effects.

7.4 Caregiver Stress and the Teacher-Family Connection

A third pattern consistent with the importance of organizational structure is the caregiver stress result. The two models preserving the teacher-family connection (T1 and T2) produce no increase in caregiver stress (−0.01 SD each), while the facilitator model — the only arm where caregivers interact with someone other than their child's teacher — raises stress marginally (0.11 SD, $p = 0.06$). While we cannot identify the precise channel, this pattern is consistent with the teacher-family relationship providing coherence or reassurance that an external worker, lacking daily contact with the child, does not. The result suggests that how an intervention is delivered may matter for caregiver wellbeing independently of the content delivered.

7.5 Workload Relief Reshaping Teacher Effort

Beyond the pooled contrasts between treatment and control, the relative performance of the three implementation models offers additional insight into how organizational structure shapes outcomes. The nested model (Section 6.1.1) establishes that adding a helper is neutral while adding a facilitator is harmful relative to training alone. But examining the full pattern of point estimates across outcome domains — even where individual differences are not statistically significant — reveals suggestive

[14]This mirrors evidence that interventions raising student effort and inputs need not translate into learning gains (Islam et al., 2024).

evidence about the channels through which each model operates.

While T1 and T2 produce statistically equivalent effects on all outcomes, the pattern of point estimates hints at how workload relief may reshape teacher effort allocation. The helper model produces higher estimates on outcomes reflecting family engagement — parental attendance, PESL, and disability attitudes — while the training-only model produces marginally higher estimates on direct child-facing outcomes such as ASQ and teacher-student interactions. This directional pattern, though not statistically distinguishable, is consistent with a capacity reallocation mechanism: when helpers absorb routine classroom tasks, teachers may redirect freed capacity toward parent mobilization rather than intensifying direct child interaction. The result is a more balanced portfolio of improvements across domains in T2 versus concentrated child development gains in T1, with neither model clearly dominating the other.

A separate pattern emerges when comparing T2 to T3. Although most individual comparisons lack statistical power, the helper model produces point estimates exceeding the facilitator model on nearly every outcome domain, with suggestive directional advantages that are consistent across classroom quality, child development, parental engagement, caregiver stress, and attitudinal shifts. This consistency suggests that the distinction between augmenting capacity within the existing role structure and reassigning a relational function to a separate worker has consequences extending beyond any single outcome.

7.6 Discussion

The mechanisms above operate not only across arms but also across families. The heterogeneity patterns suggest two distinct channels. For households with lower initial engagement, the program operated through a compensatory mechanism: improving classroom quality and providing new information offset deficits in the home environment, leading to substantial child development gains (0.22 SD) despite modest changes in measured parental engagement. For households with higher initial engagement, the program operated through reinforcement: strengthening practices already in place improved measured parental engagement (0.14 SD) but yielded smaller incremental child

development gains due to ceiling effects. These contrasting pathways indicate that knowledge transfer was most effective where information gaps were largest, while behavioral responses to new information varied by baseline family context.

These mechanisms, taken together, help explain the pattern of results across treatment arms. The training-only arm's strong child development effects reflect both the compensatory pathway (for low-engagement families) and the coordination between classroom instruction and family engagement enabled by teacher continuity. The helper arm's balanced outcomes across domains reflect successful knowledge transfer facilitated by workload relief, enabling teachers to recruit parents effectively and maintain both classroom quality and family engagement. The facilitator arm's null child development effects indicate that improved classroom quality and modest parental engagement gains do not translate into developmental effects when delivered by separate actors lacking coordination between the two domains.

8 Benchmarking Program Effects and Costs

We assess the cost-effectiveness of our intervention by comparing implementation costs to developmental effects. The intervention cost approximately USD 47,500 (NPR 6 million) to implement across 150 schools, equivalent to USD 317 (NPR 40,000) per school or USD 21 (NPR 2,600) per child. These costs covered the complete program: twenty parenting education sessions, teacher training (15 days), remuneration for 50 facilitators and 50 classroom helpers, coaching and mentoring, and monitoring and supervision.

Following standard cost-effectiveness calculations for education interventions (Ganimian, Muralidharan, and Walters, 2024), we estimate benefit-cost ratios (BCR) by comparing implementation costs to projected lifetime earnings gains from improved child development. Our observed effect sizes (0.15 SD pooled child development (ASQ), 0.21 SD for the training-only model, and 0.16 SD for the helper model) translate to BCRs in the range of 11–14 under conservative assumptions about the relationship between early childhood test scores and adult earnings. Under these estimates,

even modest tax recovery (approximately 8 percent of projected lifetime earnings gains) would fully offset program delivery costs.

Our effect sizes align with the broader distribution of early childhood interventions in LMICs. Meta-analytic evidence on ECD programs documents average effects on the order of 0.15 SD on cognitive outcomes and 0.12 SD on socio-emotional development (Holla et al., 2021). Global syntheses of education interventions report median effects of 0.10–0.15 SD across more than a thousand studies, emphasizing that even effects in this range can be meaningful when achieved at scale (Evans and Yuan, 2022). Our pooled effect (0.15 SD) falls squarely within this range, while our training-only arm's effect (0.21 SD) approaches magnitudes found in well-resourced interventions despite operating in resource-constrained rural settings. This positions our intervention between intensive urban programs like Boston Pre-K, which produced short-run gains of 0.45–0.62 SD (Weiland and Yoshikawa, 2013) and long-term gains in college enrollment (Gray-Lobe, Pathak, and Walters, 2023), and lower-intensity parenting and home-visiting models.

Appendix Table A9 benchmarks our intervention against comparable early childhood programs. Our cost per child (USD 21) and BCR (11–14) compare favorably to other teacher training and parenting interventions. Tamil Nadu's ICDS facilitator program achieved a BCR of 13:1 with effects of 0.29–0.46 SD, though at substantially higher cost per child (Ganimian, Muralidharan, and Walters, 2024). Rwanda's "First Steps" parenting program achieved effects of 0.30–0.38 SD at USD 2 per caregiver session, demonstrating exceptional cost-effectiveness for a pure parenting model (Justino et al., 2023). In Colombia, a teacher professional-development program produced 0.16 SD average gains (0.30 SD for poorest households), while additional funds for teaching materials and assistants alone had no detectable effect; the professional-development component cost roughly USD 35 upfront plus USD 13 annually per child (Andrew et al., 2024). Ghana's preschool quality package achieved 0.11–0.18 SD effects with BCRs of 11–103:1 (Wolf et al., 2019), while Peru's pedagogy reform generated 0.19 SD gains with BCRs of 29–88 (Gallego, Näslund-Hadley, and Alfonso, 2021). Bangladesh's preschool initiative produced 0.25–0.37 SD gains with BCRs of 3–28 (Spier et al., 2020), and Indonesia's ECE expansion achieved 0.16 SD improvements in socio-emotional

development (Brinkman et al., 2017).

Table 7: Program Costs and Unit Costs (Aug 2022–Feb 2023)

A. Cost Summary	
Remuneration	USD 2,508.79 [NPR 316,108]
Program costs	USD 41,539.68 [NPR 5,234,998]
Indirect costs (8%)	USD 3,523.88 [NPR 444,009]
Grand total	**USD 47,572.35 [NPR 5,994,116]**
B. Unit costs	
Centers covered (number)	150
Cost per center	USD 317.15 [NPR 39,961]
Per-child cost[b] (12/15/20)	26.43/21.14/15.86 [NPR 3,330/2,664/1,998]

[a] Exchange rate: 1 USD = 126 NPR (financial proposal). NPR shown in brackets.
[b] Terms of Reference caps Parenting Education group size at 20 per center. Per-child cost = (total cost ÷ 150 centers) ÷ assumed enrollment. *Notes:* Costs cover twenty-session Parenting Education, remuneration for 50 facilitators and 50 caregivers, coaching/mentoring, and monitoring/supervision.

By contrast, more intensive interventions achieve larger effects at substantially higher costs. India's kindergarten subsidy program produced 0.80 SD gains but required extensive per-child subsidies (Dean and Jayachandran, 2020). Honduras' conditional cash transfer cost nearly USD 500 per household annually yet yielded only 0.13 SD improvements (López Bóo and Creamer, 2019), resulting in substantially lower cost-effectiveness. The teaching-assistant variant of Colombia's Hogares Infantiles cost USD 300 per child with no measurable effects, highlighting that additional personnel do not guarantee improved outcomes without appropriate implementation design. The Gambia's expansion and quality enhancement trial produced null effects despite substantial investments (Blimpo et al., 2022).

Three features distinguish our cost structure from higher-cost interventions. First, we leveraged existing government school infrastructure and teachers rather than creating parallel delivery systems, substantially reducing fixed costs. Second, the helper and facilitator models tested whether modest additional personnel (USD 5–8 per child) could enhance effectiveness, finding that in-class helpers improved outcomes while external facilitators did not—a critical design insight for cost-conscious scaling. Third, the training-only model's strong performance (0.21 SD at lowest cost) demonstrates that meaningful effects are achievable through teacher capacity-building alone when teachers

maintain continuity between classroom instruction and parental engagement.

These results indicate that our intervention represents a cost-effective approach to improving early childhood outcomes in resource-constrained settings. The combination of moderate costs (USD 21 per child), meaningful effect sizes (0.15–0.21 SD), and favorable benefit-cost ratios (11–14:1) positions this model as a viable candidate for scaling within government education systems. Our effects place us in the mid-range of documented interventions: more effective than access-only or transfer programs, comparable to combined teacher training and parenting interventions, and approaching the effectiveness of some intensive programs but at a fraction of the cost. The training-only variant offers particular promise for contexts with severe budget constraints, while the helper model provides a pathway for incremental quality improvements where modest additional resources are available.

9 Conclusion

This paper provides experimental evidence on how organizational design shapes the effectiveness of early childhood interventions delivered through government systems. We evaluate three implementation models in Nepal's government schools, finding that program effectiveness depends critically on how responsibilities are structured among frontline teachers. Models preserving the teacher's involvement in both classroom instruction and parental engagement (T1 and T2) produce statistically equivalent child development gains, while the model that delegates parental engagement to an external facilitator (T3) eliminates developmental gains entirely and raises caregiver stress — despite improving classroom environments. A nested specification confirms that adding routine classroom support is neutral relative to training alone, while unbundling the parental engagement function to a separate worker is actively harmful. These results indicate that the organization of service delivery matters as much as program content for achieving developmental effects at scale (Bold et al., 2018; Duflo, Dupas, and Kremer, 2015).

The central mechanism is consistent with a shift in how home and school inputs interact. In control

schools, classroom quality and parental engagement appear to function as substitutes — families compensate for weak schools, or schools support children from less-engaged families. Heterogeneity patterns in treatment schools suggest this relationship weakened: gains were concentrated among children who began furthest behind, consistent with school improvements complementing rather than replacing home inputs when teachers coordinate both domains. Several findings reinforce this interpretation. Teacher-student interactions improve uniformly across all three arms, yet developmental gains emerge only where teachers also maintain family engagement — indicating that improved pedagogy is necessary but not sufficient. The helper model generates higher parental attendance, yet this does not produce proportionally larger child development effects, suggesting that the teacher's coordinating role contributes independently of engagement dosage. Caregiver stress rises only in the facilitator arm, consistent with the teacher-family relationship providing coherence that an external worker — who holds formal responsibility but lacks the informational basis from daily classroom observation — does not (Aghion and Tirole, 1997). Child development gains were concentrated among children with lower baseline development, indicating potential to narrow rather than widen developmental inequalities (Johnson and Jackson, 2019).

These findings extend understanding of human capital formation in resource-constrained settings. Rapid expansions in access to education and social services often fail to generate commensurate gains when implementation design is neglected (Glewwe and Kremer, 2006; Pritchett, 2013). We show that relatively short teacher training, when embedded in structures that support coordination between home and school, can shift both classroom practices and family behaviors — underscoring that frontline capacity and the organization of support within public systems are central determinants of effectiveness (Ganimian, Mbiti, and Mishra, 2025; Mbiti et al., 2019). By documenting that complementarities exist not only between inputs but within a single teacher's task portfolio, this study points to a mechanism potentially relevant across sectors where frontline workers perform bundles of informationally related tasks (Heckman and Karapakula, 2019; Holmström and Milgrom, 1991; Pianta, La Paro, and Hamre, 2008).

For policymakers designing staffing models in education and other frontline services, our results

offer a clear heuristic. Support that reduces a teacher's routine burden without reassigning any knowledge-intensive function — like classroom helpers, teaching assistants, or administrative aides — preserves the informational complementarities that make bundled roles productive. Our evidence indicates this is neutral at worst. Support that reassigns a complete relational function to a specialist — like outsourcing parental engagement to facilitators disconnected from classrooms — risks destroying the coordination between school and home that makes integrated programs effective. The T2-versus-T3 comparison is the cleanest test of this: same number of additional workers, similar cost, radically different organizational placement, radically different outcomes. The organization of labor at the frontline — which tasks go together, which can be separated, and at what coordination cost — is itself a first-order policy lever (Becker and Murphy, 1992; Garicano, 2000). Such approaches are particularly relevant for settings where access has expanded faster than quality (Andrew et al., 2024), and where fiscal constraints limit the feasibility of intensive external interventions. Strengthening state capacity in early childhood requires not only increasing personnel but also embedding them effectively within structures that preserve coordination between teachers and families (Fryer Jr., 2018).

Several limitations warrant attention. Our follow-up window of eight months captures immediate program effects but cannot establish long-term persistence. We note three reasons for cautious optimism: early childhood interventions of comparable magnitude have produced lasting effects on adult outcomes even when initial test-score gains appeared modest (Campbell et al., 2014; Garcıa, Heckman, and Ronda, 2023; Gertler et al., 2014; Heckman and Karapakula, 2019), and universal preschool programs have shown similar persistence, with lasting gains in college enrollment operating through behavioral rather than test-score channels (Gray-Lobe, Pathak, and Walters, 2023); the dimensions we improve — parenting practices, classroom quality, teacher skills — represent durable environmental changes rather than one-time stimuli, with evidence of sustained behavioral gains from similar programs (Justino et al., 2023; Kraft, Blazar, and Hogan, 2018); and the developmental window (ages 3–5) is one where environmental inputs have disproportionate effect due to neural plasticity, with early gains scaffolding later learning through dynamic complementarity (Johnson

and Jackson, 2019). Treatment effects on child development persist even on ASQ items requiring physical child demonstration where reporting bias is structurally precluded, strengthening confidence that measured gains reflect genuine developmental change rather than motivated reporting. The program was implemented in one district; replication across institutional contexts would strengthen external validity. Longer-term follow-up through primary school and beyond will be critical to assess whether the short-run gains observed here translate into lasting improvements in schooling trajectories and labor market outcomes (Elango et al., 2015; Heckman, 2006).

Our results underscore that effectiveness depends not only on what programs deliver but on how delivery is organized — particularly whether structures maintain or sever the connections between classroom instruction and family engagement (List, Suskind, and Supplee, 2021). By embedding reforms within government infrastructure, targeting both teachers and parents, and addressing workload constraints without fragmenting relational roles, interventions can generate meaningful developmental gains at modest cost while remaining scalable. The effect sizes observed here align with the global distribution of education program effects (Evans and Yuan, 2022) but were achieved at low cost through existing public systems serving disadvantaged populations — suggesting that even moderate early gains, when delivered through organizationally coherent models within public systems, can yield substantial long-term benefits.

References

Acemoglu, Daron and David Autor. 2011. "Skills, Tasks and Technologies: Implications for Employment and Earnings." *Handbook of Labor Economics* 4:1043–1171.

Aghion, Philippe and Jean Tirole. 1997. "Formal and Real Authority in Organizations." *Journal of Political Economy* 105 (1):1–29.

Aizer, Anna and Flávio Cunha. 2012. "The Production of Human Capital: Endowments, Investments and Fertility." NBER Working Papers 18429, National Bureau of Economic Research, Inc. URL `https://ideas.repec.org/p/nbr/nberwo/18429.html`.

Alan, Sule. 2025. "Shaping Society's Character: The Role of Schools in Developing Social and Emotional Skills." In *Handbook of Culture*, edited by Benjamin Enke, Paola Giuliano, Nathan Nunn, and Leonard Wantchekon. Elsevier. Forthcoming.

Andrew, Alison, Orazio Attanasio, Britta Augsburg, Monimalika Day, Sally Grantham-McGregor, Costas Meghir, Fardina Mehrin, Smriti Pahwa, and Marta Rubio-Codina. 2020. "Effects of a scalable home-visiting intervention on child development in slums of urban India: evidence from a randomised controlled trial." *Journal of child psychology and psychiatry* 61 (6):644–652.

Andrew, Alison, Orazio P Attanasio, Raquel Bernal, Lina Cardona Sosa, Sonya Krutikova, and Marta Rubio-Codina. 2024. "Preschool quality and child development." *Journal of Political Economy* 132 (7):2304–2345.

Araujo, M. Caridad, Pedro Carneiro, Yyannu Cruz-Aguayo, and Norbert Schady. 2016. "Teacher Quality and Learning Outcomes in Kindergarten." *The Quarterly Journal of Economics* 131 (3):1415–1453.

Araujo, M Caridad, Marta Dormal, and Norbert Schady. 2019. "Childcare quality and child development." *Journal of Human Resources* 54 (3):656–682.

Arteaga, Irma, Andreas de Barros, and Alejandro J Ganimian. 2025. "The challenges of scaling up effective child-rearing practices using technology in developing settings: Experimental evidence from India." *Journal of Research on Educational Effectiveness* :1–19.

Athey, Susan, Julie Tibshirani, and Stefan Wager. 2019. "Generalized Random Forests." *The Annals of Statistics* 47 (2):1148–1178.

Attanasio, Orazio, Sarah Cattan, Emla Fitzsimons, Costas Meghir, and Marta Rubio-Codina. 2020. "Estimating the Production Function for Human Capital: Results from a Randomized Controlled Trial in Colombia." *American Economic Review* 110 (1):48–85. URL `https://www.aeaweb.org/articles?id=10.1257/aer.20150183`.

Banerjee, Abhijit, Esther Duflo, Amy Finkelstein, Lawrence F Katz, Benjamin A Olken, and Anja Sautmann. 2020. "In praise of moderation: Suggestions for the scope and use of pre-analysis plans for rcts in economics." Tech. rep., National Bureau of Economic Research.

Bayley, Nancy. 2006. *Bayley Scales of Infant and Toddler Development: Third Edition*. San Antonio, TX: Harcourt Assessment.

Becker, Gary S. and Kevin M. Murphy. 1992. "The Division of Labor, Coordination Costs, and Knowledge." *Quarterly Journal of Economics* 107 (4):1137–1160.

Benjamini, Y. and Y. Hochberg. 1995. "Controlling for False Discovery Rate: A Practical and Powerful Approach to Multiple Testing." *Journal of Royal Statistical Society, Series B (Methodological)* 57(1):289–300.

Black, Maureen M., Susan P. Walker, Lia C. H. Fernald, Christopher T. Andersen, Ann M. DiGirolamo, Chunling Lu, Dana C. McCoy, Günther Fink, Yusra R. Shawar, Jeremy Shiffman, Amanda E. Devercelli, Quentin T. Wodon, Emily Vargas-Baron, and Sally Grantham-McGregor. 2017. "Early Childhood Development Coming of Age: Science Through the Life Course." *The Lancet* 389 (10064):77–90.

Blimpo, Moussa P, Pedro Carneiro, Pamela Jervis, and Todd Pugatch. 2022. "Improving access and quality in early childhood development programs: Experimental evidence from the Gambia." *Economic Development and Cultural Change* 70 (4):1479–1529.

Bold, Tessa, Mwangi S. Kimenyi, Germano Mwabu, Alice Ng'ang'a, and Justin Sandefur. 2018. "Experimental evidence on scaling up education reforms in Kenya." *Journal of Public Economics* 168:1–20.

Brinkman, Sally Anne, Amer Hasan, Haeil Jung, Angela Kinnell, and Menno Pradhan. 2017. "The impact of expanding access to early childhood education services in rural Indonesia." *Journal of Labor Economics* 35 (S1):S305–S335.

Brutti, Zelda and Daniel Montolio. 2021. "Preventing criminal minds: Early education access and adult offending behavior." *Journal of Economic Behavior & Organization* 191:97–126. URL https://www.sciencedirect.com/science/article/pii/S0167268121003772.

Buck, Lindsey, Nathan Fiala, Nishith Prakash, Shwetlena Sabarwal, Deepak Saraswat, and Deepika Shrestha. 2020. "Educator Knowledge of Early Childhood Development: Evidence from Eastern Nepal." Policy Research Working Paper 9381, World Bank, Washington, DC.

Campbell, Frances, Gabriella Conti, James Heckman, Seong Moon, Rodrigo Pinto, Elizabeth Pungello, and Yi Pan. 2014. "Early Childhood Investments Substantially Boost Adult Health." *Science (New York, N.Y.)* 343:1478–85.

Chen, Si and Sharon Wolf. 2021. "Measuring the Quality of Early Childhood Education in Low- and Middle-Income Countries." *Frontiers in Psychology* 12:774740.

Chernozhukov, Victor, Denis Chetverikov, Mert Demirer, Esther Duflo, Christian Hansen, Whitney Newey, and James Robins. 2018a. "Double/Debiased Machine Learning for Treatment and Structural Parameters." *The Econometrics Journal* 21 (1):C1–C68.

Chernozhukov, Victor, Mert Demirer, Esther Duflo, and Iván Fernández-Val. 2018b. "Generic Machine Learning Inference on Heterogeneous Treatment Effects in Randomized Experiments, with an Application to Immunization in India." NBER Working Paper 24678, National Bureau of Economic Research.

de Quidt, Jonathan, Johannes Haushofer, and Christopher Roth. 2018. "Measuring and bounding experimenter demand." *American Economic Review* 108 (11):3266–3302.

de Quidt, Jonathan, Lise Vesterlund, and Alistair J. Wilson. 2019. "Experimenter Demand Effects." In *Handbook of Research Methods and Applications in Experimental Economics*, edited by Arthur Schram and Aljaž Ule, chap. 20. Edward Elgar Publishing, 384–400.

Dean, Joshua T and Seema Jayachandran. 2020. "Attending kindergarten improves cognitive development in India, but all kindergartens are not equal." Tech. rep., mimeo.

Dillon, Moira R., Harini Kannan, Joshua T. Dean, Elizabeth S. Spelke, and Esther Duflo. 2017. "Cognitive science in the field: A preschool intervention durably enhances intuitive but not formal mathematics." *Science* 357 (6346):47–55.

Duflo, Esther, Pascaline Dupas, and Michael Kremer. 2015. "School governance, teacher incentives, and pupil–teacher ratios: Experimental evidence from Kenyan primary schools." *Journal of Public Economics* 123:92–110.

Elango, Sneha, Jorge Luis García, James J Heckman, and Andrés Hojman. 2015. "Early childhood education." In *Economics of Means-Tested Transfer Programs in the United States, Volume 2*. University of Chicago Press, 235–297.

Evans, David K. and Fei Yuan. 2022. "How Big Are Effect Sizes in International Education Studies?" *Educational Evaluation and Policy Analysis* 44 (3):532–540.

Fernald, Lia C. H., Patricia Kariger, Patrice Engle, and Abbie Raikes. 2009. "Examining Early

Child Development in Low-Income Countries: A Toolkit for the Assessment of Children in the First Five Years of Life." Policy research working paper, World Bank.

Filgueiras, Alberto, Priscila Pires, Sébastien Maissonette, and J. Landeira-Fernandez. 2013. "Psychometric properties of the Brazilian-adapted version of the Ages and Stages Questionnaire." *Early Human Development* 89 (8):561–576.

Fryer Jr., Roland G. 2018. "The "Pupil" Factory: Specialization and the Production of Human Capital in Schools." *American Economic Review* 108 (3):616–56.

Gallego, Francisco A, Emma Näslund-Hadley, and Mariana Alfonso. 2021. "Changing pedagogy to improve skills in preschools: experimental evidence from Peru." *The World Bank Economic Review* 35 (1):261–286.

Ganimian, Alejandro J., Isaac M. Mbiti, and Abhilash Mishra. 2025. "Leveraging Complementarities between Teachers' Content Knowledge and Pedagogical Supports: Experimental Evidence from India." Unpublished manuscript.

Ganimian, Alejandro J., Karthik Muralidharan, and Christopher R. Walters. 2024. "Augmenting State Capacity for Child Development: Experimental Evidence from India." *Journal of Political Economy* 132 (5):1565–1602.

Garcıa, Jorge Luis, James J Heckman, and Victor Ronda. 2023. "The lasting effects of early childhood education on promoting the skills and social mobility of disadvantaged African Americans and their children." *Journal of Political Economy* 131 (6):1477–1506.

Garicano, Luis. 2000. "Hierarchies and the Organization of Knowledge in Production." *Journal of Political Economy* 108 (5):874–904.

Gertler, Paul, James Heckman, Rodrigo Pinto, Arianna Zanolini, Christel Vermeersch, Susan Walker, Susan M Chang, and Sally Grantham-McGregor. 2014. "Labor market returns to an early childhood stimulation intervention in Jamaica." *Science* 344 (6187):998–1001.

Glewwe, Paul and Michael Kremer. 2006. "Schools, Teachers, and Education Outcomes in Developing Countries." In *Handbook of the Economics of Education*, vol. 2, edited by Eric A. Hanushek and Finis Welch. Elsevier, 945–1017.

Gray-Lobe, Guthrie, Parag A. Pathak, and Christopher R. Walters. 2023. "The Long-Term Effects of Universal Preschool in Boston." *Quarterly Journal of Economics* 138 (1):363–411.

Heckman, James J. 2006. "Skill formation and the economics of investing in disadvantaged children." *Science* 312 (5782):1900–1902.

Heckman, James J and Ganesh Karapakula. 2019. "Intergenerational and Intragenerational Externalities of the Perry Preschool Project." Working Paper 25889, National Bureau of Economic Research. URL http://www.nber.org/papers/w25889.

Heckman, James J., Seong Hyeok Moon, Rodrigo Pinto, Peter A. Savelyev, and Adam Yavitz. 2010. "The Rate of Return to the HighScope Perry Preschool Program." *Journal of Public Economics* 94 (1-2):114–128.

Heo, Kyung H., Jane Squires, and Paul Yovanoff. 2008. "Cross-cultural adaptation of a pre-school screening instrument: Comparison of Korean and US populations." *Journal of Intellectual Disability Research* 52 (3):195–206.

Holla, Alaka, Magdalena Bendini, Lelys Dinarte, and Iva Trako. 2021. "Is investment in preprimary education too low." *World Bank Policy Research Working Paper* (9723).

Holmström, Bengt and Paul Milgrom. 1991. "Multitask Principal-Agent Analyses: Incentive Contracts, Asset Ownership, and Job Design." *Journal of Law, Economics, & Organization* 7:24–52.

Islam, Asad, Sungoh Kwon, Eema Masood, Nishith Prakash, Shwetlena Sabarwal, and Deepak Saraswat. 2024. "All pain and no gain: When goal setting leads to more effort but no gains in test scores." *Economics of Education Review* 103:102594.

Jakiela, Pamela, Owen Ozier, Lia CH Fernald, and Heather A Knauer. 2024. "Preprimary education and early childhood development: Evidence from government schools in rural Kenya." *Journal of Development Economics* 171:103337.

Jeong, Joshua, Emily E. Franchett, Clariana V. Ramos de Oliveira, Karima Rehmani, and Aisha K. Yousafzai. 2021. "Parenting interventions to promote early child development in the first three years of life: A global systematic review and meta-analysis." *PLOS Medicine* 18 (5):1–51. URL `https://doi.org/10.1371/journal.pmed.1003602`.

Johnson, Rucker C. and C. Kirabo Jackson. 2019. "Reducing Inequality through Dynamic Complementarity: Evidence from Head Start and Public School Spending." *American Economic Journal: Economic Policy* 11 (4):310–349.

Justino, Patricia, Marinella Leone, Pierfrancesco Rolla, Monique Abimpaye, Caroline Dusabe, Marie D Uwamahoro, and Richard Germond. 2023. "Improving parenting practices for early child development: Experimental evidence from Rwanda." *Journal of the European Economic Association* 21 (4):1510–1550.

Katz, Joanne, Tsering P. Lama, Melissa N. Bates, Rebecca Heidkamp et al. 2023. "How does social desirability bias influence survey-based estimates of the use of antenatal care in rural Nepal? A validation study." *BMJ Open* 13 (7):e071511.

Kerwin, Jason, Nada Rostom, and Olivier Sterck. 2024. "Striking the right balance: Why standard balance tests over-reject the null, and how to fix it." .

Kline, Patrick and Christopher R. Walters. 2016. "Evaluating Public Programs with Close Substitutes: The Case of Head Start." *Quarterly Journal of Economics* 131 (4):1795–1848.

Knerr, Wendy, Frances Gardner, and Lucie Cluver. 2013. "Improving positive parenting skills and reducing harsh and abusive parenting in low-and middle-income countries: A systematic review." *Prevention Science* 14 (4):352–363.

Kraft, Matthew A., David Blazar, and Dylan Hogan. 2018. "The effect of teacher coaching on instruction and achievement: A meta-analysis of the causal evidence." *Review of Educational Research* 88 (4):547–588.

Kvestad, Ingrid, Sunita Taneja, Mari Hysing, Tarun Kumar, and Tor A. Strand. 2013. "Diarrhea, stimulation and growth predict neurodevelopment in young North Indian children." *PLoS ONE* 8 (1).

Lee, David S. 2009. "Training, Wages, and Sample Selection: Estimating Sharp Bounds on Treatment Effects." *Review of Economic Studies* 76 (3):1071–1102.

List, John A, Dana Suskind, and Lauren H Supplee. 2021. "The scale-up effect in early childhood and public policy." *The Scale-Up Effect in Early Childhood and Public Policy: Why Interventions Lose Impact at Scale and What We Can Do About It Routledge* .

López Bóo, Florencia and John Creamer. 2019. "Cash, conditions and child development: Experimental evidence from a cash transfer in honduras." Tech. rep., IZA Discussion Papers.

Love, John M., Ellen Eliason Kisker, Christine M. Ross, Peter Z. Schochet, Jeanne Brooks-Gunn, Diane Paulsell, Kimberly Boller, Jill Constantine, Cheri Vogel, Allison Sidle Fuligni, and Christy Brady-Smith. 2005. "Making a Difference in the Lives of Infants and Toddlers and Their Families: The Impacts of Early Head Start." Tech. rep., U.S. Department of Health and Human Services, Administration for Children and Families, Washington, DC. URL https://www.acf.hhs.gov/opre/report/making-difference-lives-infants-and-toddlers-and-their-families-impacts-early-head.

Masino, Serena and Miguel Niño-Zarazúa. 2016. "What works to improve the quality of student learning in developing countries?" *International Journal of Educational Development* 48:53–65. URL https://www.sciencedirect.com/science/article/pii/S0738059315300146. Aid, Education Policy, and Development.

Mbiti, Issac, Karthik Muralidharan, Mauricio Romero, Youdi Schipper, Constantine Manda, and Rakesh Rajani. 2019. "Inputs, Incentives, and Complementarities in Education: Experimental Evidence from Tanzania." *The Quarterly Journal of Economics* 134 (3):1627–1673.

McCoy, Dana Charles, Hirokazu Yoshikawa, Kathleen M. Ziol-Guest, Greg J. Duncan, Holly S. Schindler, Katherine Magnuson, Rui Yang, Andrew Koepp, and Jack P. Shonkoff. 2017. "Impacts of Early Childhood Education on Medium- and Long-Term Educational Outcomes." *Educational Researcher* 46 (8):474–487.

McKenzie, David. 2012. "Beyond Baseline and Follow-up: The Case for More T in Experiments." *Journal of Development Economics* 99 (2):210–221.

Milgrom, Paul and John Roberts. 1990. "The Economics of Modern Manufacturing: Technology, Strategy, and Organization." *American Economic Review* 80 (3):511–528.

Mummolo, Jonathan and Erik Peterson. 2019. "Demand Effects in Survey Experiments: An Empirical Assessment." *American Political Science Review* 113 (2):517–529.

National Association for the Education of Young Children. 2025. "Staff-to-Child Ratios and Group Sizes within NAEYC Early Learning Program Accreditation Standards." `https://www.naeyc.org`. Accessed 2026.

Paudel, Netra Raj, Prakash KC, Radhika Ghimire, Clas-Håkan Nygård, and Subas Neupane. 2024. "Occupational Burnout and Their Determinants among Schoolteachers in Nepal: A Cross-Sectional Study." *BMC Public Health* 24:472.

Phillips, Deborah A and Jack P Shonkoff. 2000. "From neurons to neighborhoods: The science of early childhood development." .

Phuyal, Sabin, Ashish Shrestha, and Manish Bhattarai. 2024. "Prevalence of Occupational Voice Disorder among Teachers of Kathmandu District, Nepal." *Indian Journal of Otolaryngology and Head & Neck Surgery* .

Pianta, Robert C., Karen M. La Paro, and Bridget K. Hamre. 2008. *Classroom Assessment Scoring System (CLASS) Manual, Pre-K*. Baltimore, MD: Paul H. Brookes Publishing Co.

Poudyal, Niraj, Mahesh Banskota, and Dipesh Khadka. 2019. "Disability in Nepal." Report, UNICEF.

Pritchett, Lant. 2013. *The Rebirth of Education: Schooling Ain't Learning*. Washington, DC: CGD Books.

Schaffner, Julie, Paul Glewwe, and Uttam Sharma. 2025. "Why Programs Fail: Lessons for Improving Public Service Quality from a Mixed-Methods Evaluation of an Unsuccessful Teacher Training Program in Nepal." *The World Bank Economic Review* 39 (2):473–496.

Schonhaut, Luisa, Iván Armijo, Marianne Schönstedt, Jorge Alvarez, and Miguel Cordero. 2013. "Validity of the Ages and Stages Questionnaires in term and preterm infants." *Pediatrics* 131 (5):e1468–e1474.

Shonkoff, Jack P. and Philip A. Fisher. 2013. "Rethinking evidence-based practice and two-generation programs to create the future of early childhood policy." *Development and Psychopathology* 25 (4pt2):1635–1653.

Spier, Elizabeth, Kevin Kamto, Adria Molotsky, Azizur Rahman, Najmul Hossain, Zannatun Nahar, and Hosneara Khondker. 2020. "Bangladesh early years preschool program impact evaluation." *World Bank, Washington, DC* .

Squires, Jane and Diane Bricker. 2009. *Ages & Stages Questionnaires, Third Edition (ASQ-3): A Parent-Completed Child Monitoring System*. Baltimore, MD: Paul H. Brookes Publishing Co.

Squires, Jane, Diane Bricker, and LaWanda Potter. 1997. "Revision of a Parent-Completed Developmental Screening Tool: Ages and Stages Questionnaires." *Journal of Pediatric Psychology* 22 (3):313–328.

Vermeer, Harriet J., Marinus H. van IJzendoorn, Rodrigo A. Cárcamo, and Linda J. Harrison. 2016. "Child-Staff Ratios in Early Childhood Education and Care Settings and Child Outcomes: A Systematic Review and Meta-Analysis." *PLOS ONE* 11 (6):e0157608.

Wager, Stefan and Susan Athey. 2018. "Estimation and Inference of Heterogeneous Treatment Effects using Random Forests." *Journal of the American Statistical Association* 113 (523):1228–1242.

Walters, Christopher R. 2015. "Inputs in the Production of Early Childhood Human Capital: Evidence from Head Start." *American Economic Journal: Applied Economics* 7 (4):76–102.

Weaver, Jeffrey, Sandip Sukhtankar, Paul Niehaus, and Karthik Muralidharan. 2025. "Maternal Cash Transfers for Gender Equity and Child Development: Experimental Evidence from India." *Working Paper* .

Weiland, Christina and Hirokazu Yoshikawa. 2013. "Impacts of a Prekindergarten Program on Children's Mathematics, Language, Literacy, Executive Function, and Emotional Skills." *Child Development* 84 (6):2112–2130.

Winichakul, K Pun, Guillermo Lezama, Priyoma Mustafi, Marissa Lepper, Alistair Wilson, David Danz, and Lise Vesterlund. 2024. "The effect of experimenter demand on inference." .

Wolf, Sharon, J. Lawrence Aber, Jere R. Behrman, and Edward Tsinigo. 2019. "Experimental impacts of the Quality Preschool for Ghana interventions on teacher professional well-being, classroom quality, and children's school readiness." *Journal of Research on Educational Effectiveness* 12 (1):10–37.

World Bank. 2020. "Nepal Early Childhood Education Diagnostic." Tech. rep., Washington, D.C.

World Bank and UNICEF. 2020a. "Costing Study on Early Childhood Education and Development (ECED) in Nepal: A Case for Investment in ECED." Tech. rep., World Bank, Washington, DC. URL `https://doi.org/10.1596/34574`.

———. 2020b. "Costing Study on Early Childhood Education and Development (ECED) in Nepal: A Case for Investment in ECED." Tech. rep., World Bank and UNICEF, Washington, DC.

Yousafzai, Aisha K., Muneera A. Rasheed, Arjumand Rizvi, Robert Armstrong, and Zulfiqar A. Bhutta. 2014. "Effect of integrated responsive stimulation and nutrition interventions in the Lady Health Worker programme in Pakistan on child development, growth, and health outcomes: a cluster-randomised factorial effectiveness trial." *The Lancet* 384 (9950):1282–1293.

Appendices

Table of Contents

A Appendix Tables **ii**

B Appendix Figures **x**

C Teacher Education Training Sessions **xv**
C.1 Overview xv
C.2 Session Details xv

D Participatory Education Training Sessions **xvii**
D.1 Overview xvii
D.2 Session Details xvii

E ASQ-3 Validation and Measurement **xxi**
E.1 Overview xxi
E.2 Psychometric Properties xxi
E.3 Cross-Cultural Validation xxi
E.4 Use in Economic Research - A few Examples xxii
E.5 Structural Features Reducing Susceptibility to Reporting Bias xxii
E.6 Domain Composition xxii

F ASQ Decomposition: Item Classification **xxiv**
F.1 Overview xxiv
F.2 Summary xxiv
F.3 36-Month Items xxiv
F.4 48-Month Items xxv
F.5 60-Month Items xxvi

G Outcome Variables **xxviii**
G.1 Overview xxviii
G.2 Secondary Outcomes: Learning Environment & Practices and Teacher-Student Interaction xxviii
G.3 Child Primary Outcomes: Ages and Stages Questionnaire Score xxix
G.4 Caregiver Primary Outcomes: Index of Stress and Practice of Early Stimulation and Learning xxxvi
G.5 Care Givers' Secondary Outcomes xxxvii

A Appendix Tables

Table A1: Effects on Parental Education Session Attendance

	Attendance	Attendance Consistency
	(1)	(2)
Training + Helper	1.24*	-0.02
	(0.70)	(0.03)
Training + Facilitator	0.07	-0.02
	(0.72)	(0.02)
Session Number	-0.03*	0.00
	(0.01)	(0.00)
Observations	2,600	2,600
R^2	0.37	0.27

Notes: The table presents results on Parental Education session attendance. The outcome variable is the number of parents attending each session in column (1), and the within-school coefficient of variation in session attendance in column (2). All regressions include municipality fixed effects. *Training Only* is the reference category. *Training + Helper* equals one if a school was assigned to the training plus classroom helper arm, and zero otherwise. *Training + Facilitator* equals one if a school was assigned to the training plus outside classroom helper (facilitator) arm, and zero otherwise. The data come from Parental Education Session Attendance Records, covering 130 treatment schools that implemented parental education sessions and for which attendance data were collected, totaling 2,600 session-level observations. Of the 150 treatment schools, 12 did not implement Parental Education sessions (9 had not sent teachers for Parental Education training, and 3 had facilitators who opted out), and attendance data were not collected for 8 additional schools. Standard errors are clustered at the school level and reported in parentheses. ***, **, and * denote statistical significance at the 1, 5, and 10 percent levels, respectively.

Table A2: Balance Table

	Overall	Control	Training Only	Training + Helper	Training + Facilitator	Training Only vs. Control	Training + Helper vs. Control	Training + Facilitator vs. Control
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Demographics								
Child's Age (months)	47.644	48.131	48.095	47.781	46.570	-0.035	-0.350	-1.560
	(12.000)	(12.550)	(11.872)	(12.220)	(11.279)	(1.100)	(1.178)	(1.073)
Child's Gender (female)	0.503	0.495	0.499	0.519	0.500	0.005	0.024	0.005
	(0.500)	(0.500)	(0.500)	(0.500)	(0.500)	(0.024)	(0.024)	(0.022)
Number of Adults in HH	3.135	3.115	3.007	3.233	3.189	-0.108	0.118	0.074
	(1.533)	(1.433)	(1.442)	(1.581)	(1.661)	(0.093)	(0.095)	(0.130)
Number of Siblings	1.253	1.185	1.338	1.225	1.265	0.153	0.039	0.079
	(1.130)	(1.039)	(1.189)	(1.154)	(1.129)	(0.099)	(0.090)	(0.104)
Asset Index	7.800	7.803	7.721	7.836	7.841	-0.082	0.033	0.038
	(3.315)	(3.319)	(3.241)	(3.501)	(3.203)	(0.306)	(0.313)	(0.286)
Caregiver's Gender (female)	0.835	0.816	0.847	0.833	0.845	0.031	0.017	0.029
	(0.371)	(0.388)	(0.360)	(0.373)	(0.362)	(0.025)	(0.027)	(0.025)
Caregiver's Education	2.834	2.898	2.777	2.810	2.849	-0.121	-0.089	-0.049
	(1.374)	(1.330)	(1.389)	(1.399)	(1.379)	(0.118)	(0.129)	(0.120)
Panel B: Developmental Outcomes and Caregiver Attitude								
Baseline ASQ Score	0.036	0.000	0.108	0.099	-0.061	0.108	0.099	-0.061
	(0.950)	(0.999)	(0.895)	(0.919)	(0.983)	(0.078)	(0.082)	(0.078)
Confidence towards Disability Index	-0.005	0.000	0.071	-0.052	-0.041	0.071	-0.052	-0.041
	(1.033)	(1.000)	(1.101)	(0.980)	(1.042)	(0.078)	(0.073)	(0.081)
Attitude towards Disability Index	-0.096	0.000	-0.076	-0.134	-0.175	-0.076	-0.134*	-0.175**
	(0.929)	(1.000)	(0.917)	(0.903)	(0.887)	(0.069)	(0.075)	(0.068)
Baseline Communication Index	0.044	0.000	0.106	0.116	-0.043	0.106	0.116	-0.043
	(0.950)	(0.999)	(0.878)	(0.930)	(0.977)	(0.078)	(0.084)	(0.085)
Problem Solving Index	-0.018	0.000	0.018	-0.015	-0.076	0.018	-0.015	-0.076
	(0.997)	(0.999)	(0.990)	(0.972)	(1.024)	(0.077)	(0.077)	(0.072)
Baseline PESL Index	0.037	0.000	-0.040	0.043	0.146	-0.040	0.043	0.146**
	(0.989)	(1.000)	(0.985)	(0.972)	(0.990)	(0.061)	(0.072)	(0.070)
Baseline Stress Index	0.066	0.000	-0.009	0.116	0.160	-0.009	0.116	0.160
	(1.073)	(1.000)	(1.059)	(1.077)	(1.145)	(0.091)	(0.098)	(0.099)
Respondent thinks child has Disability	0.096	0.164	0.024	0.164	0.034	-0.141	0.000	-0.130
	(2.559)	(3.580)	(0.152)	(3.679)	(0.182)	(0.127)	(0.186)	(0.127)
Observations	3,004	766	757	725	756			

Notes: The table reports descriptive statistics and balance tests, showing group means and mean differences between treatment and control schools. Columns (1)–(5) present means for the *Overall* sample, the *Control* group, the *Training Only* group, the *Training + Helper* group, and the *Training + Facilitator* group, respectively. Columns (6)–(8) report differences in means between each treatment arm and the control group. Standard errors clustered at the school level are reported in parentheses. The analysis sample comprises 3,004 student respondents from the Endline Survey. ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively.

Table A3: Treatment Effects on Child Development and Caregiver Outcomes with Baseline Controls

	ASQ Index	PESL Index	Caregiver Stress Index
	(1)	(2)	(3)
Panel A: Pooled Treatment Arms			
[0.5em] Any Treatment	0.13***	0.11**	0.03
	(0.05)	(0.05)	(0.05)
Observations	3,004	3,003	2,987
Panel B: Combining Training and Help			
[0.5em] Training Only	0.16***	0.14**	-0.01
	(0.05)	(0.06)	(0.06)
Training plus any help	0.10*	0.09**	0.05
	(0.05)	(0.05)	(0.05)
Observations	3004	3003	2987
Panel C: Individual Treatment Arms			
[0.5em] Training Only	0.17***	0.13**	-0.01
	(0.05)	(0.06)	(0.06)
Training + Helper	0.12*	0.14**	-0.01
	(0.06)	(0.06)	(0.06)
Training + Facilitator	0.10	0.06	0.10*
	(0.06)	(0.05)	(0.06)
Observations	3004	3003	2987
Baseline Control	Yes	Yes	Yes

Notes: The table reports results for the main outcomes estimated using regression equation (5.1), with baseline controls included for each outcome. Outcomes are standardized relative to the control group: the *ASQ Index* (Ages and Stages Questionnaire, a measure of child development), the *PESL Index* (Practices of Early Stimulation and Learning, capturing parental engagement), and the *Caregiver Stress Index*. Baseline controls correspond to the same measure at baseline (e.g., baseline ASQ score for ASQ outcomes, baseline PESL score for PESL outcomes, and baseline Stress score for Stress outcomes). All regressions include municipality fixed effects. *Any Treatment* equals one if a school received *Training Only*, *Training + Helper*, or *Training + Facilitator*, and zero otherwise. The analysis sample comprises 3,004 student respondents from the Endline Survey. Standard errors are clustered at the school level and reported in parentheses. ***, **, and * denote statistical significance at the 1, 5, and 10 percent levels, respectively.

Table A4: Treatment Effects on Child Development across ASQ Index Sub-Components

	ASQ	Communication	Gross Motor	Fine Motor	Problem-solving	Personal-social
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Pooled Treatment Arms						
Any Treatment	0.13*	0.15**	0.09*	0.12**	0.10	0.08
	(0.07)	(0.06)	(0.04)	(0.06)	(0.07)	(0.06)
Observations	3,004	1,455	3,004	3,004	3,004	3,004
Panel B: Combining Training and Help						
Training Only	0.20**	0.22***	0.13***	0.16**	0.17**	0.14*
	(0.08)	(0.07)	(0.05)	(0.07)	(0.08)	(0.07)
Training plus any help	0.09	0.12*	0.06	0.10	0.06	0.05
	(0.07)	(0.07)	(0.05)	(0.06)	(0.08)	(0.07)
Observations	3004	1455	3004	3004	3004	3004
Panel C: Individual Treatment Arms						
Training Only	0.20**	0.22***	0.13***	0.16**	0.17**	0.14*
	(0.08)	(0.07)	(0.05)	(0.07)	(0.08)	(0.07)
Training plus Helper	0.15*	0.17**	0.02	0.13	0.13	0.12
	(0.08)	(0.07)	(0.06)	(0.08)	(0.09)	(0.08)
Training plus Facilitator	0.04	0.07	0.10**	0.08	-0.01	-0.01
	(0.08)	(0.08)	(0.05)	(0.07)	(0.08)	(0.08)
Observations	3004	1455	3004	3004	3004	3004

Notes: The table reports results for the main outcomes estimated using regression equation (5.1), with ASQ sub-components. Outcomes are standardized relative to the control group: the *ASQ* (Ages and Stages Questionnaire, a measure of child development) is an aggregated score across its sub-components presented in columns (2) through (6). All regressions include municipality fixed effects. *Any Treatment* equals one if a school received *Training Only*, *Training + Helper*, or *Training + Facilitator*, and zero otherwise. The analysis sample comprises 3,004 student respondents from the Endline Survey. Standard errors are clustered at the school level and reported in parentheses. ***, **, and * denote statistical significance at the 1, 5, and 10 percent levels, respectively.

Table A5: Pairwise Between-Arm Differences and Statistical Power

Outcome	Comparison	Difference	SE	p-value	MDE	Power
ASQ Index	T1 vs T2	0.053	0.060	0.382	0.169	14%
ASQ Index	T1 vs T3	0.139**	0.058	0.017	0.163	67%
ASQ Index	T2 vs T3	0.087	0.066	0.185	0.184	26%
PESL Index	T1 vs T2	−0.047	0.059	0.430	0.165	12%
PESL Index	T1 vs T3	0.005	0.062	0.930	0.173	3%
PESL Index	T2 vs T3	0.052	0.059	0.375	0.164	14%
Stress Index	T1 vs T2	−0.000	0.059	0.998	0.164	3%
Stress Index	T1 vs T3	−0.117**	0.054	0.031	0.151	58%
Stress Index	T2 vs T3	−0.117*	0.061	0.055	0.170	48%

Notes: Differences computed using `lincom` after regressing each outcome on treatment arm indicators with municipality fixed effects and standard errors clustered at the school level. MDE = minimum detectable effect at 80% power (= 2.80 × SE). Power = ex-post power at the observed difference (= $\Phi(|\text{diff}|/\text{SE} - 1.96)$). Positive ASQ/PESL differences indicate T1 outperforms the comparator; negative Stress differences indicate T1 has lower (better) stress. ***, **, * denote significance at 1, 5, and 10 percent levels.

Table A6: Attrition

Predictor Variable	Coefficient	P-value
	(1)	(2)
Panel A: Differential Attrition (Treatment vs. Control)		
Training Only	0.029	0.20
	(0.023)	
Training + Helper	0.022	0.25
	(0.019)	
Training + Facilitator	0.006	0.52
	(0.010)	
Panel B: Baseline Predictors of Attrition		
Age of the Child	-0.000	0.42
	(0.000)	
Gender of Child (Female = 1)	-0.010*	0.09
	(0.006)	
Number of Adults in HH	-0.000	0.99
	(0.002)	
Number of Children under 18	0.003	0.42
	(0.004)	
Asset Index	-0.003**	0.01
	(0.001)	
Gender of Caregiver (Female = 1)	-0.015	0.12
	(0.010)	
Education of Caregiver	-0.004*	0.07
	(0.002)	
Baseline PESL Index	-0.003	0.40
	(0.003)	
Baseline Stress Index	-0.003	0.44
	(0.003)	
Baseline ASQ Index	-0.003	0.54
	(0.004)	

Notes: The table reports attrition results from the Endline Survey. Panel A presents coefficients from OLS regressions of an attrition indicator, equal to one if a student was not observed at endline and zero otherwise, on treatment arm indicators. Panel B reports coefficients from OLS regressions of the same attrition indicator on baseline child and household characteristics. All regressions include municipality fixed effects. The *ASQ Index* refers to the Ages and Stages Questionnaire (a measure of child development), the *PESL Index* refers to the Practices of Early Stimulation and Learning (capturing parental engagement), and the *Caregiver Stress Index* measures self-reported stress among primary caregivers. Standard errors are clustered at the school level and reported in parentheses, with corresponding p-values shown in Column (2).

Table A7: Lee Bounds Estimates of Treatment Effects

Outcome	Lower Bound [95% Conf. Interval]	Upper Bound [95% Conf. Interval]
	(1)	(2)
ASQ Index	0.066	0.229
	[-0.022, 0.153]	[0.136, 0.322]
PESL Index	-0.116	0.121
	[-0.189, -0.043]	[0.042, 0.199]
Stress Index	-0.119	0.051
	[-0.220, -0.019]	[-0.049, 0.151]

Notes: The table reports upper and lower bounds on treatment effects for the *ASQ Index* (Ages and Stages Questionnaire, a measure of child development), the *PESL Index* (Practices of Early Stimulation and Learning, capturing parental engagement), and the *Caregiver Stress Index*, accounting for attrition from the Endline Survey. Bounds are estimated following Lee (2009). Ninety-five percent confidence intervals are reported in brackets beneath each estimate.

Table A8: Non-Compliance in Parental Education Sessions

Baseline Characteristic	Coefficient	P-value
	(1)	(2)
Age of the Child	0.894	0.42
	(1.105)	
Gender of Child (Female = 1)	0.031	0.30
	(0.030)	
Number of Adults in HH	0.313**	0.03
	(0.143)	
Number of Siblings	0.102	0.28
	(0.094)	
Asset Index	0.023	0.94
	(0.297)	
Gender of Caregiver (Female = 1)	-0.067*	0.05
	(0.035)	
Education of Caregiver	-0.356**	0.01
	(0.140)	
Baseline ASQ Index	0.031	0.67
	(0.073)	
Baseline PESL Index	0.019	0.79
	(0.072)	
Baseline Stress Index	-0.012	0.92
	(0.125)	

Notes: The table reports associations between non-compliance in parental education sessions and baseline child and household characteristics. Each row reports coefficients from regressing the corresponding baseline characteristic on an indicator variable equal to one if a student's school did not conduct parental education sessions and zero otherwise. Coefficients are estimated using OLS with municipality fixed effects. The *ASQ Index* refers to the Ages and Stages Questionnaire (a measure of child development), the *PESL Index* refers to the Practices of Early Stimulation and Learning (capturing parental engagement), and the *Caregiver Stress Index* measures self-reported stress among primary caregivers. Standard errors are clustered at the school level and reported in parentheses, with corresponding p-values shown in Column (2).

Table A9: Benchmarking Program Effects and Costs

Study	Country	Main Outcomes	Effect Size (SD)	Program Type/Cost-Effectiveness
A. Teacher Professional Development Center-Based Quality Reforms				
Nepal ECD (this study)	Nepal	Ages and Stages Questionnaire (ASQ); Practices of Early Stimulation and Learning (PESL)	+0.13–0.20 (ASQ); +0.11–0.20 (PESL)	Teacher training + parental engagement; modest cost; no formal benefit–cost ratio (BCR) yet
Andrew et al. (2024)	Colombia	Cognitive latent factor	+0.16 overall; +0.30 poorest	Teacher professional development (Hogares Infantiles); $35 upfront + $13 annually per child
Wolf et al. (2019)	Ghana	School readiness; teacher well-being	+0.11 (cognitive); +0.18 (socio-emotional)	Quality preschool reform; BCR ≈ 11–103:1
Gallego, Näslund-Hadley, and Alfonso (2021)	Peru	Cognitive test scores	+0.19	Pedagogical reform in preschools; BCR ≈ 29–88
Brinkman et al. (2017)	Indonesia	Socio-emotional skills	+0.16	ECE expansion in rural areas; cost-effective at scale
Spier et al. (2020)	Bangladesh	Cognitive and socio-emotional	+0.25 (cognitive); +0.37 (socio-emotional)	Preschool initiative; BCR ≈ 3–29
Dillon et al. (2017)	India	Intuitive vs. formal mathematics	+0.29–0.32 (nonsymbolic, durable); +0.13 (symbolic, fades); no durable gains on formal math	Preschool math enrichment; ≈ $316 per group of 6 children over 4 months
B. Parenting Groups/Home-Visiting				
Justino et al. (2023)	Rwanda	ASQ indices; parenting practices	+0.30–0.38 (ASQ); +0.47–0.62 (parent–child activities)	Parenting group model (First Steps); ≈ $2 per caregiver per session
Andrew et al. (2020)	India (urban)	Bayley-III developmental indices	+0.30 overall; +0.35 cognitive	Home-visiting program; ≈ $148 per child per year
C. Financial Access/Scholarship Subsidies				
Dean and Jayachandran (2020)	India	Cognitive composite; school attendance	+0.80 at end of kindergarten	Kindergarten scholarship subsidy; cost not reported
D. Cash Transfers				
López Bóo and Creamer (2019)	Honduras	ASQ (cognitive)	+0.13	Conditional cash transfer; ≈ $500 per household per year
E. State Capacity/System Reform				
Ganimian, Muralidharan, and Walters (2024)	India (Tamil Nadu)	Math, language, executive function, nutrition	+0.29–0.46 (tests); +0.10–0.11 (nutrition, HH tests)	ICDS facilitator model; BCR ≈ 13:1–22:1
F. Access/Quality Expansion with Null Effects				
Blimpo et al. (2022)	Gambia	Motor and language/hearing (MDAT)	Null	Expansion/quality trial; low cost, no effects

Notes: ASQ = Ages and Stages Questionnaire. PESL = Practices of Early Stimulation and Learning. MDAT = Malawi Developmental Assessment Tool. ICDS = Integrated Child Development Services. BCR = Benefit–Cost Ratio. Effect sizes reported in standard deviations (SD). Costs expressed in 2020 US dollars where available.

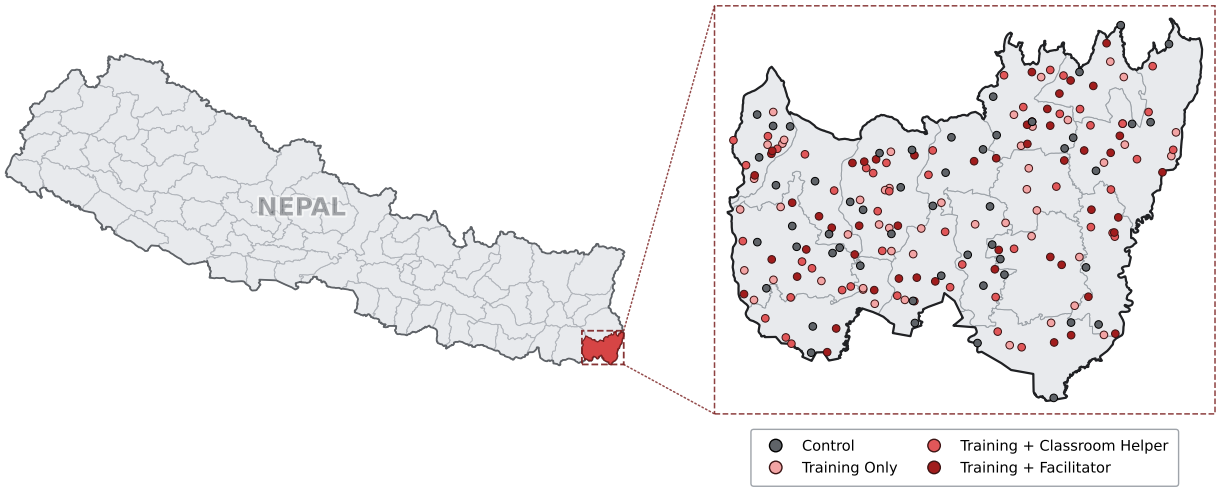
B Appendix Figures

Figure A1: Location of Jhapa District within Nepal



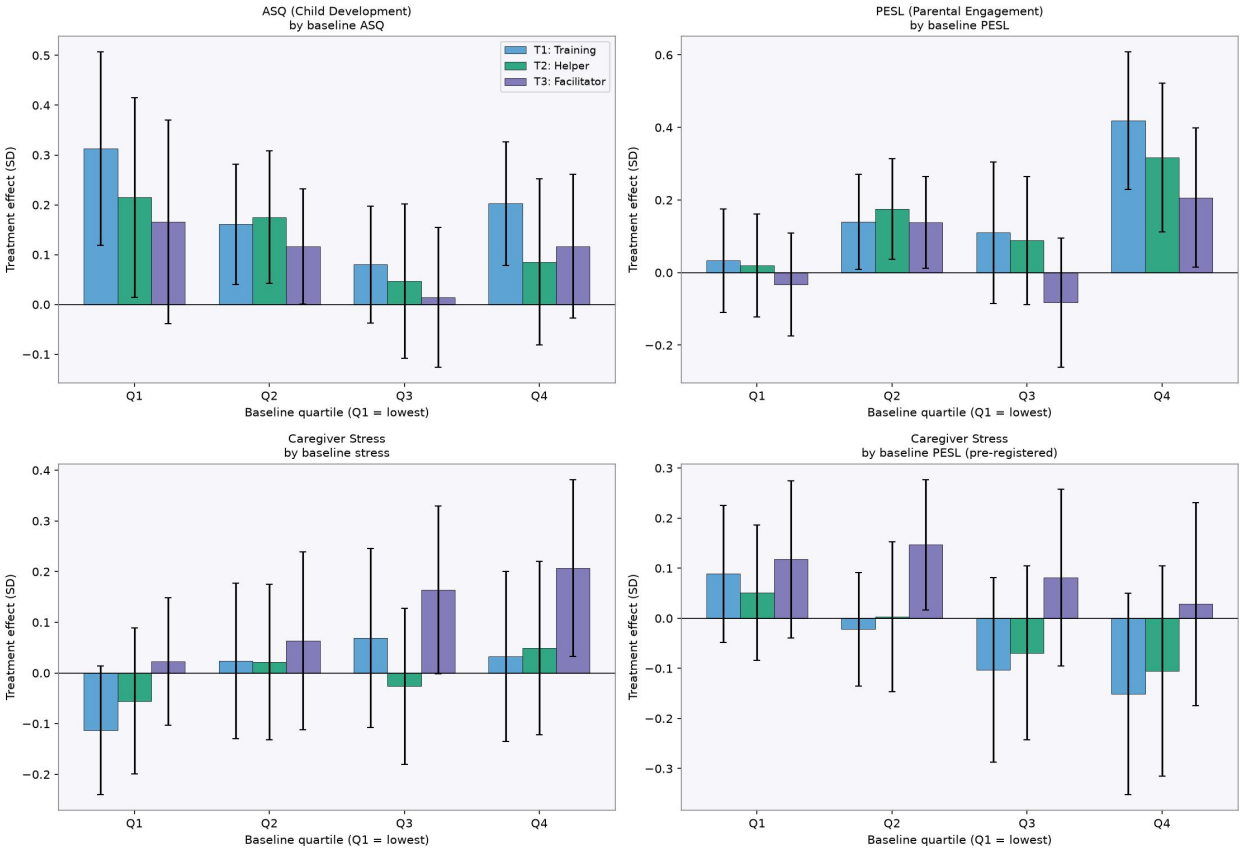
Notes: This figure shows the location of Jhapa District (highlighted), the study site, within Nepal. Jhapa is Nepal's easternmost district, in the southeastern corner of the country, bordering India.

Figure A2: Map of Distribution of Treatment and Control Schools across Jhapa District



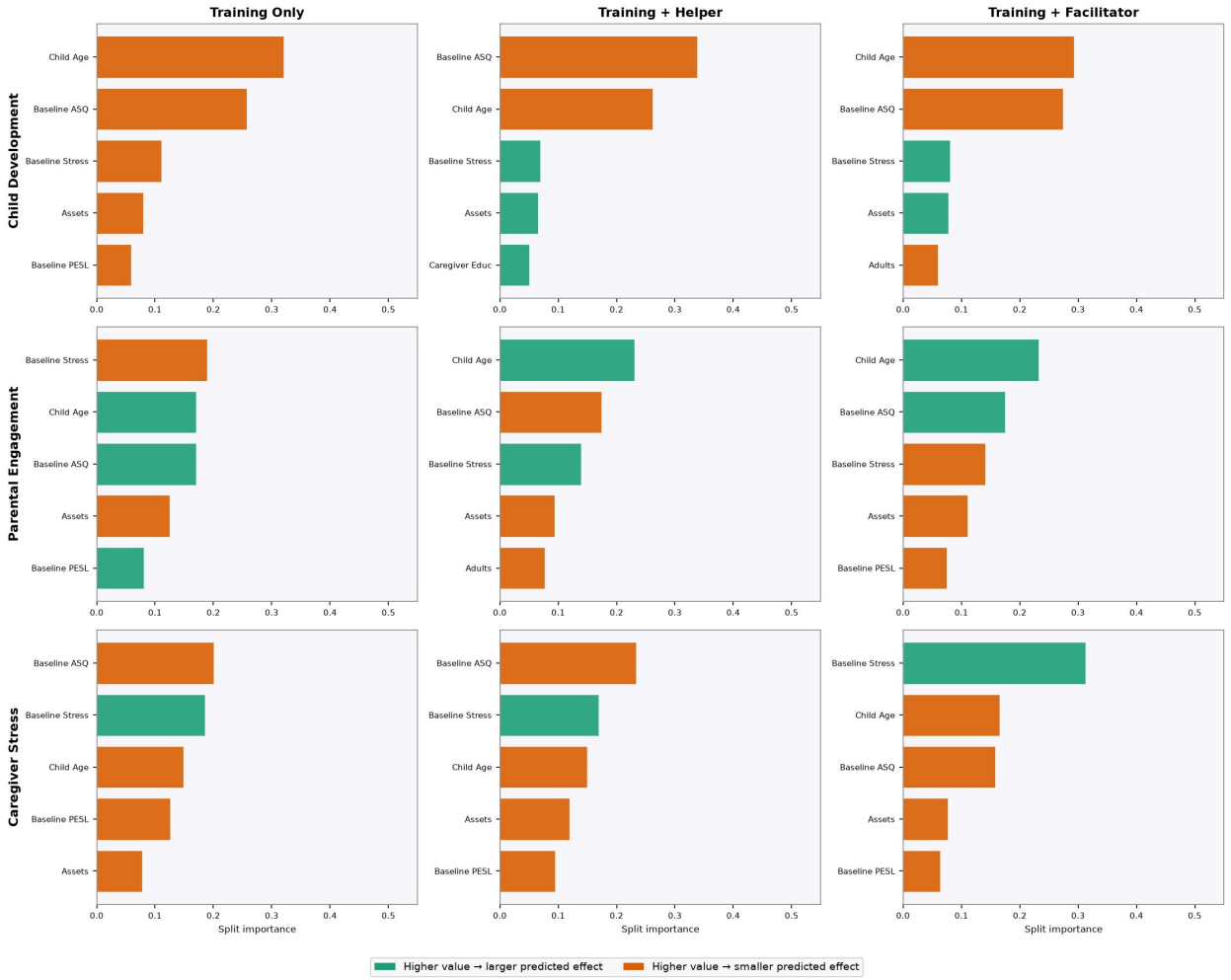
Notes: This figure plots the location of the 51 *Control* and 150 *Treatment* schools in this study. *Treatment* schools include 50 schools from each of the three treatment arms (*Training Only*, *Training + Helper*, and *Training + Facilitator*).

Figure A3: Pre-Specified Heterogeneity: Treatment Effects by Baseline Quartile



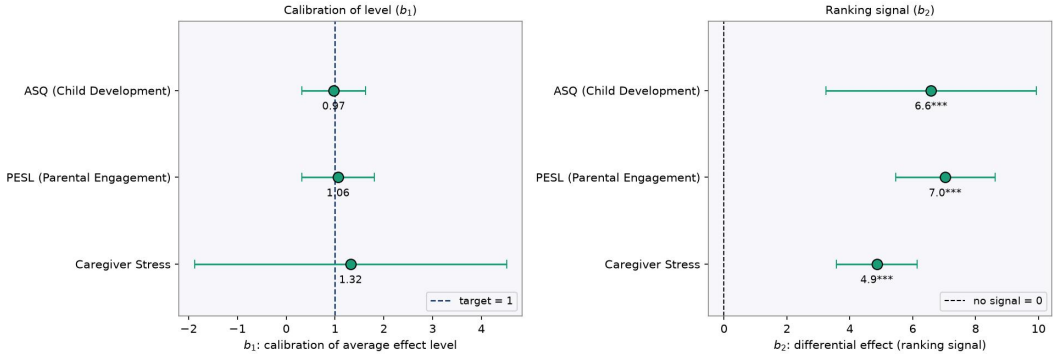
Notes: Each bar is the treatment effect for one arm within a quartile of the baseline distribution, from an OLS regression of the endline outcome on arm-by-quartile interactions, quartile indicators, and municipality fixed effects, with standard errors clustered at the school level (90% confidence intervals). Panels one to three split each outcome by quartiles of its own baseline; the fourth panel splits caregiver stress by baseline parental engagement, the pre-registered dimension. For caregiver stress, a positive effect denotes an increase.

Figure A4: Importance and Direction of Baseline Characteristics in the Causal Forest



Notes: Each panel corresponds to one outcome (rows) and one treatment arm relative to control (columns). Within a panel, the bars report the five baseline characteristics on which the forest splits most often; the length of each bar is that characteristic's variable importance, defined as its share of forest splits. The color of each bar records the direction in which the characteristic enters, measured by the sign of the correlation between the characteristic and the predicted treatment effect across children: a characteristic shaded teal predicts a larger effect at higher values, and one shaded orange predicts a smaller effect at higher values. For child development and parental engagement a larger effect is beneficial; for caregiver stress a larger effect denotes a greater increase in stress. The figure is the source of the importance shares and signed directions discussed in the text.

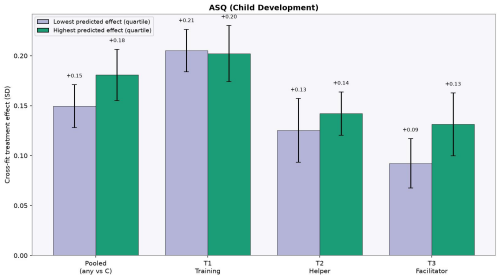
Figure A5: Calibration of Causal-Forest Predictions



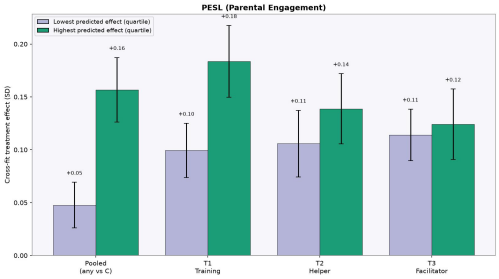
Notes: The figure implements the best-linear-predictor (BLP) test of Chernozhukov et al. (2018b), which assesses whether the causal forest's predicted effects are informative about actual treatment effects. Because the forest selects the dimensions of heterogeneity from the data, its predictions require external validation before they can be interpreted. The outcome is regressed on the average predicted effect and on each child's deviation from that average, yielding two coefficients. The coefficient on the average predicted effect (b_1) measures calibration of the overall level; a value near one, as obtained for child development and parental engagement, indicates that the forest recovers the average treatment effect. The coefficient on the deviation (b_2) tests for genuine heterogeneity; a positive and significant value, obtained for all three outcomes, indicates that children predicted to benefit more do benefit more. Only the sign and significance of b_2 are interpretable, as its magnitude is not normalized to one. Bars denote 90 percent confidence intervals.

Figure A6: Cross-Fit Group Treatment Effects (Bottom vs. Top Quartile of Predicted Effect)

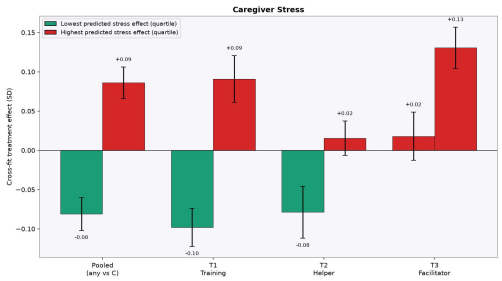
(a) Child Development



(b) Parental Engagement



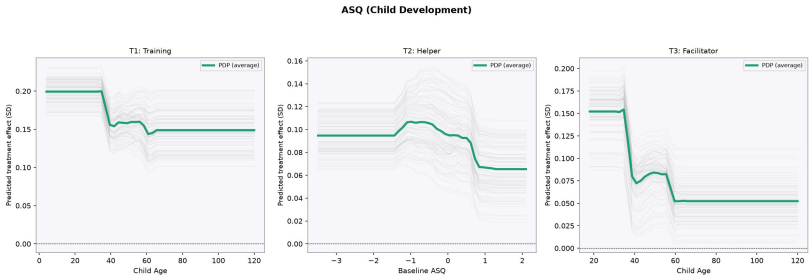
(c) Caregiver Stress



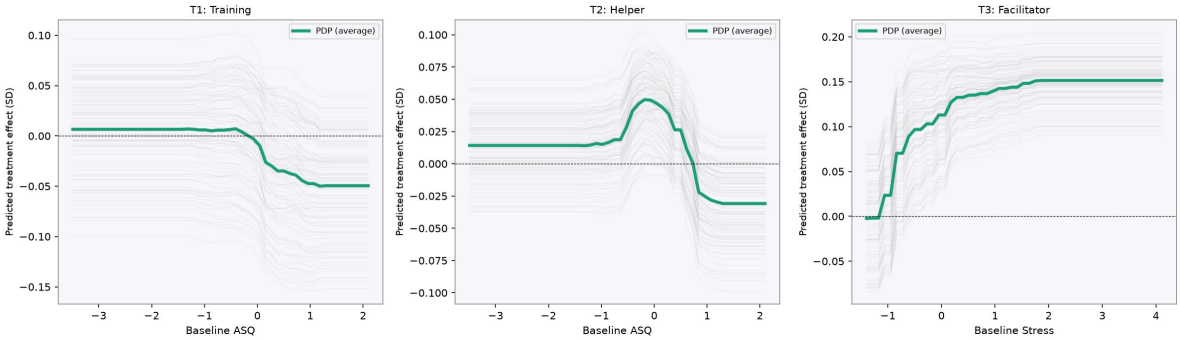
Notes: The figure reports treatment effects for children in the bottom and top quartiles of the forest's predicted effect, estimated separately within each group, and provides a cross-fitted validation of the estimated heterogeneity. Sorting observations by an estimated effect and re-estimating within the resulting groups can induce mechanical differences, because the prediction is constructed from the same outcomes used in estimation. To address this, the sample is repeatedly partitioned in half; the forest is trained on one partition and used to assign observations in the other to groups, so that the observations defining the groups are never those used to estimate the group effect. The predicted ordering is preserved. For child development, the bottom and top quartiles gain 0.21 and 0.20 SD under Training Only, with all groups gaining; for parental engagement the corresponding estimates are 0.10 and 0.18 SD; and for caregiver stress the top quartile under the Facilitator arm increases by 0.13 SD, while the bottom quartile under Training Only declines. Effects are estimated by OLS within each group, with municipality fixed effects and standard errors clustered at the school level; bars denote 90 percent confidence intervals from twenty repeated sample partitions. For caregiver stress, a positive value denotes an increase.

Figure A7: Partial Dependence of the Predicted Effect on the Leading Characteristic, by Arm

(a) Child Development

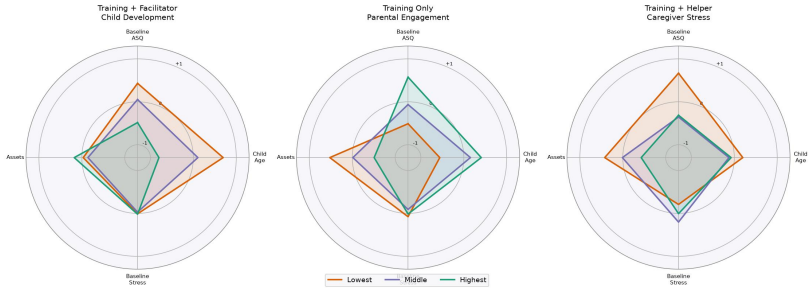


(b) Caregiver Stress



Notes: The figure plots the partial dependence of the predicted treatment effect on each arm's most important baseline characteristic. In each panel, the bold line traces the average predicted effect as that characteristic varies, holding the remaining characteristics at their observed values, and the faint lines trace the same relationship for a random sample of individual children. The near-parallel individual curves indicate that the characteristic moves the predicted effect in the same direction across children rather than through a small number of influential cases, and the smooth, monotone average relationship indicates no threshold at which the effect changes discontinuously. Panels are shown for child development and caregiver stress, the outcomes discussed in the text.

Figure A8: Baseline Profile of Predicted Groups: Remaining Arm–Outcome Combinations



Notes: The figure reports the arm–outcome combinations not shown in Figure 6: child development under the Facilitator arm, parental engagement under Training Only, and caregiver stress under the Helper arm. Each panel plots the average baseline profile of children in the lowest, middle, and highest tercile of the forest-predicted effect for that combination. Each axis is one baseline characteristic, with baseline ASQ, child age, and baseline stress held in fixed positions across panels and the fourth position occupied by that panel's next most important characteristic. Values are standardized across the estimation sample, so a value of +1 indicates the group averages one standard deviation above the sample mean and a larger radius indicates a higher value. For caregiver stress, the highest tercile is the group with the largest predicted increase in stress. The patterns match those in Figure 6: children with lower baseline ASQ and younger children are predicted to gain most in child development, and families with higher baseline stress are those whose stress rises most.

C Teacher Education Training Sessions

C.1 Overview

This appendix provides a summary of the 10-day education training sessions for teachers and facilitators of the parental education sessions.

C.2 Session Details

1. **Conceptual Understanding of Early Childhood Development.** On Day I, teachers engaged in exploring the foundations of child development. They began with introductions and a pre-test, then delved into understanding ECD concepts from multiple perspectives including brain development and early learning. The day focused on major domains of child development (physical, social, emotional, cognitive, and language), with interactive activities such as group presentations, brainstorming sessions, and a debate on biological versus environmental factors influencing child growth.

2. **Child Growth and Development.** On Day II, teachers explored developmental milestones from prenatal stages through early childhood, covering key periods from birth to 96 months. The session progressed to childhood disabilities, focusing on types, causes, and preventive strategies. Teachers engaged in discussions about inclusive education and analyzed child behavior and disorders, developing skills to recognize and address behavioral challenges.

3. **ECD Policies, Child Health, Nutrition, Care.** On Day III, teachers explored ECD policies, child rights, health, and nutrition. They discussed constitutional provisions, child rights based on UNCRC, and the holistic development of children in the Nepalese context. The day covered the ECD National Strategy 2077, the Nurturing Care Framework, and roles of local authorities, parents, and stakeholders in ECD implementation.

4. **ECD Programs.** On Day IV, teachers explored ECD program management and implementation strategies, examining various program models including parenting programs, home-based, community-based, and school-based initiatives. The session covered ECD center management, daily operations, physical and learning environments, and roles of teachers, parents, and community members.

5. **ECD Curriculum.** On Day V, teachers explored the ECD national curriculum, its components, implementation strategies, and holistic child development approach. They examined the 2061 National Curriculum, its goals, learning environments, and instructional strategies. Activities included paired discussions on integrated thematic learning and collaborative creation of theme-based web charts linking developmental domains and learning concepts.

6. **Early Learning and Development.** On Day VI, teachers focused on learning assessment, play-based pedagogical approaches, and practical implementation strategies. They explored assessment techniques, observation methods, and documentation strategies for tracking developmental progress. The day included hands-on activities designing observation tools and understanding how to use assessment data to inform teaching strategies.

7. **Instruction and Learning / Child Learning and Development Assessment.** Day VII was divided into two parts. In the first half, teachers examined child-centered instruction strategies, early learning principles, and effective approaches to educational engagement through hands-on group activities. In the second half, teachers explored comprehensive child assessment strategies focusing on continuous and holistic evaluation methods, including direct and indirect approaches such as observation, interaction, questioning, and performance analysis.

8. **Teacher's Performance Development (TPD) for ECD Facilitators.** On Day VIII, teachers focused on professional development for ECD facilitators, examining TPD conceptual frameworks, importance, and practical implementation. The session included presentations from HCRD authorities on performance evaluation methods, test types, and assessment tools for educators.

9. **Model Training Curriculum.** On Day IX, teachers explored the TPD Model Training Curriculum, developing a detailed 10-day facilitator training program with session contents and contextual facilitation plans. Activities included practical simulation of demonstration teaching and facilitation techniques.

10. **Project Work, Field-Based Learning, and Evaluation.** On Day X, teachers engaged in practical application, project implementation, and final assessment. They developed project work plans for ECD centers, prepared comprehensive action plans, and participated in a multi-faceted evaluation process. The training concluded with feedback collection and a closing ceremony.

D Participatory Education Training Sessions

D.1 Overview

This appendix provides a summary of the 20 participatory education training sessions, outlining key topics and learning objectives.

D.2 Session Details

1. **Introduction and Importance of Early Child Development.** Introduces the critical importance of early childhood (pregnancy to age 8) for physical, cognitive, emotional, and social development. Key points include rapid brain growth, building a foundation for lifelong learning, and parents' role in providing nurturing care, safety, and equal opportunities. Discussion topics include perceptions of early childhood and challenges, with a focus on the responsibility of parents to foster healthy development through active participation and positive role modeling.

2. **Role of Father in Early Child Development.** Defines the role of the father in early childhood development and shared parental responsibilities because a nurturing father-child relationship is beneficial to the well-being of a child. Fathers should provide prenatal support, emotional bonding, be actively involved in supporting mother and child. Discussion focuses on challenges and ways in which fathers can stay engaged.

3. **Early Child Development Programs and Responsible Sections and Institutions.** This session informs parents about local government and non-government programs that support early childhood development. These include parenting education, home-based development, entry-level childcare, and early childhood programs focused on children's physical, cognitive, and emotional growth. Key takeaways for parents include identifying available programs and ensuring that essential needs such as vaccination, healthcare, nutrition, clean drinking water, security, and early education are met.

4. **Care of Mothers, Children During Pregnancy, and Newborn Infants.** Provides a comprehensive guide for caring for mother and child from pregnancy to early infancy, covering maternal and child health, family support, preparation for birth, and early newborn needs. Parents learn about fetal development, nutrition, safe delivery practices, early infant care, and developmental milestones. The session also addresses postpartum recovery and emotional well-being, ensuring parents feel confident in their journey from pregnancy to parenthood.

5. **Stimulation In Early Childhood.** Emphasizes the importance of sensory and kinesthetic stimulation in brain development. Parents learn how visual, auditory, tactile, gustatory, and kinesthetic stimulation promotes brain development and learning. The discussion covers the benefits of stimulation, the risks of its absence, and the caregivers' role in creating an environment that fosters cognitive, emotional, and physical development.

6. **Components of Early Child Development.** Details five key aspects of child development: physical, cognitive, linguistic, emotional, and social. Parents learn all aspects are interrelated and that nutrition, stimulation, health services, protection, care, and learning opportunities

encourage holistic development. Parental responsibilities include providing proper health and education, as well as fostering a supportive environment through consistent engagement and role modeling. Discussion explores challenges and how to best support development.

7. **Early Childhood Development Stages and Characteristics.** Outlines developmental phases of early childhood, as well as milestones and age-appropriate activities that align with each stage. Parents are taught to identify each stage and what factors contribute to well-rounded growth. This session encourages discussion on observed child behaviors, ways to accelerate development, and importance of engagement, protection, and proper care in early childhood.

8. **Early Detection of Children with Special Needs.** Parents learn how to identify disabilities in infants and children, along with the importance of early detection and appropriate support. Learning objectives include familiarity with different types of disabilities, their causes, preventative measures, and indicators for early diagnosis. Preventive measures include maternal care, vaccinations, and safe childbirth practices. Parents and guardians are encouraged to focus on their children's abilities rather than their limitations, as it is important to create an inclusive environment where children with disabilities receive equal opportunities for development. Parents of children with disabilities are encouraged to seek available resources, including disability identity cards and support services.

9. **Causes and Prevention of Child Malnutrition.** Malnutrition is taught as an imbalance between the essential elements required by the body that can be identified as acute malnutrition (wasting), chronic malnutrition (stunting), underweight, or micronutrient deficiencies. Parents learn its causes, symptoms, and prevention strategies. Causes include hereditary factors, lack of awareness, insufficient maternal nutrition, food insecurity, vitamin and mineral deficiencies, and frequent illness. Prevention methods involve maternal and child nutrition, exclusive breastfeeding, regular health check-ups, hygiene, and utilizing government nutrition programs. Parents play a key role in providing balanced diets and monitoring growth.

10. **Child Health and Care.** Parents learn the importance of nurture and timely treatment in protecting children from illnesses. This involves the ability to identify symptoms, their causes, preventive measures, and provide appropriate care—including emotional support. Common childhood diseases are fever, pneumonia, diarrhea, and infections. Home remedies and medical interventions for various illnesses are discussed, stressing the need for attentive caregiving and prompt medical consultation when necessary.

11. **Prevention of Harmful Contagious Diseases In Children.** Brings awareness to child-fatal contagious diseases and their prevention through full immunization. Diseases such as tetanus, tuberculosis, diphtheria, polio, and measles, along with their symptoms, transmission methods, and available vaccines are discussed. Parents are encouraged to follow immunization schedules, recognize early symptoms, and seek immediate medical care when necessary. The government provides free vaccines to all children, so parents must get their children vaccinated.

12. **Child Behavior and Positive Discipline; Protection of Children from Violence, Abuse, and Punishment.** Children develop habits and behavior by observing their parents and environment. Parents learn their importance as role models and reinforcers of good habits through praise, consistency, and positive engagement. Harsh discipline and neglect negatively

effect personality development, while encouragement and structured support help children manage emotions and behavior effectively. Violence, abuse, and discrimination have lasting detrimental effects and are violations of children's rights. A nurturing, non-discriminatory environment with clear expectations, encouragement, and legal protections is essential for healthy child development.

13. **Child Safety, Protection from Accidents, and Learning in Emergencies.** Creating a safe environment and emergency preparedness at home and in the community protects children from accidents, ensures their safety and continued development. Children are naturally curious and may not foresee dangers so childproofing and risk education through stories and demonstrations are essential. Preventative measures include securing hazardous objects and keeping them out of reach, teaching road safety, and supervising child activities. In emergencies, knowing basic first aid—such as treating burns, cuts, and fractures—can be lifesaving. Disaster preparedness involves identifying hazards, planning emergency responses, storing essential supplies, and practicing evacuation drills.

14. **Importance of Early Childhood Learning.** Introduces the importance of early childhood learning through play and experience-based methods. Babies learn from birth by engaging with their environment using their senses, imitation, problem-solving, and exploration. These early experiences significantly effect brain development due to rapid formation of neural connections in the first three years. Age-appropriate sports and activities are suggested for cognitive and social development. The role of parents is to provide a stress-free, nurturing environment, engage in the play, and encourage curiosity through storytelling and conversation. Discussion includes children's learning processes and suitable play activities.

15. **Child-Centered Learning Activities.** Children learn best when provided with opportunities and an environment suited to their age, interests, and abilities. This session focuses on child-centered learning that is safe, healthy, inclusive, and stimulating. Child-centered activities include sports, storytelling, music, dance, drawing, and nature observation. These help develop creativity, communication skills, confidence, problem-solving abilities, and leadership. Parents' role is to create spaces for play and ensuring learning aligns with a child's natural curiosity.

16. **Utilization of Local Learning Materials.** Discusses the use and availability of homemade toys and locally sourced learning materials in a child's development. Homemade toys crafted from readily available resources are cost-effective and culturally relevant and promote creativity, emotional bonding, and cognitive skills. Locally made toys like dolls, puppets, soft balls, and pictorial storybooks aid in language development, motor skills, and social learning. Parents are responsible for ensuring these materials are safe, age-appropriate, and accessible while creating a supportive play environment.

17. **Early Literacy and Numeracy Skills in Children.** Explains the importance of early exposure to language and numeracy skills through everyday activities rather than formal instruction before school. Children develop communication, literacy, and numeracy skills naturally through interactions, storytelling, counting household objects, and recognizing shapes in their surroundings. Early literacy elements include awareness of print and signs, listening,

speaking, and recognizing letters, while early numeracy focuses on measurement, sorting, and geometric concepts. Parents play a crucial role in fostering these skills by integrating learning into daily routines, reading to children, and encouraging exploration through play.

18. **Child Development Center and Pre-Primary Classes.** Child Development Centers (CDCs) provide early learning experiences that are accessible for all children before formal schooling. Parents are instructed to enroll children above three years and below five years of age at CDCs where they can participate in pre-primary classes and learn in a safe and stimulating environment through activities like storytelling, music, and creative play. Parents can reinforce education received at CDCs by talking to their children about lessons and focusing on holistic development, not just educational progress.

19–20. **Disability Modules.** Raise awareness about childhood disabilities—their causes, types, and ways to support children effectively. Interactive methods such as storytelling and simple games are used to help caregivers understand the functional challenges children may face and how to address them with empathy and inclusion. Sessions emphasize the importance of early identification, reducing stigma, and creating supportive environments at home and in the community. Real-life examples are used to spark discussion and reflection, encouraging caregivers to embrace inclusive practices. Participants also share key definitions from Nepal's disability framework and guide families on how to seek relevant services and support.

E ASQ-3 Validation and Measurement

E.1 Overview

The Ages and Stages Questionnaire is a parent-completed developmental screening instrument originally developed by Squires, Bricker, and Potter (1997); we use its third edition (ASQ-3), documented in Squires and Bricker (2009). It is designed to identify children who may need further evaluation for developmental delays and to measure developmental progress at the population level. The ASQ has been translated into over 20 languages, normed in more than 15 countries, and adopted for national screening programs in the United States, Canada, Australia, South Korea, Brazil, and several European countries. Its widespread adoption reflects two strengths: feasibility (parent-completed, no clinician required, 10–15 minutes per administration) and psychometric validity (strong agreement with gold-standard clinical assessments).

E.2 Psychometric Properties

The ASQ-3 technical manual reports the following psychometric properties based on standardization samples and validation studies:

Table A10: ASQ-3 Psychometric Properties

Property	Value
Test-retest reliability (2-week interval)	92% agreement
Inter-observer reliability (parent vs. trained examiner)	93% agreement
Overall sensitivity (detecting true delays)	86%
Overall specificity (correctly identifying non-delays)	85%
Concurrent validity with Bayley Scales	85% agreement (range 76–91%)
Internal consistency (Cronbach's α)	0.51–0.87 by domain and age

Notes: Values represent overall averages reported in Squires and Bricker (2009). Specific figures vary by age interval and developmental domain. Inter-observer reliability refers to agreement between parents completing the ASQ at home and trained professional examiners independently assessing the same children — particularly relevant for assessing the scope of reporting bias.

The inter-observer agreement of 93% is particularly relevant: it indicates that parent reports overwhelmingly align with independent professional assessment, leaving limited scope for systematic reporting bias to drive measured treatment effects.

E.3 Cross-Cultural Validation

The ASQ has been validated across diverse cultural and economic contexts, demonstrating that its measurement properties generalize beyond Western standardization samples:

- **Latin America:** Filgueiras et al. (2013) validate the Brazilian adaptation of the ASQ, reporting acceptable psychometric properties. Schonhaut et al. (2013) confirm validity in Chilean populations including preterm infants.
- **South Asia:** Kvestad et al. (2013) document ASQ performance in Indian populations, noting that while absolute cutoff scores may require cultural adaptation, the instrument's ability to detect *within-population variation* — which is what our study relies on — is preserved.

- **East Asia:** Heo, Squires, and Yovanoff (2008) validate cross-cultural adaptations in Korean populations, demonstrating that the five-factor domain structure holds across settings.
- **Global synthesis:** Fernald et al. (2009) review child development measurement instruments for the World Bank and include the ASQ as a low-cost, parent-report screening tool usable across diverse low-resource settings.

E.4 Use in Economic Research - A few Examples

The ASQ has been employed as a primary or secondary outcome measure in rigorous economic evaluations of early childhood programs. A few examples are below:

- Araujo, Dormal, and Schady (2019) — childcare quality effects on development in Peru (Cuna Más)
- López Bóo and Creamer (2019) — conditional cash transfer effects in Honduras
- Weaver et al. (2025) — maternal cash transfers and child development in India
- Justino et al. (2023) — parenting program in Rwanda

E.5 Structural Features Reducing Susceptibility to Reporting Bias

Several design features of the ASQ make it structurally less susceptible to the reporting biases that affect attitudinal or recall-based self-report measures:

1. **Concrete behavioral anchors.** Each item describes a specific, observable action ("Does your child kick a ball by swinging his leg forward?" "Does your child string small items onto a string?") rather than eliciting attitudes or global impressions.
2. **Real-time performance protocol.** The administration manual instructs: "Try each activity with your baby before marking a response." Many items contain explicit test instructions ("Ask your child to...," "Do not repeat the numbers," "Give your child two or three tries before marking") that require the child to demonstrate in real time.
3. **Age-specific calibration.** Items are designed for skills that are *emerging* at each age (36, 48, or 60 months) — not already mastered — reducing scope for over-reporting.
4. **Three-point scale.** The Yes/Sometimes/Not Yet response format provides an intermediate option that absorbs uncertainty without forcing binary inflation.
5. **No obvious "correct" answer.** A respondent wishing to inflate scores would need to know which specific motor or cognitive tasks indicate age-appropriate development — knowledge that even professionals require scoring manuals to determine.

E.6 Domain Composition

The ASQ-3 assesses five developmental domains. Each age-specific questionnaire (36, 48, or 60 months) contains six items per domain (30 items total):

- **Communication:** Expressive and receptive language, following directions, vocabulary use
- **Gross Motor:** Large muscle movements — kicking, jumping, throwing, balancing, hopping
- **Fine Motor:** Small muscle control — drawing, cutting, stringing beads, copying shapes
- **Problem-Solving:** Cognitive skills — memory, counting, pattern recognition, following sequences
- **Personal-Social:** Self-care, social interaction, emotional regulation, independence

Items vary in their measurement approach. Some require real-time physical demonstration by the child (drawing a circle, catching a ball), some require verbal responses to structured prompts (repeating numbers, naming colors), and some rely on caregiver report of behaviors occurring in other contexts (home routines, playground behavior). In Section 6.4.1, we exploit this variation to construct sub-indices that isolate items where reporting bias is structurally precluded.

F ASQ Decomposition: Item Classification

F.1 Overview

The decomposition starts from a single tenet. We retain only items whose performance the child can be asked to demonstrate and that the enumerator can directly observe during the assessment, and we then tighten this requirement in stages. Items whose scoring depends on caregiver recall of behavior outside the assessment, such as home routines, habitual speech, and playground or social behavior, are set aside from the outset.

This appendix provides the complete item-by-item classification used to construct the three alternative ASQ indices reported in Table 4. Each of the 30 items per age group (36, 48, and 60 months) is classified into one of the following progressively stricter categories:

- **ASQ-O (Likely Observed):** Items where the child's behavior can be elicited and observed during the assessment visit. Excludes items requiring caregiver recall of behavior in other settings (home routines, habitual speech, playground behavior, social interactions with other children).
- **ASQ-T (Strict Test):** Items where the questionnaire contains at least one of: (i) an explicit imperative verb ("Ask your child to..."), (ii) an instruction to record the child's response, or (iii) a test protocol instruction ("Do not repeat the numbers," "Give 2–3 tries").
- **ASQ-P (Performance):** Items where the child must produce a physical/motor output (drawing, cutting, jumping, catching, building). Excludes all verbal responses and gestural responses.

Each index is a strict subset, with ASQ-P $\subset$ ASQ-T $\subset$ ASQ-O $\subset$ Original.

Because of this screen, the loosest index, ASQ-O, is already a strict subset of the Original ASQ. The Original includes recall-based items, for example *uses a spoon with little spilling*, *takes turns waiting*, and *walks up stairs at home*, that ASQ-O excludes. In the item tables below, such items carry no mark in any of the O, T, or P columns. They remain part of the Original ASQ but enter none of the three constructed indices, and this is by construction rather than an omission.

F.2 Summary

Age Group	ASQ-O (Observed)	ASQ-T (Strict Test)	ASQ-P (Performance)
36-month	23/30 (77%)	11/30 (37%)	7/30 (23%)
48-month	21/30 (70%)	12/30 (40%)	6/30 (20%)
60-month	23/30 (77%)	20/30 (67%)	12/30 (40%)

F.3 36-Month Items

#	Domain	Item Description	O	T	P
C1	Comm.	Point to at least 7 body parts when asked	✓		
C2	Comm.	Makes sentences of 3–4 words (habitual)			
C3	Comm.	Put book on table, shoe under chair (explicit "ask")	✓	✓	✓
C4	Comm.	Tell what is happening in a picture	✓		
C5	Comm.	Move zipper up/down (explicit "show...ask")	✓	✓	✓

#	Domain	Item Description	O	T	P
C6	Comm.	Say first and last name when asked	✓		
GM1	Gross Motor	Kick a ball by swinging leg forward	✓		
GM2	Gross Motor	Jump with both feet leaving floor	✓		
GM3	Gross Motor	Walk up stairs one foot per stair (home/store)			
GM4	Gross Motor	Stand on one foot for 1 second	✓		
GM5	Gross Motor	Throw ball overhand	✓		
GM6	Gross Motor	Jump forward 6 inches	✓		
FM1	Fine Motor	Draw vertical line (explicit “ask her to make”)	✓	✓	✓
FM2	Fine Motor	String beads onto string	✓		
FM3	Fine Motor	Draw circle (explicit “ask him to make”)	✓	✓	✓
FM4	Fine Motor	Draw horizontal line (explicit “ask her to make”)	✓	✓	✓
FM5	Fine Motor	Try to cut paper with scissors	✓		
FM6	Fine Motor	Hold pencil between fingers like adult	✓		
PS1	Prob-Solv.	Copy 4 objects lined up (explicit “line up...copies”)	✓	✓	✓
PS2	Prob-Solv.	Find chair to reach things (habitual)			
PS3	Prob-Solv.	Identify figure (“point and ask...write response”)	✓	✓	
PS4	Prob-Solv.	Repeat “7–3” (“do not repeat...try another”)	✓	✓	
PS5	Prob-Solv.	Copy bridge with blocks (explicit “show your child”)	✓	✓	✓
PS6	Prob-Solv.	Repeat “5–8–3” (same protocol)	✓	✓	
S1	Pers-Soc.	Use spoon with little spilling (daily routine)			
S2	Pers-Soc.	Push wagon, steer around objects (habitual play)			
S3	Pers-Soc.	Say “me” or name in mirror	✓		
S4	Pers-Soc.	Put on coat by himself (daily routine)			
S5	Pers-Soc.	Answer “are you a girl or boy?” (explicit “ask”)	✓	✓	
S6	Pers-Soc.	Take turns waiting (social behavior)			
Totals			**23**	**11**	**7**

Note. Items with no mark in any column are recall-based (home routine, habitual, playground, or social) and are excluded from all three indices. They remain part of the Original ASQ only.

F.4 48-Month Items

#	Domain	Item Description	O	T	P
C1	Comm.	Name 3 items from a category	✓		
C2	Comm.	Answer “what do you do when hungry/tired?” (write response)	✓	✓	
C3	Comm.	Tell 2 things about common objects	✓		
C4	Comm.	Use word endings -s, -ed, -ing (habitual)			
C5	Comm.	Follow 3 unrelated directions (“give all three”)	✓	✓	✓
C6	Comm.	Use all words in a sentence (habitual)			
GM1	Gross Motor	Catch large ball (“give 2–3 tries”)	✓	✓	✓
GM2	Gross Motor	Climb playground slide (playground)			

#	Domain	Item Description	O	T	P
GM3	Gross Motor	Throw ball overhand 6 feet	✓		
GM4	Gross Motor	Hop on one foot	✓		
GM5	Gross Motor	Jump forward 20 inches	✓		
GM6	Gross Motor	Stand on one foot 5 seconds (“give 2–3 tries”)	✓	✓	✓
FM1	Fine Motor	Put together 5–7 piece puzzle	✓		
FM2	Fine Motor	Cut paper on line (“draw a line” — imperative)	✓	✓	✓
FM3	Fine Motor	Copy at least 3 shapes (child copies)	✓	✓	✓
FM4	Fine Motor	Unbutton one or more buttons	✓		
FM5	Fine Motor	Draw person with 3+ features	✓		
FM6	Fine Motor	Color within lines (habitual)			
PS1	Prob-Solv.	Repeat “5–8–3” (“do not repeat...try another”)	✓	✓	
PS2	Prob-Solv.	Point to smallest circle (“ask without pointing”)	✓	✓	
PS3	Prob-Solv.	Follow 3 directions with under/between/middle	✓	✓	✓
PS4	Prob-Solv.	Name 5 colors (“shown and asked”)	✓	✓	
PS5	Prob-Solv.	Dress up and play-act (habitual)			
PS6	Prob-Solv.	Count 5 objects (“place objects...ask without help”)	✓	✓	
S1	Pers-Soc.	Serve food from container (daily routine)			
S2	Pers-Soc.	Tell name/age/gender/city	✓		
S3	Pers-Soc.	Wash hands with soap (daily routine)			
S4	Pers-Soc.	Name 2+ playmates (“ask without telling names”)	✓	✓	
S5	Pers-Soc.	Brush teeth without help (daily routine)			
S6	Pers-Soc.	Dress/undress without help (daily routine)			
Totals			**21**	**12**	**6**

Note. Items with no mark in any column are recall-based (home routine, habitual, playground, or social) and are excluded from all three indices. They remain part of the Original ASQ only.

F.5 60-Month Items

#	Domain	Item Description	O	T	P
C1	Comm.	Follow 3 unrelated directions (“give all three”)	✓	✓	✓
C2	Comm.	Use 4–5 word sentences (habitual)			
C3	Comm.	Use past tense (“ask your child questions”)	✓	✓	
C4	Comm.	Complete comparisons (“ask your child questions”)	✓	✓	
C5	Comm.	Answer “what do you do when hungry/tired?” (write response)	✓	✓	
C6	Comm.	Repeat sentences back (“read one sentence at a time”)	✓	✓	
GM1	Gross Motor	Throw ball overhand 6 feet	✓		
GM2	Gross Motor	Catch large ball (“give 2–3 tries”)	✓	✓	✓
GM3	Gross Motor	Stand on one foot 5 seconds (“give 2–3 tries”)	✓	✓	✓
GM4	Gross Motor	Walk on tiptoes 15 feet	✓		
GM5	Gross Motor	Hop on one foot 4–6 feet (“give two tries”)	✓	✓	✓

#	Domain	Item Description	O	T	P
GM6	Gross Motor	Skip using alternating feet	✓		
FM1	Fine Motor	Trace on line (“ask your child to trace”)	✓	✓	✓
FM2	Fine Motor	Draw person (“ask your child to draw”)	✓	✓	✓
FM3	Fine Motor	Cut paper on line (imperative setup)	✓	✓	✓
FM4	Fine Motor	Copy shapes (child copies)	✓	✓	✓
FM5	Fine Motor	Copy letters (“cover up all except” — protocol)	✓	✓	✓
FM6	Fine Motor	Copy own name (“print...can child copy?”)	✓	✓	✓
PS1	Prob-Solv.	Point to smallest circle (“when asked...point”)	✓	✓	
PS2	Prob-Solv.	Name 5 colors (“when shown and asked”)	✓	✓	
PS3	Prob-Solv.	Count to 15 without mistakes	✓	✓	
PS4	Prob-Solv.	Complete opposite sentences (sentences given)	✓	✓	
PS5	Prob-Solv.	Identify numbers 3, 1, 2 (shown, child names)	✓	✓	
PS6	Prob-Solv.	Name 4+ letters in name (“point and ask”)	✓	✓	
S1	Pers-Soc.	Serve food using utensils (daily routine)			
S2	Pers-Soc.	Wash hands and face (daily routine)			
S3	Pers-Soc.	Tell name/age/gender/city	✓		
S4	Pers-Soc.	Dress/undress including buttons (daily routine)			
S5	Pers-Soc.	Use toilet independently (daily routine)			
S6	Pers-Soc.	Take turns and share (social behavior)			
Totals			**23**	**20**	**12**

Note. Items with no mark in any column are recall-based (home routine, habitual, playground, or social) and are excluded from all three indices. They remain part of the Original ASQ only.

G Outcome Variables

G.1 Overview

This appendix provides details on the pre-specified outcomes in the pre-analysis plan registered at RCT ID: AEARCTR-0011318.

G.2 Secondary Outcomes: Learning Environment & Practices and Teacher-Student Interaction

Outcome	Definition
Teacher–Student Interaction	*Response Options: 1 if Yes, 0 if No.* *Note:* There are two indices: (i) Positive index *The sum of responses to B1, B2, B3, B4, B5, B6, B7, B9, and B10 across 20 rounds of observation.* (ii) Negative index *The sum of responses to B8, B11, B12, B13, and B14 across 20 rounds of observation.* 1. Reading to a small/large group. 2. Teachers or Students singing. 3. Circle time: Child leading a whole class activity. 4. Back and forth discussion/conversation in a circle time. 5. Back and forth discussion/interaction/conversation with a small group or one child. 6. Engaging in pretend play with children/child. 7. Telling child or children what to do (not back-and-forth discussion). 8. Negative discipline. 9. Calmly negotiating conflict between peers. 10. Smiling/positive affect. 11. Negative affect/yelling. 12. Teacher is sitting at her desk or not engaged with children. 13. Teacher is talking with another adult. 14. Teacher out of the classroom.

Learning Environment and Practices Index

Coded: 1 if Yes, 0 if No, discarded if N/A
The index will be the additive sum of the responses

1. Teacher has prepared yearly, monthly and daily plans.
2. Teachers follows standard daily time table (as learnt during the training).
3. Stories, poems and songs are regularly used in ECD center.
4. ECD center and classroom are regularly cleaned.
5. Access to safe drinking water.
6. Access to Child friendly washroom and toilets.
7. Child-friendly, clean (free from dust); adequate space classroom; Proper seating arrangement in ECD center (Use of carpet, pre fab, cushions, round tables etc.
8. Six learning areas management and arrangement of learning materials accordingly.
9. Language learning areas arrangement and management.
10. Science learning areas arrangement and management.
11. Play/Act/role-play skills related learning areas arrangement and management.
12. Creativity learning areas arrangement and management.
13. Math learning areas arrangement and management.
14. Construction/Materials development learning areas arrangement and management.
15. Use of learning and play materials and conducts activities accordingly.
16. Teacher takes daily attendance of all children (attendance registers maintained).
17. Teacher also conducts appropriate and disable-friendly activities for children with disability.
18. Use of ECD Kit Box in children's daily activities.
19. Regular observation/monitoring/supervision and guidance by HT.
20. Teachers have maintained records on students' progress (as per the prescribed formats).

G.3 Child Primary Outcomes: Ages and Stages Questionnaire Score

Note: *Each response in this subsection will be scored in accordance with the standard methodology used for scoring ASQ in the literature. Aggregate score in this section will be the sum of individual scores for all these questions.*

Aggregate ASQ score is the sum of scores in all five subsections: Communication, Gross Motor, Fine Motor, Problem- solving, and Personal-social.

Subsection of ASQ (36 month)	Definition
Communication	1. When you ask your child to point to her nose, eyes, hair, feet, ears, and so forth, does she correctly point to at least seven body parts? 2. Does your child make sentences that are three or four words long? 3. Without giving your child help by pointing or using gestures, ask him to "put the book on the table" and put the shoe under the chair". Does your child carry out both of these directions correctly? 4. When looking at a picture book, does your child tell you what is happening or what action is taking place in the picture (for example, "barking","running", "eating" or "crying"?) 5. Show your child how a zipper on a coat moves up and down, and say, "See, this goes up and down". Put the zipper to the middle and ask our child to move the zipper down. Return the zipper to the middle and ask your child to move the zipper up. Do this several times, placing the zipper in the middle before asking you child to move it up or down. Does your child consistently move the zipper up when you say "up" and down when you say "down"?
Gross Motor	1. Without holding onto anything for support, does your child kick a ball by swinging his leg forward? 2. Does your child jump with both feet leaving the floor at the same time? 3. Does your child walk up stairs, using only one foot on each stair? 4. Does your child stand on one foot for about 1 second without holding onto anything? 5. While standing, does your child throw a ball overhand by raising his arm to shoulder height and throwing the ball forward? 6. Does your child jump forward at least 6 inches with both feet leaving the ground at the same time?
Fine Motor	1. After your child watches you draw a line from the top of the paper to the bottom with a pencil, crayon, or pen, ask her to make a line like yours. Do not let your child trace your line. Does your child copy you by drawing a single line in a vertical direction? 2. Can your child string small items such as beads, macaroni, or pasta “wagon wheels” onto a string or shoelace?

3. After your child watches you draw a single circle, ask him to make a circle like yours. Do not let him trace your circle. Does your child copy you by drawing a circle?
4. After your child watches you draw a line from one side of the paper to the other side, ask her to make a line like yours. Do not let your child trace your line. Does your child copy you by drawing a single line in a horizontal direction?
5. Does your child try to cut paper with child-safe scissors? He does not need to cut the paper but must get the blades to open and close while holding the paper with the other hand.
6. When drawing, does your child hold a pencil, crayon, or pen between her fingers and thumb like an adult does?

Problem-Solving

1. While your child watches, line up four objects like blocks or cars in a row. Does your child copy or imitate you and line up four objects in a row?
2. If your child wants something he cannot reach, does he find a chair or box to stand on to reach it (for example, to get a toy on a counter or to "help" you in the kitchen?
3. When you point to the figure and ask your child, "What is this?" does our child say a word that means a person or something similar? Please write your child's response.
4. When you say, "Say 'seven three'", does your child repeat just the two numbers in the same order? Do not repeat the numbers. If necessary, try another pair of numbers and say, "Say 'eight two'".
5. Show your child how to make a bridge with blocks, boxes, or cans, like the example. Does our child copy you by making one like it?
6. When you say, "Say 'five eight three'", does our child repeat just the three numbers in the same order? Do not repeat the numbers. If necessary, try another series of numbers and say. "Say 'six nine two'".

Personal-Social

1. Does your child use a spoon to feed herself with little spilling?
2. Does your child push a little wagon, stroller, or toy on wheels, steering it around objects and backing out of corners if he cannot turn?
3. When your child is looking in a mirror and you ask, "Who is in the mirror?" does she say either "me" or her own name?
4. Does your child put on a coat, jacket, or shirt by himself?
5. Using these exact words, ask your child, "Are you a girl or a boy?" Does our child answer correctly?

	6. Does your child take turns by waiting while another child or adult takes a turn?

Subsection of ASQ (48 month)	**Definition**
Communication	1. Does your child name at least three items from a common category? 2. Does your child answer the following questions? (Mark "sometimes" if your child answers only one question.) a. "What do you do when you are hungry?" Please write your child's response. b. "What do you do when you are tired?" Please write your child's response. 3. Does your child tell you at least two things about common objects? 4. Does your child use endings of words, such as "-s," "-ed," and "-ing"? 5. Without your giving help by pointing or repeating, does your child follow three directions that are unrelated to one another? Give all three directions before your child starts. 6. Does your child use all of the words in a sentence (for example, "a," "the," "am," "is," and "are") to make complete sentences, such as "I am going to the park," or "Is there a toy to play with?" or "Are you coming, too?"
Gross Motor	1. Does your child catch a large ball with both hands? 2. Does your child climb the rungs of a ladder of a playground slide and slide down without help? 3. While standing, does your child throw a ball overhand in the direction of a person standing at least 6 feet away? To throw overhand, your child must raise his arm to shoulder height and throw the ball forward. 4. Does your child hop up and down on either the right or left foot at least one time without losing her balance or falling? 5. Does your child jumpforward a distance of 20 inches from astanding position, starting with his feet together? 6. Without holding onto anything, does your child stand on one foot for at least 5 seconds without losing her balance and putting her foot down?

Fine Motor

1. Does your child put together a five- to seven-piece interlocking puzzle?
2. Using child-safe scissors, does your child cut a paper in half on a more or less straight line, making the blades go up and down?
3. Using the shapes below to look at, does your child copy at least three shapes onto a large piece of paper using a pencil, crayon, or pen, without tracing?
4. Does your child unbutton one or more buttons?
5. Does your child draw pictures of people that have at least three of the following features: head, eyes, nose, mouth, neck, hair, trunk, arms, hands, legs, or feet?
6. Does your child color mostlywithin the lines in a coloring book or within the lines of a 2-inch circle that you draw?

Problem-Solving

1. When you say, "Say 'five eight three,'" does your child repeat just the three numbers in the same order? Do not repeat the numbers. If necessary, try another series of numbers and say, "Say 'six nine two.'"
2. When asked, "Which circle is the smallest?" does your child point to the smallest circle?
3. Without your giving help by pointing, does your child follow three different directions using the words "under," "between," and "middle"?
4. When shown objects and asked, "What color is this?" does your child name five different colors, like red, blue, yellow, orange, black, white, or pink?
5. Does your child dress up and "play-act," pretending to be someone or something else?
6. If you place five objects in front of your child, can he count them by saying, "one, two, three, four, five," in order?

Personal-Social

1. Does your child serve herself, taking food from one container to another using utensils?
2. Does your child tell you at least four of the following?
3. Does your child wash his hands using soap and water and dry off with a towel without help?
4. Does your child tell you the names of two or more playmates, not including brothers and sisters?

5. Does your child brush her teeth by putting toothpaste on the toothbrush and brushing all of her teeth without help?
6. Does your child dress or undress himself without help (except for snaps, buttons, and zippers)?

Subsection of ASQ (60 month)	Definition
Communication	1. Without your giving help by pointing or repeating directions, does your child follow three directions that are unrelated to one another? Give all three directions before your child starts. 2. Does your child use four- and five-word sentences? For example, does your child say, "I want the car"? Please write an example. 3. When talking about something that already happened, does your child use words that end in "-ed," such as "walked," "jumped," or "played"? Ask your child questions, such as "How did you get to the store?" ("We walked.") "What did you do at your friend's house?" ("We played.") Please write an example. 4. Does your child use comparison words, such as "heavier," "stronger," or "shorter"? Ask your child questions, such as "A car is big, but a bus is ..." (bigger); "A cat is heavy, but a man is ..." (heavier); "A TV is small, but a book is ..." (smaller). Please write an example. 5. Does your child answer the following questions? (Mark "sometimes" if your child answers only one question.) a. "What do you do when you are hungry?" Please write your child's response here: b. "What do you do when you are tired?" Please write your child's response. 6. Does your child repeat the sentences shown below back to you, without any mistakes? a. Jane hides her shoes for Maria to find. b. Al read the blue book under his bed.
Gross Motor	1. While standing, does your child throw a ball overhand in the direction of a person standing at least 6 feet away? To throw overhand, your child must raise his arm to shoulder height and throw the ball forward. 2. Does your child catch a large ball with both hands?

3. Without holding onto anything, does your child stand on one foot for at least 5 seconds without losing her balance and putting her foot down?
4. Does your child walk on his tiptoes for 15 feet (about the length of a large car)?
5. Does your child hop forward on one foot for a distance of 4–6 feet without putting down the other foot?
6. Does your child skip using alternating feet?

Fine Motor

1. Ask your child to trace on the line below with a pencil. Does your child trace on the line without going off the line more than two times?
2. Ask your child to draw a picture of a person on a blank sheet of paper. You may ask your child, "Draw a picture of a girl or a boy."
3. Draw a line across a piece of paper. Using child-safe scissors, does your child cut the paper in half on a more or less straight line, making the blades go up and down?
4. Using the shapes below to look at, does your child copy the shapes in the space below without tracing?
5. Using the letters below to look at, does your child copy the letters without tracing? Cover up all of the letters except the letter being copied.
6. Print your child's first name. Can your child copy the letters? The letters may be large, backward, or reversed.

Problem Solving

1. When asked, "Which circle is the smallest?" does your child point to the smallest circle?
2. When shown objects and asked, "What color is this?" does your child name five different colors, like red, blue, yellow, orange, black, white, or pink?
3. Does your child count up to 15 without making mistakes?
4. Does your child finish the following sentences using a word that means the opposite of the word that is italicized? For example: "A rock is hard, and a pillow is soft." Please write your child's responses.
    a. A cow is *big* and a mouse is ...
    b. Ice is *cold*, and fire is ...
    c. We see stars at *night*, and we see the sun during the ...
    d. When I throw the ball *up*, it comes ...
5. Does your child know the names of numbers?

	6. Does your child name at least four letters in her name? Point to the letters and ask, "What letter is this?" (Point to the letters out of order.)
Personal-Social	1. Can your child serve himself, taking food from one container to another, using utensils? 2. Does your child wash her hands and face using soap and water and dry off with a towel without help? 3. Does your child tell you at least four of the following? Please mark the items your child knows. 4. Does your child dress and undress himself, including buttoning medium-size buttons and zipping front zippers? 5. Does your child use the toilet by herself? (She goes to the bathroom, sits on the toilet, wipes, and flushes.) 6. Does your child usually take turns and share with other children?

G.4 Caregiver Primary Outcomes: Index of Stress and Practice of Early Stimulation and Learning

Outcome	Definition
Parent (Caregiver) Stress Index 1. My child rarely does things that make me feel good.	*Response Options: 1- Strongly disagree; 5-Strongly agree.* 2. Sometimes I feel my child doesn't like me and doesn't want to be close to me. 3. My child smiles at me much less than expected. 4. When I do things for my child, I get the feeling that my efforts are not appreciated very much. 5. I expected to have closer and warmer feelings for my child than I do, and this bothers me. 6. Sometimes my child does things that bother me just to be mean. 7. I feel that I am ... *(Response Options: 1-A very good parent; 5-Not very good at being a parent.)*

Practice of early stimulation and learning (Index) 1. Read books or looked at picture books with the child. 2. Told stories to the child.	*Response Options: 1 if Yes, 0 if No; The index will be the additive sum of the responses.* 3. Sang songs to or with the child (including lullabies)? 4. Took the child outside the home? 5. Played with the child. 6. Named, counted, or drew things for or with the child.

G.5 Care Givers' Secondary Outcomes

Outcome	Definition	Coding
Attitude to disciplining the Child	Do you think it's necessary or/and normal to spank to discipline your child?	0 - No, 1 - Yes
Intensity of discipline	In the past 14 days, how often have you spanked to discipline your child?	0 - Never, 1 - Once, 2 - Twice, 3 - More than twice
Intensity of early stimulation and learning (index)	1. How often did you engage in this activity with the child?	The index will be the additive sum of the responses. 1 - Only several times (1-3 days a week), 2 - Often (4-5 days a week), 3 - Nearly every day (6 days a week)
Attitude to Disability 2. Do you think schools should accommodate children with disabilities? 3. Do you think disabilities can be helped?	1. Do you believe children with disabilities should be separated from their parents?	

4. I believe that children with disabilities deserve access to educational opportunities	The index will be the additive sum of the responses. 1: 0 - Yes, 1 - No 2 & 3: 1 - Yes, 0 - No 4: 5 - Strongly Agree, 4 - Agree, 3 - Not sure, 2 - Disagree, 1 - Strongly Disagree
Feeling confident to help with disability 2. I feel confident in my ability to support children with disabilities.	1. I have the necessary knowledge and skills to support children with disabilities The index will be the additive sum of the responses. 5 - Strongly Agree, 4 - Agree, 3 - Not sure, 2 - Disagree, 1 - Strongly Disagree